2022

Engineering *Eubacterium limosum* towards commodity production

Alexander Agius

30th Of March, 2022

University of Nottingham
UK | CHINA | MALAYSIA

Declaration

Unless otherwise acknowledged, the work presented in this thesis is my own. No part has been submitted for another degree in the University of Nottingham or any other institute of learning.

Alexander Agius, 2022

Abstract

Climate change and sustainability have come to the fore in recent years, with a greater interest in transition away from fossil usage where possible. One of the most difficult uses of fossil fuels to replace is the production of precursors for the synthesis of chemicals essential for our modern civilization. C1 utilising organisms may be able to aid in the replacement of fossil fuels for this purpose, with the added benefit of potentially producing carbon neutral or carbon negative products.

Eubacterium limosum is an acetogenic bacterium which is not used currently in any industrial setting. It possesses interesting properties which may make it suitable for future use, from its utilisation of multiple energy sources, such as sugars, synthesis gas, carbon monoxide/CO_2 mixes and methanol, to its native production of chemicals such as butanol and vitamin B12. This work sought to develop this organism to be more amenable to industrial fermentation with the goal of addressing some of the organism's limitations and introducing a non-native chemical to its product spectrum. These limitations included the organism requiring yeast extract supplementation of growth media, significant adhesive properties, and a lack of a robust genetic toolset. All three of these issues were addressed as part of this work.

An improved strain was created by directed evolution using continuous fermentation. This resulted in a strain which no longer required complex medium supplementation (with yeast extract), and no longer had significant extracellular adhesive properties. In the investigation of the cause of the altered phenotype observed changes in the epigenome, hitherto not considered, suggested themselves as a likely explanation. Improved CRISPR-Cas9 tools were developed, and a toxin-antitoxin-based counter selection system employed. The use of and development of genetic modification tools produced many mutants with the following mutants created as part of this work, Δ*act*, Δ*cop*, Δ*hisI*, Δ*mti*, Δ*mtiI*, Δ*pbt1*, Δ*pbt2*, Δ*act*Δ*pbt1*, Δ*act*Δ*pbt2* and Δ*act*Δ*pbt1*Δ*pbt2*. The generated mutants bettered the understanding of the internal biochemistry of the organism as related to butyrate production and methanol usage, with the essentiality of the gene pair *metVmetF* confirmed by an inability to create a knockout strain. The creation of auxotrophic mutants allowed for use of allele-coupled exchange, a genetic modification technique, culminating in the genomic integration of genes from

Clostridium acetobutylicum which allowed for the production acetone and isopropanol. This work serves as a proof of concept for *E. limosum*, with the primary conclusion of this work being that *E. limosum* is an organism which is suitable for further development towards industrial use.

Acknowledgements

I would like to thank the people that helped make this work possible, in many cases this work could not have occurred or proceeded as it did without them.

I would like to thank my friends and family who supported me and helped me to be as involved in this as I was, this sort of work becomes your life for the period you're doing it and without them I doubt that I could have finished it.

Doctor Robert Mansfield was my initial bench supervisor. He got me set up and started the ball rolling on the techniques that I would use throughout this work. His rapid responses to questions proved very useful as I was finding my feet in a field, I had no prior experience in. He left as bench supervisor to start his own company and I wish him and his partners well.

Doctor James Millard took over as my bench supervisor and was very helpful throughout this work, not only for his understanding of both practical techniques and the theory behind them, but also for his understanding of the molecular biological processes. Indeed, his adaptation of the toxin-antitoxin system for *E. limosum* allowed work to accelerate to somewhat compensate for the Covid-19 shutdown that occurred while this work was underway. His help was above and beyond what would be expected of a bench supervisor, and he is an exemplar in this regard.

Professor Nigel Minton was my primary supervisor on this work, and despite his many commitments he was able to take the time to help when needed and meet with me throughout the project. This work could not have occurred without him, from his securing of funding for this work, via the Era CoBiotech BioMetChem project, to his

willingness to describe projects to me at a poster session which was displaying the PhD projects available.

Doctor Ying Zhang was a helpful secondary supervisor and helped me not only with her recommendations on both my work and thesis but also with my search for a place to live as my deadline was extended.

Professor Philippe Soucaille was, in addition to a friendly and helpful tertiary supervisor and contributor to the Era CoBiotech BioMetChem project, an essential contributor to this work. The supplying of the NG-6849 strain engineered in his lab allowed for much time saving and allowed for useful genetic comparisons to be performed.

I would also like to thank the lab users of B60, many of whom provided useful guidance and help on various parts of this project, from training on equipment to health and safety briefings.

Finally, I would like to thank the team at Deep Seq in Nottingham who performed my genomic sequencing and methylation analysis, the data they generated was key to this work.

Table of contents

List of figures

List of tables

Abbreviations

%	Percentage
℃	Degrees Celsius
AE	Allelic exchange
ACE	Allelic coupled exchange
BLAST	Basic Local Alignment Search Tool
BLASTp	Basic Local Alignment Search Tool Protein
BLASTn	Basic Local Alignment Search Tool Nucleotide

colony PCR	Colony Polymerase Chain Reaction
crRNA	CRISPR RNA
dH_2O	Distilled Water
ELGA Water	Water purified using an ELGA water purification system.
FAST	Fluorescence-Activating and Absorption-Shifting Tag Protein
FOA	Fluorinated nucleotide 5-Fluoroorotic Acid
GABA	Gamma Aminobutyric Acid
GC	Gas chromatography
GCMS	Gas chromatography with mass spectrometry
gDNA	Genomic DNA
sgRNA	Single guide RNA
HPLC	High Performance Liquid Chromatography
Indel	Insertion/Deletion
kV	Kilovolt
KAIST	Formerly known as “Korea Advanced Institute of Science and Technology”
mA	Milliamps
mL	Millilitre
mm	Millimetre
mM	Millimolar
ms	Millisecond
MS	Mass spectrometry
nm	Nanometre
ng	Nanogram
ng/µL	Nanograms per microlitre
MtaABC	Methanol:coenzyme M Methyltransferase
PAM	Protospacer adjacent motif
PCR	Polymerase Chain Reaction
P_{araE}	*Clostridium acetobutylicum araE* promoter
P_{fdx}	*Clostridium sporogenes fdx* promoter

P_{thl}	*Clostridium acetobutylicum thl* promoter
SNPs	Single nucleotide polymorphisms
SOE PCR	Splicing by overhang extension PCR
TAKOS	Toxin-antitoxin knockout system
THF	Tetrahydrofolate
tracrRNA	Trans-activating crRNA
V	Volts
μg	Microgram
μL	Microlitre
Ω	Ohm

Chapter 1

Introduction

Introduction

1.1 Background

Fossil fuels are the primary drivers of our industrialised civilisation. They are used in a wide number of processes, from logistics to the production of chemicals, and for energy production (1). Unfortunately, there are significant drawbacks to their use. The supply of fossil fuels is finite and diminishing, with some estimates suggesting approximately 50 years of oil at current usage rates (2). As poorer nations develop, it is likely that demand for fossil fuels will increase. In addition, there are the geopolitical issues surrounding fossil fuels like oil (3). Some nations possess vast reserves, while others possess none, causing power differences and creating a resource-based incentive for conflict. These problems are likely to worsen as reserves run out and demand increases.

Further to these concerns is climate change. The prevailing theory is that rising CO_2 emissions, since the start of the industrial revolution, have increased the greenhouse effect of our atmosphere and resulted in an increase in global temperature (4). These rising global temperatures have become a concern for the international community, with recent international agreements being created to try to curtail this changing climate. The Paris agreement specifically seeks to limit global temperature rise to 2°C above that of pre-industrial temperatures (5). The target, therefore, is to reduce greenhouse gas emissions, primarily from fossil fuel extraction, processing, and combustion. Furthermore, particulate pollution and other pollutants are also of concern, with particulate pollution being detrimental to human health, especially for those with respiratory conditions (6), resulting in many deaths every year (7).

While emission reduction will play an important role, reducing the CO_2 in the atmosphere to pre-industrial levels is key to limiting temperature increases or reversing the increase. Artificial carbon capture and storage is likely to play an important role in climate change. Indeed, the Intergovernmental Panel on Climate Change (IPCC) suggests carbon capture will be a useful tool in achieving the target of limiting climate change (8). Some companies with significant funding have already made some headway into the endeavour of carbon capture technologies (9). The

primary obstacle is the commercial viability of carbon capture processes. Ideally a saleable product would be produced to offset the costs of carbon capture and make it profitable. It is worth noting, however, that if the produced product is a fuel that will be combusted, releasing CO_2, the process will not have the effect of reducing CO_2 present in the atmosphere. While carbon-neutral fuels would help emissions to plateau, they would not reduce atmospheric CO_2 levels by themselves. However, such an outcome is not the desired endpoint of carbon capture, where the goal is the reduction of atmospheric CO_2 levels.

If carbon capture is used to produce precursor chemicals for non-combustion processes, such as plastic production, then carbon will be sequestered into these chemicals, thereby reducing atmospheric CO_2. If large amounts of carbon are sequestered, such as into disused oil and gas wells (10), carbon capture could have a significant impact on atmospheric CO_2. Carbon capture can become an important part of future climate strategy, especially if there is a shift to carbon-neutral energy production using nuclear or renewables technologies, thereby reducing emissions (8). Natural means of carbon capture are also possible, with soil sequestration (11,12) and planting of trees both being able to sequester carbon (13). However, trees need land. This land must be able to support plant life well enough to grow large forests. In a world where land is needed for human agriculture, it is unlikely to be tenable to use trees alone to sequester carbon (8). So, a combination of multiple approaches is likely to be the most effective.

This goal presents challenges. If we are to reduce or remove the cause of most pollution, we necessarily need to limit the primary driver of our civilisation. We will still need energy, the ability to transport goods and people, and precursor chemicals. Further, it is likely demand for these will increase as the population increases (14). For all the reasons stated thus far, we need alternatives to fossil fuels. Alternatives must have as minimal negative effect on the environment as possible. They must be able to compete with fossil fuels by keeping up with the vast demand, and they must be economically viable. Competing economically is a difficult proposition, with fossil fuels already having its technology developed, its infrastructure developed, and in many cases, having state subsidies (14). While some processes such as energy

production may be replaced with existing technology, primarily nuclear and renewable energy (15), the other functions of fossil fuels require new technology developed.

Liquid fuels are likely to be of major importance for many years to come. While electric vehicles are likely to dominate in the future, they are hampered by the use of lithium-based batteries, which are expensive, heavy, are highly polluting to produce and dangerous should an accident occur (16). Further, such systems are impractical for large container ships and aircraft. As such, alternatives to petrochemical-based liquid fuels are desirable. Some progress has been made with the production of ethanol, primarily in Brazil using sugarcane (17–19) and via carbon capture (9), as an alternative source for combustion-powered road vehicles. While ethanol may be useful in smaller sea going vessels it is currently unsuitable for large container ships (20). Biodiesel, however, may be suitable for shipping, especially when using diesel-style engines (20). Large aircraft cannot make use of ethanol; however, other biofuel alternatives may be viable, with blended fuels using kerosene and a biofuel being effective (21). Further. bio-derived synthetic paraffinic kerosene, kerosene derived from triglycerides and free fatty acids, offers a renewable source for kerosene, and they are under investigation by many aviation companies (22).

Chemical precursors are another important function of petrochemicals. Precursor chemicals are used in various processes, from the production of lubricants and paint to pharmaceuticals and plastics (23). All these needs will continue, and with development of many nations demand is likely to increase. This use is the least publicised yet is arguably the most important. Humans can live in a world without electricity and cars, but many humans cannot live without the life-saving medicines and drugs that some precursor chemicals are used to make.

Alternative methods to produce liquid fuels and important chemicals and precursors are needed, preferably technologies that allow for renewable production of said chemicals and ideally those with as minimal and impact on the environment as possible. The solution to these problems may lie with microorganisms. Bioprocesses have gained attention in recent years, though they have existed for many years.

Putting aside the pre-industrial uses in food products, bioprocesses have been used industrially for over 100 years. Chaim Weizmann famously used acetone-butanol-ethanol fermentation of *C. acetobutylicum* to produce acetone to support the British war effort in the First World War, where, just as today, acetone was an important precursor chemical (24). While this process was abandoned on an industrial scale as fossil fuel-based methods were found, primarily the cumene process (25), there are several important differences between the interwar period and today. Firstly, fossil fuel stocks were much larger in the early 20th century. Secondly climate change was not understood as it is today, and pollution mitigation was not the priority it is today. Finally, improvements in technology have allowed for techniques that were not possible in 1916. Not only has fermentation technology improved due to automation and the use of computers, but the technology regarding strain development has also improved. Adaptive fermentation was the only effective means to bring about phenotypical change before the current understanding of genetics and genetic modification (26). The tools now available to researchers allow for specific directed genetic changes to be made and allow for a better understanding of the metabolism and genetics of industrial strains. Further, with a greater catalogue of organisms known, there are far more starting points for strain development, with novel strains natively producing products on desired feedstocks, and thermophilic organisms able to reduce cooling costs and reduce contamination risks by operating at high temperature.

Bioenergy with carbon capture and storage is an interesting proposal that seeks to be significantly carbon negative, rather than just mitigating emissions or somewhat counteracting emissions of the process. This is achieved by the sequestration of carbon out of the atmosphere. Sequestration is likely to be important if the damage already done by CO_2 emissions is to be undone and levels returned to pre-industrial levels. Sequestration can take many forms, including the production of chemicals that are not intended for use in combustion (27,28). These chemicals can be used as precursors or used in their raw form, potentially becoming an economic building block and may act as part of the move towards a carbon-neutral economy.

1.2 Feedstock

Organisms need energy and nutrients to grow, and bacteria are no different in this regard. Feedstocks supply bacteria with the resources to grow, reproduce, and produce the desired product. Feedstock is a significant cost involved in industrial-scale production (29). Many organisms can natively produce useful products, such as acetone (24,30), on a small scale in a laboratory. Large-scale production in a commercially viable manner to compete with petrochemicals depends primarily on product yield and the feedstock cost. Selecting an appropriate feedstock before developing an industrial strain is, therefore, of paramount importance. It must be available, cheap and must be useable to produce the product of interest; be that biomass or chemicals produced by cells. While sugar feedstocks have been widely used for many organisms (29), they have the distinct disadvantage of being a food product and are processed from crops (31). If sugars are being grown to supply industry, then the land used to produce the sugar is not being used for growing crops to feed the growing population (31).

Therefore, utilising arable farming land to produce feedstocks for fermentation is not tenable due the increasing human population and the increased need for food (31). This is further compounded by the likelihood of sea levels rising as climate change progresses, decreasing the land that will be accessible (32). While research is being done into improving plant yields (33–35), improving the resilience of plants to make more land usable (36,37) and in actually trying to increase the amount of land that is useable (38,39), the increasing population makes it likely that these developments are essential to sustain sustenance for the population. Therefore, it is preferable to use feedstocks from alternative sources, such as waste.

Single carbon chemicals, or C1, such as methane and methanol, are attractive alternatives to conventional feedstocks. Not only do C1 feedstocks not waste valuable farming land but using them can make use of chemicals considered waste. Conversion of waste into C1 can reduce the cost and complexity of making use of waste (40) and make engineering organisms simpler, with only one feedstock needing to be accounted for when designing an industrial strain. The viability of using waste as a

means of C1 generation will depend on the waste source, with waste gas from steel mills providing an excellent source of CO_2, and anaerobic digestion providing a source of methane. The use of biomass waste converted into C1 could therefore be far cheaper than sugars. It is unlikely that waste streams alone would be sufficient to replace petrochemicals, though they may be of use in developing the technology to eventually supplant them. Direct air carbon capture could be a viable source of uncontaminated CO_2 for methanol production (27,28). This carbon capture is likely to be scalable (41,42), and if powered by either renewable or atomic energy, could be sufficient to capture enough CO_2 as to be a viable, though likely more expensive, alternative to petrochemicals, at least as regards carbon-based precursor chemicals (43).

1.2.1 Syngas

Synthesis gas (or syngas) is a mixture of the gases carbon monoxide (CO), carbon dioxide (CO_2) and hydrogen (H_2). Syngas is most commonly produced from methane via steam reforming of the methane gas (44,45). Syngas can be generated from many carbon sources, including biomass, and is a waste product of steel mills (46). Syngas can also be produced from CO_2 and water using high-temperature electrolysis. Syngas can hence be produced renewably via electrolysis, assuming the electricity is generated through renewable means (47). Syngas is useful as a carbon source for fermentation (48) and can be converted into methanol via hydrogenation of CO_2 and carbon monoxide by hydrogen, all of which are present in syngas (44,45). Biological conversion of syngas offers several advantages compared to chemical conversion. Biological processes operate at far lower pressures and temperatures, are able to deal with minor contamination of the syngas, such as a sulphur, which is generally very damaging to catalysts (49,50), and are better able to deal with variances of gas composition (51).

1.2.2 Methanol

C1 compounds are of interest as a feedstock for acetogens for the production of valuable chemicals (52). Although methane may be the cheapest C1 carbon source (53), it has the significant disadvantage of being a gas. Being a gas makes its storage and transport far more difficult as it must be transported as a cryogenic liquid, which requires specialised infrastructure and carries with it significant safety concerns (54). Methanol, on the other hand, is a liquid at room temperature and so avoids the issues faced by cryogenic liquids, though it is a very volatile liquid. The majority of methanol production comes from the use of synthesis gas, but it is also possible to produce it from other sources such as coal and biomass (44,55). There are also suggestions that, utilising renewable energy, electrolysis of water, and carbon capture, methanol could be produced in a more environmentally friendly and renewable manner (56,57). Methanol can also be produced from solid organic wastes, such as lignocellulosic biomass, using thermochemical decomposition under oxygen-limited conditions (40).

Methanol has been proposed as the future economic currency, just as fossil fuels are today, with its main competition being the “hydrogen economy” and the “ethanol economy” (58,59). Methanol and ethanol have significant advantages over hydrogen, as liquids they are easier to transport, maintain, and can utilise existing infrastructure currently used for petroleum (60). Due to the small size of hydrogen molecules, it is very difficult to maintain leak-proof containers (61). Even if this is done, the hydrogen must be stored pressured as a cryogenic liquid (61), which raises serious safety concerns with gas leakage, leading to fires and explosions (54), and with cryogenic liquid, causing severe cold burns and structural damage (54). The economics of hydrogen are also not favourable, with the immense cost of new infrastructure and its storage (61,62). Methanol also has advantages over ethanol. The first advantage is the improved production of methanol from existing fossil fuels when compared to ethanol (63). This is of interest as methane, which may normally be vented or wasted could be used to produce methanol, decreasing the environmental impact of existing fossil fuel extraction. The second advantage is that renewable methanol can be produced without the need for arable land, whereas ethanol generally

requires it (64); if ethanol is produced from a methanol feed acetogen, this is not the case, although strictly speaking, it would still qualify as part of a methanol economy as methanol is the building block for that ethanol. It is worth noting however, that ethanol is currently being produced from a C1 feedstock. Lanzatech has made great strides in the use of waste gases from steel mills, feeding *C. autoethanogenum* to produce ethanol (9,55). This system can produce large quantities of ethanol and acts to lessen the pollution effects of the plants it is attached to. Such a system may serve as a form of commercially viable carbon capture.

When used as a feedstock for biological processes, methanol has distinct advantages over syngas and gas feedstocks generally, primarily due to it being a liquid. Water and methanol are miscible with one another and mix with ease, such that in a conventional bioreactor, it is relatively probable that a consistent concentration can be equalised across the reactor vessel. With all gases, there is a limiting factor of the rate at which a gas can dissolve in water (65). This gas-liquid transfer must occur for the gas to act as a feedstock and be taken up by organisms in the reactor. This can result in more feedstock being wasted when compared to miscible liquids, though this can be mitigated recycling gas in the reactor setup (66). Reactors can also be designed increase the gas-liquid transfer, with membranes (67) and microbubble systems (66), but these add to the cost and complexity of a reactor. Biomass can also be retained to increase the biomass concentration within the bioreactor (68–70). Increased cost decreases the economic viability of an endeavour and complexity increases the chance of breakdown, which again hampers economic viability.

1.3 Bacterial hosts

The choice of bacterial host is vitally important. The host determines the feedstocks that are natively useable, the native product spectrum, and the availability of molecular biology tools. It also determines how resilient a fermentation setup may be to variations in conditions and the risk of contamination. While it is possible to genetically modify organisms or to create new molecular biology tools, it takes time and money to do so, the expenditure of which can be avoided if a more suitable organism is chosen in the first instance. Well-characterised organisms or organisms

which have extensive industrial use may be of great interest for this work if they can be used with the project constraints.

1.3.1 *Clostridium acetobutylicum*

Clostridium acetobutylicum is a famous anaerobe that was first used industrially in the early 20th century (24). The organism has been genetically modified using ClosTron (71), allelic exchange, and CRISPR-Cas9 (72). However, it has primarily been used with a sugar feedstock and there is no evidence that the organism can make use of methanol. Although the organism was not selected for use in this work, its genes would still play a role in this project as they will be used to attempt to engineer acetone production (5.1).

1.3.2 Acetogens

Acetogens are anaerobic bacteria that utilise the Wood-Ljungdahl pathway (73) for energy conservation and carbon fixation, reducing CO or CO_2 and producing acetyl-CoA from the assimilated carbon. Their primary fermentation product is acetate (74,75). They are of interest with respect to C1 feedstock due to their ability to natively grow on C1. In addition to acetate, some acetogens can also produce additional products such as ethanol, butanol, and butyrate (74). Acetogens are found all over the world, generally in ecological niches (76) such as in the digestive tracts of animals and in decomposing organic matter (77). This likely owing to the acetogen's use of the Wood-Ljungdahl pathway, which was critical to the early evolution of life on earth (78,79). Since their initial discovery, an early isolate being *Clostridium aceticum* (80), much research has been performed on acetogens, from metabolic characterisation (81) to genomic and phenotypical characterisation (82). Acetogens are currently being used on an industrial scale today. Ethanol production by *Clostridium autoethanogenum* has been piloted by Lanzatech, utilising waste gas from steel mills, they are able to produce over 100,000 gallons of ethanol per year (83). Genetic modification of acetogens offers a potential route towards the production of more complex chemicals from C1 feedstocks. With the advantages

methanol offers as a feedstock (1.2.2), acetogenic bacteria that can make use of methanol as a feedstock are of interest for development as part of this work.

1.3.2.1 *Moorella thermoacetica*

Moorella thermoacetica, previously known as *Clostridium thermoaceticum*, is a model acetogen that has been used to investigate various acetogenic pathways (84). Its use as a model organism offers advantages when considering it as a platform for chemical production. The existence of a sequenced genome (85) and a genome-scale model (86) are a good start. *M. thermoacetica* is known to be able to use methanol, as well as sugars and other alcohols (84).

However, *M. thermoacetica* is a thermophile (84), and grows optimally between 55 and 60°C. While not a problem in and of itself, issues arise when considering the application of this project, which involves gas fermentation with methanol as a carbon source. Methanol has a boiling point of 64.7°C, which is only a few degrees above the maximum temperature of *M. thermoacetica* optimal range. This would result in more methanol evaporation in a bioreactor than would occur with a mesophilic organism, wasting the carbon source and increasing operating costs, thereby decreasing commercial viability.

Furthermore, this thermophilic nature causes issues for many lab-based techniques. Viable antibiotic resistance genes for use as counter-selection would be more limited and would require growth at the lower end of optimal growth to be functional (87). Additionally, *M. thermoacetica* cannot grow on agar plates, and roll tubes pressurised with gas must be used (88–91), making picking of individual colonies very difficult. Although transformation of plasmids has occurred, this had only been done with suicide vectors which are unable to self-replicate (91). No clean in-frame scarless genome editing technology had been published in this organism at the commencement of this work. As this project was not intended to centre around tool building, *M. thermoacetica* was not selected.

1.3.2.2 *Clostridium autoethanogenum*

Clostridium autoethanogenum is an anaerobic bacterium currently in use industrial-level production of products (83). Genetic modification, using ClosTron, allelic exchange, and CRISPR-Cas9, has been demonstrated in *C. autoethanogenum*, allowing for metabolic engineering to be performed (92,93). It is able to grow on synthesis gas and carbon monoxide in addition to sugars (92). However, it is not able however to grow on methanol, the feedstock of interest for this work, and despite its long-standing reputation and mature genetic tools, it was not selected for use in this work.

1.3.2.3 *Clostridium carboxidivorans*

Clostridium carboxidivorans is an acetogen with commercial potential (94). It is capable of producing much more complex products than many other acetogens, including butanol, hexanoate, and hexanol (95) in addition to typical products such as acetate, ethanol and butyrate. While it is able to make use of C1 chemical such as CO and CO_2/H_2 (96) it cannot make use of methanol (94). It is likely that *C. carboxidivorans* lacks a MtaABC complex homologue. This complex is present in methanol utilising acetogens, such as *Eubacterium limosum*, where it acts to assimilate methanol into the Wood-Ljungdahl pathway (97,98). While it may be possible to introduce this complex via genetic modification, *C. carboxidivorans* had not been successfully genetically modified at the time this work started. Development of a genetic toolset that would work in *C. carboxidivorans* would be a significant piece of work in of itself and indeed it was, with researchers at the University of Nottingham working on that very problem at the time this research was performed. The lack of methanol usage, and the lack of successful genetic modification system present at the time of the start of this work resulted in this organism not being used in this work, despite its very desirable product spectrum, though it is likely of interest for future researchers.

1.3.2.4 *Acetobacterium woodii*

Acetobacterium woodii is an acetogenic bacterium that has been well characterised in the literature (75,99–105), and its genetic modification is well established (92). While it can make use of methanol (106), its ability to grow on methanol alone for long periods of time is limited and unreliable. Therefore, despite the positives of *A. woodii*, it was not the ideal organism for this work and was not used.

1.3.2.5 *Eubacterium limosum*

Eubacterium limosum is an acetogenic bacterium that can natively produce acetate, butyrate, ethanol and butanol while transforming syngas and methanol (107), and phytoestrogens while transforming isoflavonoids (108). The genome of the ATCC 8486 strain has been sequenced and published (108). The primary interest in *E. limosum* for this work is its ability to use methanol as a carbon source (52). By introducing synthetic pathways, valuable products such as γ-aminobutyric acid and 1,4-butanediol could theoretically be produced from a low-cost feedstock. As methanol can be produced in a sustainable way, chemicals produced by introduction of synthetic pathways into *E. limosum* can also be produced sustainably. Of key importance, genetic modification techniques and associated processes (transformation protocols, plasmid replicons, selective markers) are all available for *E. limosum.*

E. limosum is also likely of interest to the healthcare industry due to its presence as part of the gut microbiome. In recent years, it has become clear that the gut microbiome of both animals and humans (109) is of great importance, and has been proposed as an instigator for many diseases, including depression and obesity (110). *E. limosum* has been found to be present in high quantities in the microbiomes of centenarians, suggesting a role in maintaining a long life in humans (111).

These factors combined resulted in *E. limosum* being selected as the organism used in this work. There are many strains of *E. limosum* available, and while ATCC 8486 was initially chosen for this work, NG-6849 was later adopted due to its superior phenotype.

1.4 Wood-Ljungdahl pathway

The Wood-Ljungdahl pathway enables acetogens to use CO or CO_2/H_2 as their sole sources of carbon and energy. Their ability to accept electrons from a wide array of donors and the ability to produce complex organic molecules from C1 compounds (103) gives acetogens an advantage over other organisms (75). This pathway likely evolved very early in the history of life and allowed life to survive in the less desirable conditions of early Earth (98,103). The Wood-Ljungdahl pathway has two main branches: a methyl branch and a carbonyl branch, also referred to as the eastern and western branches respectively (85), as summarised in Figure 1.

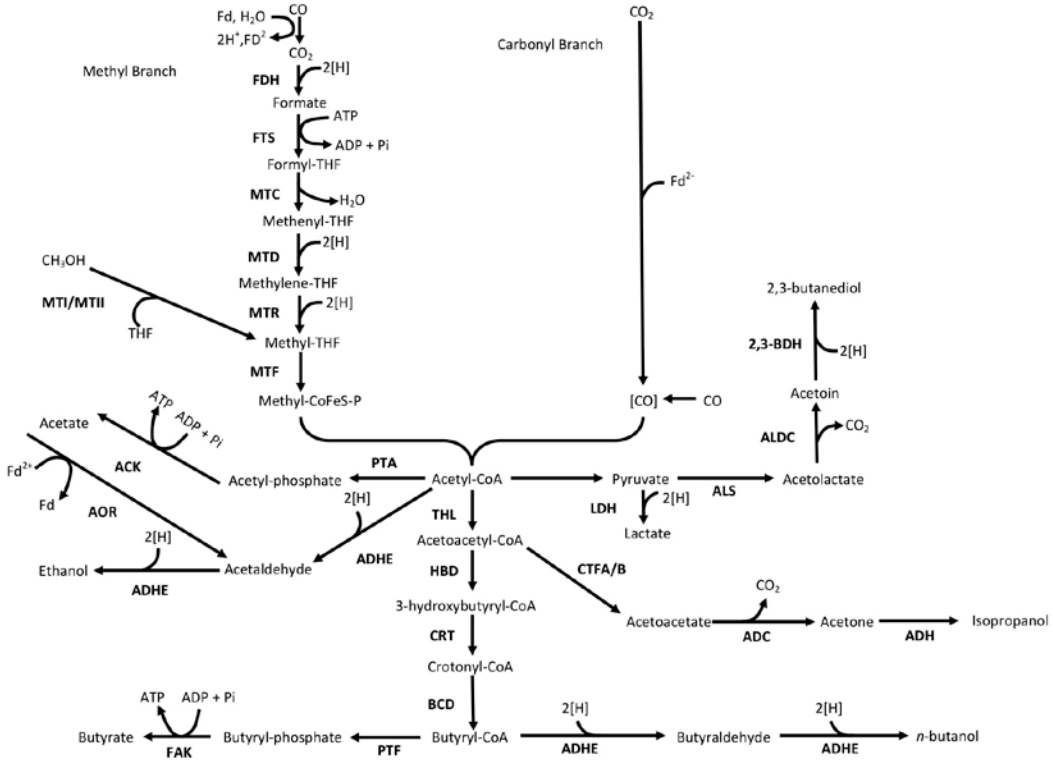

Figure 1 - The Wood-Ljungdahl pathway and representative downstream metabolic pathways in an acetogen

ACK, acetate kinase; ADC, acetone decarboxylase; ADHE, aldehyde/alcohol dehydrogenase; ALDC, acetolactate decarboxylase; ALS, acetolactate synthase; AOR, aldehyde:ferredoxin oxidoreductase; BCD, butyryl-CoA dehydrogogenase; CoFeS-P, corrinoid iron-sulphur protein; CRT, crotonase; CTFA/B, coenzyme A/B transferase; FAK, fatty acid kinase; Fd, oxidized ferredoxin; $Fd^{2+,}$ reduced ferredoxin; FDH, formate dehydrogenase; FTS, formyl-THF synthetase; HBD, 3-hydroxybutyryl-CoA dehydrogenase; LDH, lactate dehydrogenase; mti, methyltransferase I; mtiI, methyltransferase II; MTC, methenyl-THF cyclohydrolase; MTD, methylene-THF dehydrogenase; MTF, methyltransferase; MTR, methylene-THF reductase; PTA, phosphotransacetylase; PTF, phosphotransferase; THF: tetrahydrofolate; THL, thiolase, 2,3-BDH: 2,3-butanediol dehydrogenase; 2 [H], reducing equivalents (e.g. NADH or NADPH). Figure adapted from Humphreys & Minton (112).

In brief, the fixation of CO_2 starts with the reduction of one CO_2 molecule by six electrons to a methyl group in the methyl branch of the pathway. In the carbonyl branch, a CO_2 molecule is reduced to CO, which is used in combination with the methyl group from the methyl branch process to make acetyl-CoA. This acetyl-CoA is then either used to generate ATP and acetate, by conversion to acetyl phosphate and the transfer of the phosphoryl group to ADP, or is incorporated into cell carbon (85,113,114). This is the process that enables acetogens to fix CO_2 and is where the by-product for which they are named is produced. Interestingly, the Wood-Ljungdahl pathway can be used in two directions: one oxidative, using acetate as a carbon source, and the other reductive, for energy conservation and autotrophic carbon assimilation (85,114–116).

Regarding to the methyl branch of the pathway, genes are spread throughout the *M. thermoacetica* genome (98), with the genes encoding the N^{10}-formyltetrahydrofolate synthetase enzyme being located over 300 genes away from the *acs* gene cluster. The carbonyl branch is quite the opposite, however, with the genes for the pathway enzymes being co-localized in the *acs* gene cluster with genes organized by the functional sequentially of the enzymes; genes ordered in parallel with their respective enzymes located in the Wood-Ljungdahl pathway; the first gene in the cluster encodes the first enzyme in the carbonyl branch of the Wood-Ljungdahl pathway.

1.4.1 Use of methanol by acetogens

Methanol metabolism in acetogens occurs via the Wood-Ljungdahl pathway, and this metabolism has been best characterised in *M. thermoacetica* (85). For methanol usage to be possible in an acetogen, it must have the ability to produce the complex methanol:coenzyme M methyltransferase MtaABC. This 3-subunit complex is well-conserved throughout acetogens (98). Methanol is processed in both directions in the methyl branch of the Wood-Ljungdahl pathway, with three molecules of methanol proceeding in the forward direction, while one proceeds in the reverse direction. The use of methanol in the reverse may seem counterintuitive but by doing this an organism can produce a molecule of CO_2, which it can uptake via the carbonyl branch (85), and six reducing equivalents. These reducing equivalents are then used in the

carbonyl branch to drive conversion of CO_2 to CO. Across both pathways, this reaction results in the production of three acetate molecules for every four methanol molecules in the presence of an excess of CO_2.

If we assume an excess of CO_2, the following reaction occurs:

$$4CH_3OH + 2CO_2 + nADP + nP_i \rightarrow 3CH_3COOH + 2H_2O + nATP$$

Equation 1: Equation for the fixation of methanol and CO_2 to acetate
An empirical reaction formula displaying the overall reaction that occurs for autotrophic growth on methanol and CO_2 (117).

The conversion of methanol to acetate allows for the regeneration of ATP from ADP. This core reaction is essential to regenerate ATP while feeding on methanol and CO. If it is removed, it is improbable that a cell line would retain its ability to grow on methanol. This pathway has the distinct advantage of no carbon loss, as the above anaerobic process uses CO_2. Aerobic processes for metabolising methanol exist, such the serine cycle and ribulose monophosphate pathway (118,119). These aerobic processes have the distinct disadvantage of carbon loss via CO_2, which is due to the biochemistry of those pathways not fully fixing carbon into either biomass or metabolites (118,119). Attempts have been made to engineer additional pathways through the use of carboxylation reactions to try to remedy this in an aerobic process (120). When compared to the aerobic pathways, the Wood-Ljungdahl pathway offers a more promising way to process methanol as a carbon source and would theoretically result in greater product yields in an industrial application due to the lack of carbon loss.

In many cases, this metabolism ends with the production of acetate via the precursor acetyl-CoA (74). As stated previously, some acetogens, either natively or via metabolic engineering, produce products other than acetate (74) such as ethanol. The ethanol production route having an intermediate of acetaldehyde (74), see Figure 1. The downside of this metabolic system is primarily the low amount of excess energy, in the form of free ATP, as compared with aerobic bacteria (121). This energy

limitation severely limits the viability of fermentation towards more complex compounds.

1.5 The Genome and its modification

In the mid-1990s, the first bacterial genome was sequenced (122) and in the following years genome sequencing technology has advanced and become much cheaper (123). With genome sequencing becoming more affordable, it can be used by a wider array of researchers in many more countries and done more routinely. Methods that were not viable or common due to the cost of genome sequencing are now routine because of decreased costs. For example, sequencing plasmids to ensure inserts have been inserted faithfully and in the correct position, or as the last stage in confirming a genomic mutant. Having a complete annotated genome sequence is essential for directed genetic development and so is an important foundation, and it can be a crucial step in developing an organism into an industrial strain. The genome sequence will be used throughout the development of a strain, from identifying potential targets to modify to the creation of genome-scale models, which are a useful tool for genetic engineering (124–126).

1.5.1 Sequencing technology

The first step in obtaining a completely annotated genome is nucleotide sequencing. High-throughput “next-generation sequencing” has made complete genome sequencing cheap enough to be practical. This is primarily done through the parallel processing of many thousands of nucleotide sequences, with older techniques such as the dye-based Sanger sequencing being performed in a serial manner, which is much slower (127–129). The first generation of high-throughput sequencing technologies, such as Illumina MiSeq, generated quite short average reads of between 50 and 700 base pairs (130,131).

Due to the parallel sequencing technique, many incomplete reads are generated, and these must be threaded together into the full sequence using appropriate software (132). If the organism has been previously characterised, templates can be used to

map the sequences to approximate positions within the genome. For uncharacterised organisms, however, overlapping reads must be used (133,134). Issues arise, primarily in prokaryotes, where there are the large regions of very similar sequences or repeat sequences, which are often much longer than the read length, and can make it difficult to determine full length and position (130,131,135). Software algorithms can somewhat deal with this and can be used in *de novo* assembly (134,136). Sequencing technologies with longer read lengths can compensate in such circumstances, with Oxford Nanopore MinION able to generate read lengths of up to 150,000 bases long (137).

1.5.2 Genome annotation

After a full-length read is generated, the next step is to annotate the genome. The individual genes must be labelled and annotated as the raw nucleotide sequence alone is of little use. Gene annotation is performed by looking for coding sequences and genomic features. The gene function is suggested based on similarity with existing known protein sequences (138,139). This process is automated and done by the use of software (140,141). This software works by taking the DNA sequence, translating it into a protein sequence and matching it against a database of known proteins to elucidate function.

Genome annotation using this approach has several issues. Any errors in sequencing will be compounded, causing issues in annotation, and so are cumulative with inherent inaccuracies of the process (142). This is a wide reaching problem that affects the whole research community (143). Without performing multiple full genome sequences with both high-accuracy short-read and lower-accuracy long-read technologies and comparing them extensively (which is very expensive and time consuming) it can be difficult to initially identify errors in sequencing, especially in a previously uncharacterised organism (144). This means that many erroneous genomes may be published in the literature. This is an issue, as these erroneous genomes enter databases that are referenced for use in automated annotation. Erroneous genome sequencing causes erroneous annotation, which itself causes yet more errors due to it being referenced. This affects databases like UniProt, which compile data from many

sources and are often used as a resource by researchers (145). The only current way of dealing with this issue is manual curation which is limited due to the small number of qualified individuals relative to the work load, the slow nature of manual curation and likely the lack of interest in performing such a task.

1.5.3 Genetic modification of acetogens

Although some acetogens naturally produce products other than acetate (146), most require genetic modification with additional metabolic pathways to allow for the production of other chemicals (74). With a significant investment of time, organisms could be modified to produce higher carbon alcohols such as 1-pentanol (147), which are useful as fuels, solvents, and precursors (148–150), or other valuable chemicals such as acetone.

Genetic modification is the process by which the genes of an organism are modified. Generally, this is done in a targeted fashion with a set of target genes modified to alter the phenotype either for study or to engineer a cell for product formation (151,152)(152)(152)(152)(152)(152)(152). Initially, genetic modification was very primitive, mostly random, and limited to UV or chemical mutagens (153,154). Technology has improved dramatically from the early days, and full genome sequencing is now cheap and rapid (155). When combined with new genetic modification tools (151), this has allowed for significant progress to be made in the field of genetic engineering (151). Many genetic modification technologies have been created and have proved successful as a means of generating stable mutants in bacterial species (156,157).

Genetic modification of acetogens has been used to allow for the production of non-native products. *Clostridium aceticum* has been genetically modified to introduce a novel pathway using a thioesterase enzyme from *Bacillus subtilis* and an acyl-CoA thioesterase from *Haemophilus influrenzae* to produce acetone (74). *Clostridium ljungdahlii* has been genetically modified by the introduction of genes from the *C. acetobutylicum* encoding the butanol synthesis pathway (158). Much of the genetic modification undertaken in acetogens (especially in *Clostridium*) has been

accomplished using allele exchange (92,159) and ClosTron technology (160–162). However, in recent years, more advanced Counter-selection-based techniques such as CRISPR and toxin-antitoxin-based systems have come to the fore, offering dependable, scarless, in-frame genetic modification with a high degree of efficiency and specificity.

1.5.3.1 ClosTron

ClosTron is a genetic modification system developed for *Clostridium* bacteria (160,162). It enables directed mutagenesis in *Clostridium* by making use of mobile genetic elements known as group II introns (156). ClosTron is much faster than previous methods, with mutants being created in under two weeks and resulting in stable mutations. The system represents a significant improvement on previously used systems. In fact, in the original ClosTron research paper demonstrating this technology had more mutants of *C. acetobutylicum* and *C. difficile* than had been made in all previous studies combined (159,160).

The system works by exploiting type I and II introns to ensure that only cells that have successfully undergone mutation integrate an antibiotic resistance gene. This allows for a selection force, the addition of antibiotics to the growth media, to eliminate non-mutant cell lines. The system is carried on a pMTL80000 series plasmid into the cell. This plasmid contains a type I intron which acts to silence the antibiotic resistance gene *ermB*; *ermB* confers antibiotic resistance to erythromycin. Once the type II intron, the LI.LtrB intron of *Lactococcus lactis*, has integrated, *ermB* becomes active, conferring resistance to cells and allowing for rapid identification of mutants.

ClosTron has some downsides. Firstly, the antibiotic resistance conferred by ClosTron requires additional steps to remove (163). If markers are not removed, additional genetic modifications require a different antibiotic to act as a selection marker, necessitating growth medium with multiple antibiotics, a problem that only increases as more modifications are done. Secondly, there is a possibility for genes downstream of the genomic insertion site to suffer from polar effects as a result of insertions into

operonic regions (161). In addition, ClosTron is far less specific than newer genetic modification tools and can insert into other locations, sometimes in multiple locations, requiring significant work (such as Southern-blot analysis) to confirm that the appropriate insertion occurred.

By using alternative counter-selection systems and antibiotic resistance genes carried only on a plasmid, which is later lost, it is possible to create clean in-frame deletions without requiring the insertion of antibiotic resistance genes into the genome. This is how many newer genetic modification techniques function and in so doing render older genetic modification techniques such as ClosTron obsolescent. Due to the availability of these newer techniques ClosTron was not used as part of this work.

1.5.3.2 Allele exchange (AE) and Allele-coupled exchange (ACE)

Mutants can be created though homologous recombination. This DNA repair mechanism is exploited to allow for recombination events between the genomic wild-type allele and mutant allele templates on supplied plasmids. These plasmids contain homologous DNA that is shared between the mutant and wild-type allele flanking the region of mutation, which are referred to in this study as homology arms (164–166). In the case of the creating of in-frame deletions, homology arms are designed with the central region of the gene missing, containing only the start and stop codons to ensure an in-frame deletion. For the inserting material into the chromosome, homology arms are created in the inverse fashion, with the payload being placed between the flanking sequences. If designed properly, these mutations can be considered “clean” as they are in-frame, do not leave residual selective markers in the genome, and should have no polar effects. This provides a significant advantage over more primitive systems like that of ClosTron (164).

Generating of mutants using AE relies on naturally existing but infrequent processes. Therefore, most transformed cells within a population will not contain the desired mutation, requiring significant screening for mutants. Counter-selection can provide a means to improve effectual efficiency. This does not increase the number of cells which undergo allelic exchange, but rather killing cells that do not.

Allele-coupled exchange (ACE) is a form of allelic exchange in which the second recombination event is used to cause a selectable phenotype to eliminate cells from the population that have not undergone double crossover (Figure 2), with asymmetrical homology arms ensuring the order of recombination events (164,165). ACE occurs in two steps. First, antibiotic selection eliminates cells that do not contain or properly maintain the ACE plasmid. Resistance is conferred by the initial recombination event resulting in the integration of a resistance marker into the genome (164). The second step relies on counter-selectable chromosomal gene targets, such as *pyrE*. Such a gene should be at the end of its operon so that an in-frame insertion should not affect its operon or the following operon or create polar effects. Mutant variants of *pyrE* confer fluorinated nucleotide 5-fluoroorotic acid (FOA) resistance and make the host an uracil auxotroph, while the wild-type is FOA sensitive and a uracil prototroph (159). By cycling between these two phenotypes, one can “knock-in” and “knock-out” genes (159).

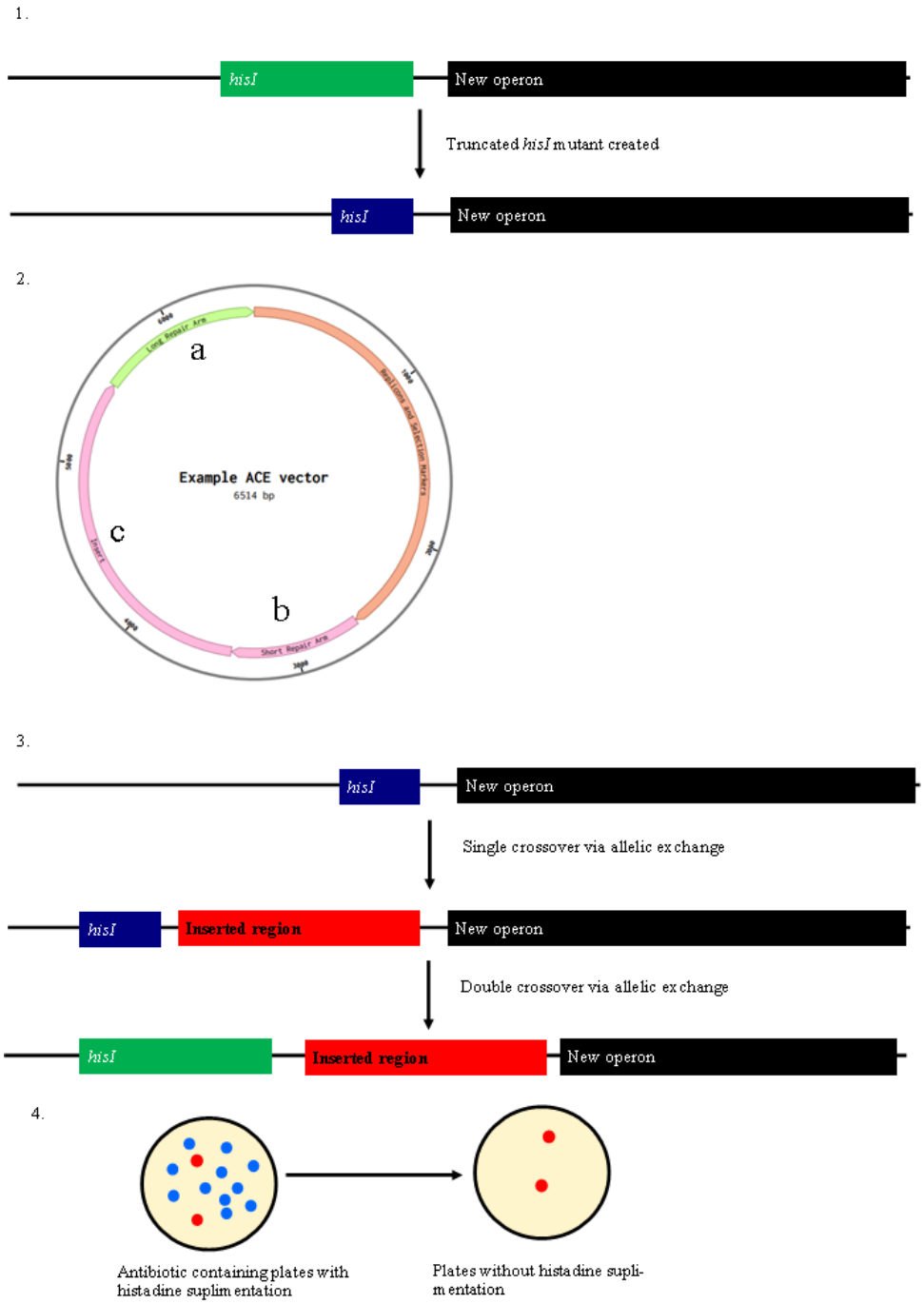

Figure 2 - Allele-coupled exchange in an auxotrophic background

1. *ACE depends on a counter-selectable phenotype, in this example an auxotrophic mutant with a truncated hisI gene that requires medium supplementation with histidine for growth.*
2. *An ACE plasmid is introduced into the mutant with the goal of insertion of an operon.*
   a. *A long homology arm dictates where the first cross over event takes place.*
   b. *A short homology arm ensures repair of hisI through the second recombination event; this ensures cells which regain hisI have also integrated the insert region.*
   c. *The insert region is self-contained with its own promoters and terminators.*
3. *Recombination occurs, first inserting the cargo (2 c.) then a double crossover occurs repairing the hisI gene and eliminating the vector backbone.*
4. *The plasmid is initially selected on the basis of antibiotic resistance, followed by ability to grow without histidine selecting cells in which double crossovers has occurred. Blue circles represent cells maintaining the plasmid which have not undergone recombination, the red representing those which undergo recombination.*

ACE has been used in acetogens of the genus *Clostridium* as well as in *E. limosum* (117,159). While ACE can work with any counter-selectable phenotype, in the case of the prior work performed in *E. limosum*, the dual phenotypes of the wild-type and mutant *pyrE* alleles were exploited. This prior work relied on a mutant strain with a truncated *pyrE*, so a mutant strain needs to be created. This can be created using AE (159). It also relied upon the pCB102 replication-deficient vector, which only a minority of cells can fully integrate, while the remainder maintain the free plasmid and are limited by its replication speed. The vector carried a cassette which not only repaired the *pyrE* gene but also inserted the lac-tcdr gene from *C. difficile* (117). Cells which have integrated the plasmid gain a growth advantage and could be identified as larger colonies on plates (159).

The mutant *pyrE*, having a distinct phenotype compared with the wild-type *pyrE*, enables the selection of cells that have successfully repaired the gene. Mutants cannot synthesise uracil, so growth on uracil-deficient medium will select for those cells that have repaired the *pyrE* gene and presumably integrated the gene carried on the cassette. This can be tested for using colony PCR and appropriate primers (117,159). To ensure the initial set of cells had a mutant *pyrE*, cells could be cultured on medium with fluoroorotic acid, which would kill wild-type cells (159).

The advantage of this system is its relative ease of use and the inability to get false-positive results. The *pyrE* gene cannot be repaired by any mechanism other than ACE, and those cells that have not had their *pyrE* repaired cannot grow on uracil-deficient medium (159). The disadvantage is the additional time needed over standard plasmid transformations due to the requirement to grow on an additional set of medium plates, uracil-deficient, after antibiotic selection on rich medium plates (159). This requires the availability of a defined minimal medium, which is not always available.

1.5.3.3 CRISPR

Homologous recombination is the natural process that allelic exchange relies upon (Figure 3). It is exploited as a means of genetic modification in many systems including CRISPR. Of note CRISPR systems for use in organisms capable of performing NHEJ function differently (167). As discussed previously, homologous recombination is infrequent, and most of a transformed population will not be mutants (159). CRISPR-Cas systems act as a counter-selection, in which guideRNA-mediated targeting of the wild-type allele selectively kills cells that have not undergone homologous recombination due to genomic damage caused by the Cas. By causing damage to the genome, organisms must repair in response to survive. Allelic exchange is used to direct this repair and so causes the desired mutation.

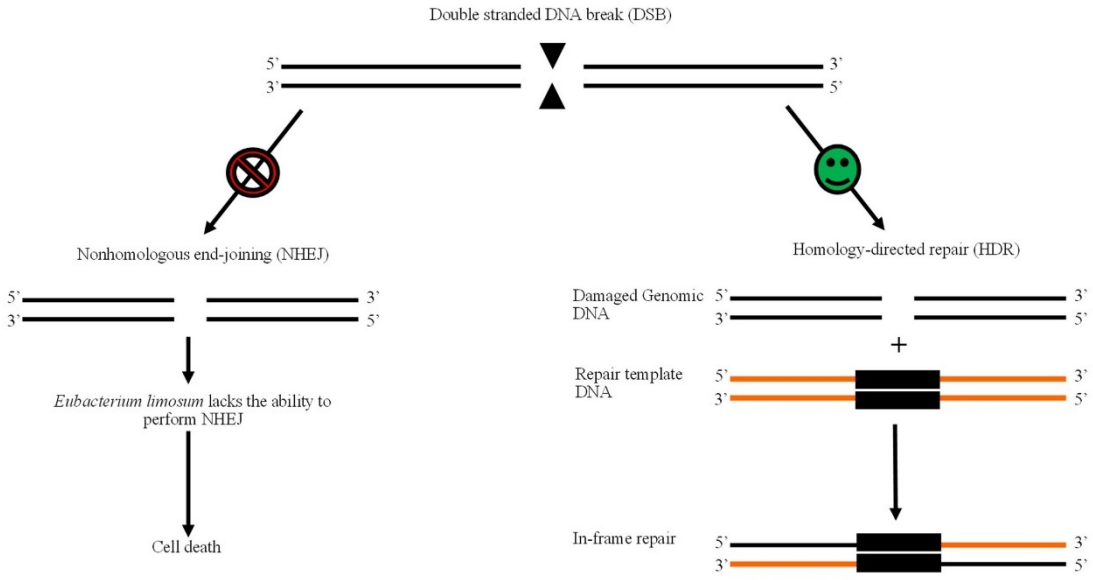

Figure 3 - Nonhomologous end-joining (NHEJ) and homology-directed repair (HDR) in E. limosum (based on a figure from Hsu et al. (2014)(167)***)***

E. limosum lacks mechanisms for nonhomologous end-joining making double-stranded breaks lethal, therefore homology-directed repair is the only way to modify the genome without killing the cell.

Bacterial cells are under threat from bacteriophages, viruses that target prokaryotic organisms (168–172). Some bacteria have evolved an adaptive immune system known as the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated protein) system. These systems consist of a *cas* operon (which varies in composition depending on the type of CRISPR system) and associated CRISPR arrays. CRISPR arrays consist of direct repeats with spacers between them.

These spacers are typically derived from the genomes of invading phages. Spacers are processed into CRISPR RNA (crRNA) molecules, which allow for the targeting of the viral genetic material and its degradation by the nuclease Cas (168–172). Most Cas proteins require a protospacer adjacent motif (PAM), which is a short region of DNA that is close to the region that Cas proteins will cleave. The directionality and relative position of the PAM will vary for differing Cas systems and proteins (168–172). Different Cas proteins have different characteristics, and so, different advantages and disadvantages regarding their exploitation as part of molecular tools. CRISPR-Cas systems are divided into class and type. This class/type system consists of two classes and six types (173) .

Class 1 systems, which make up the majority of CRISPR systems in archaea and bacteria (174,175), consist of multiple effector proteins that function as a complex. These systems encompass different types, including type I, type III, and type IV systems (176,177). Type I systems, characterised by the presence of Cas3 (178), exhibit variations in cascade proteins across different subtypes. They exclusively target DNA (179,180). In contrast, type III systems contain Cas10 (181–183) and possess the ability to target both DNA and RNA for cleavage. Type IV systems employ smaller effector molecules and utilise crRNA molecules as guides for the effector complex (184). It is worth noting that type IV systems have relatively less published material compared to type I and type III systems within the class 1 category.

Class 2 systems, characterised by a single effector protein, are extensively described in the literature primarily due to the widespread use of Cas9. However, they represent the minority of CRISPR systems found in nature. Unlike class 1, most class 2 systems cannot target both DNA and RNA, with the specific system type determining the target specificity. Class 2 encompasses type II, type V, and type VI systems (176,177). Type II systems employ Cas9 as the effector molecule and are widely recognized as highly characterised systems, commonly utilised for genetic modification purposes (167,172,185–189). In addition to Cas9, ancillary proteins such as Cas1 and Cas2 play a role in spacer insertion, while Cas4 is implicated in

regulation alongside its potential involvement in spacer insertion (190). tracrRNA acts as a guide for type II systems.

Type V systems include CRISPR-Cas14 and CRISPR-Cas12. CRISPR-Cas12 (formerly known as Cpf1) is a class 2 type V system where Cas12 binds and cleaves DNA with tracrRNA as a guide (191). Although CRISPR-Cas12 has not been extensively used for genetic modification, it is of great interest for disease diagnostics. During the COVID-19 pandemic, a CRISPR-Cas12 system were developed to detect the presence of the SARS-CoV-2 virus (202). The effective detection of specific viruses may be the primary role for which CRISPR-Cas12 is used in the future. CRISPR-Cas14 systems make use of Cas14, smallest Cas effectors, with gene sizes varying from 400 to 700 bases. Cas14s have their origins in archaea (192). Although Cas14 may seem like a prime candidate for usage in genetic modification systems due its small size resulting in smaller plasmids and thus higher efficiency (193) in a plasmid-based system, its targeting of only single-stranded DNA precludes its use as a genetic modification tool. Like Cas12, Cas14 is more useful in diagnostics than in genetic modification, as it does not require a PAM sequence and is more specific in its recognition of ssDNA than Cas12 or Cas13 are for dsDNA and RNA respectively (192).

Type VI systems are the only class 2 system that can cleave RNA, and they make use of Cas13 as their effector protein (194,195). CRISPR-Cas13 based systems are class 2 type VI (194–196). Their main research use is targeting RNA for RNA knockdown (196), a process where a specific RNA is decreased in number within a cell, preventing gene expression, which has a useful role in phenotype analysis. Such systems offer significant advantages over traditional RNA interference due to their much higher specificity and decreased off-target activity (196). Although not useful for DNA genetic modification, a Cas13 tool would be very useful in phenotype analysis. Unfortunately, such a tool has not been demonstrated in acetogens, and while CRISPR-Cas13 use has been demonstrated for use in mammalian cells (196), it is likely that significant work would be needed to allow for use in acetogenic bacteria.

CRISPR-Cas systems can be endogenous to an organism (180,197) or may be introduced, along with the associated RNA-guide, on a vector (185,186,198,199), generally a plasmid (174). There have also been suggestions of using a phage as a vector, especially when targeting pathogenic bacteria (175)

Plasmids used for CRISPR-Cas mediated genetic modification are made of multiple parts. For non-endogenous systems the Cas protein would be present on plasmids in addition to the guide RNA required to direct the Cas. Plasmids also need to display successful transformation into the cell, this is generally done using antibiotic resistance. In general sequences to ensure replication in the appropriate cell types are also present. Some vectors lack the ability to replicate and are known as “suicide” vectors. Suicide vectors have been used to ensure that Cas9 activity is temporary, limiting the risk of off-site activity (200) and aiding in plasmid loss after the desired modification has occurred. If used in prokaryotes which lack the ability to perform non-homologous end joining (201), sequences to allow for DNA repair must be included. Homology arms are DNA sequences up and downstream of the target area and act as a template to guide homologous recombination (167,202). To perform a knock-in, sequences are inserted between homology arms to ensure correct matching and repair of the double strand break (203).

Plasmid design will result in either a constitutive or inducible CRISPR-Cas system. Constitutive systems are easier to design and do not require the development of a system that is highly active when induced and highly repressed when not. However, they do have drawbacks. If Cas expression is too high, it can cause issues in plasmid assembly in *E. coli*, either killing the cells outright or selecting for cells that have in some way mutated the Cas protein (204) as CRISPR-Cas systems are a stress upon cells (187). If said system is constitutive, it is always active, and so may be active at an inopportune time, such as at a time of high cellular stress. Cells typically have significant stress as a result of transformation itself and the initial growth in the presence of antibiotics (205,206). Multiple stresses may result in decreased efficiency and decreased numbers of cells surviving. This is especially an issue if a CRISPR-Cas protocol is poorly optimised or results in low efficiency, with this being most noticeable in mutations that confer a disadvantage.

Inducible systems have many advantages primarily that Cas activity is not activated continually. Further by having a tight system of repression and expression, Cas activity can be varied throughout an experiment, which may be of great utility in CRISPR-interference (CRISPRi). An inducible CRISPR system known as RiboCas has been developed and has proven to be successful in many organisms, primarily the anaerobic *Clostridium* (185). This system makes use of theophylline-dependent riboswitches to control the expression of the *Streptococcus pyogenes cas9*, resulting in a wide dynamic range and tight repression, especially when compared to constitutive systems.

This system had not been demonstrated in *E. limosum*. However, the fact that normal constitutive CRISPR has worked in *E. limosum* (117) indicated that the barriers to this system would most likely not be fundamental. Barriers may include the specifics of the transformation and induction protocol, the concentration of the inducer, the riboswitch used, and the specifics of the plasmid construction. The fact that this system has worked in various *Clostridia* (anaerobes) and *Parageobacillus thermoglucosidasius* (an aerobe) (185) indicates that this system should work across a diverse range of organisms.

1.5.3.3.1 CRISPR-Cas9

CRISPR-Cas9 is a class 2 type II system, with a single effector protein, Cas9, acting to perform both guide and target binding in addition to cleavage (190). The Cas9 protein, adapted from a bacterial type II CRISPR immune system (168,207), has been extensively used as a tool for genetic modification (207). It has two nuclease domains, which when targeted to a sequence by an RNA-guide, cause a double-strand break (208,209). Without the presence of supplemented template DNA, to allow for homologous recombination to occur, double-stranded DNA breaks are often lethal due to many bacteria, including *E. limosum,* being unable to perform non-homologues end-joining (201). CRISPR-Cas9 acts as a form of counter-selection, with cells that receive a double-stranded break, but do not undergo allelic exchange as part of the homologous recombination to repair the double-stranded break, dying. This selects for the mutant genotype and kills the wild type of genotype.

A nickase-Cas9 (nCas9) is a Cas9 protein that has had one of its two nuclease domains inactivated (188,210). This can be used for more specific editing of genes. It is possible for RNA-guides to target off-site (i.e., as in not the target gene). Using an nCas9 requires a pair of nCas9 to bind before a double-strand break occurs, increasing on-target activity (211,212).

A nuclease-deactivated Cas9 (‘dead’ Cas9 or dCas9) has had both of its nuclease domains inactivated and so can be used to affect expression levels of genes, inhibit a gene’s transcription (213,214), or in the case of both bacterial (213) and mammalian cells (171,188,207,215,216), nuclease-deactivated Cas9 has been fused with transcription factors to allow for increased expression of a gene. This makes this variant of Cas9 useful in CRISPRi.

CRISPR-Cas9 is well reported and has been used in a wide range of organisms (199,217,218). A CRISPR-Cas9 system has also been developed for *E. limosum* and has been demonstrated (117). As CRISPR-Cas9 is not native to *E. limosum*, it must be introduced and expressed from a plasmid containing Cas9, a suitable guide to direct Cas9, and a pair of homology arms to allow for homologous recombination and allelic exchange (Figure 4). The existence of a pre-existing system that has been developed for *E. limosum* made CRISPR-Cas9 the obvious initial choice for the project.

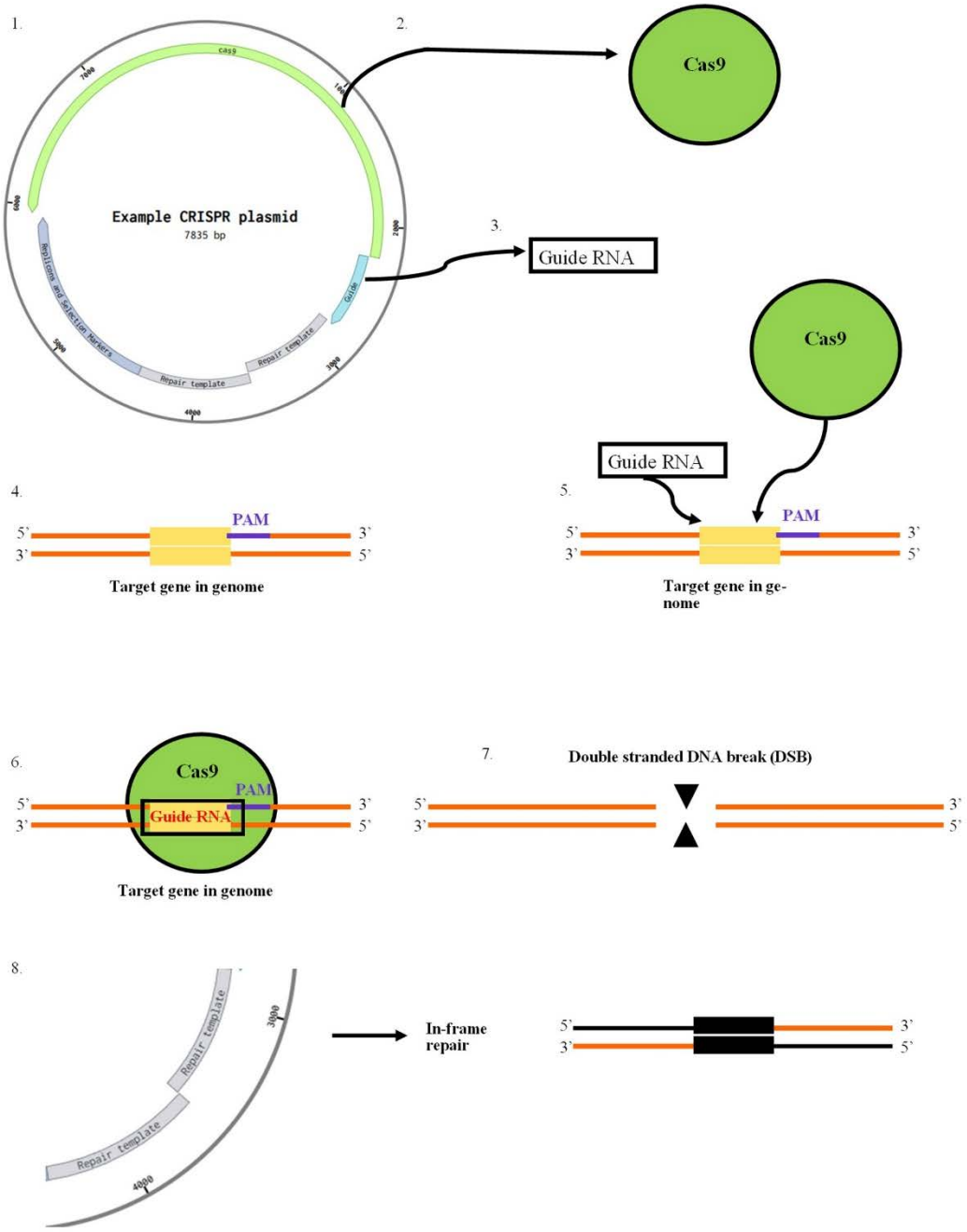

Figure 4 - Non-native CRISPR-Cas9 functioning off a plasmid (based on a figure from Hsu et al. (167)***)***

1. *A plasmid containing Cas9, guides to direct Cas9 and repair template sequences is introduced into the cell.*
2. *Cas9 is expressed from the plasmid.*
3. *Guide is expressed from the plasmid.*
4. *The target region for modification requires the presence of a PAM.*
5. *Guide directs Cas9 to site preceding PAM.*
6. *Cas9 binds*
7. *Cas9 forms a double-stranded break at site.*
8. *Repair template sequences on the plasmid act to allow homologous recombination and allow the repair of the double-stranded break using allelic exchange, introducing the mutation.*

1.5.3.3.2 CRISPR-Cas3

CRISPR-Cas3 is a class 1 type I system that uses multiple effector proteins acting as a complex. CRISPR-Cas3 makes use of a different style of guides as compared with CRISPR-Cas9. Instead of a single shorter guide, it makes use of several guides that are enclosed by repeat sequences. *C. difficile* has been successfully edited multiple times using CRISPR-Cas3 and phage-delivered crRNA. This crRNA was used to direct the endogenous Type I CRISPR system of *C. difficile* (219).

E. limosum has an endogenous CRISPR-Cas3 system, which was harnessed by a fellow researcher (Dr. James Millard, University of Nottingham, personal communication) towards the end of this study (Figure 5). It offered many advantages over a traditional CRISPR-Cas9 system. Firstly, vectors could be made smaller as they would no longer require the ~4 kilobases of Cas9 to be carried. This could either allow multiplexing with plasmids of similar size to the CRISPR-Cas9 plasmids or could allow for much higher transformation efficiency with a smaller plasmid (193). Secondly, the endogenous nature of the system removes the requirement for promoter optimisation for the Cas protein's production (as is needed for Cas9 carried on a vector). Thirdly, there is no concern about mutation of the Cas in the steps leading up to transformation, for instance, while being maintained in *E. coli*. Finally, the endogenous CRISPR-Cas3 system should be more efficient than the currently existing CRISPR-Cas9 system using in *E. limosum*, which had an efficiency of ~10%. The CRISPR-Cas3 system developed made use of tetracycline-controlled transcriptional (TET) activation (220,221), specifically a tetracycline on system (222) referred to in this study as TET, with transcription being turned on in the presence of anhydrotetracycline (Figure 5). This granted inducible control over the system, a feature that was desirable. While the CRISPR-Cas3 system is endogenous to *E. limosum*, plasmids must be introduced to use said system as a tool to cause genetic modification. Plasmids were of a similar makeup to those of CRISPR-Cas9 but lacked Cas9 and used TET to control guide expression.

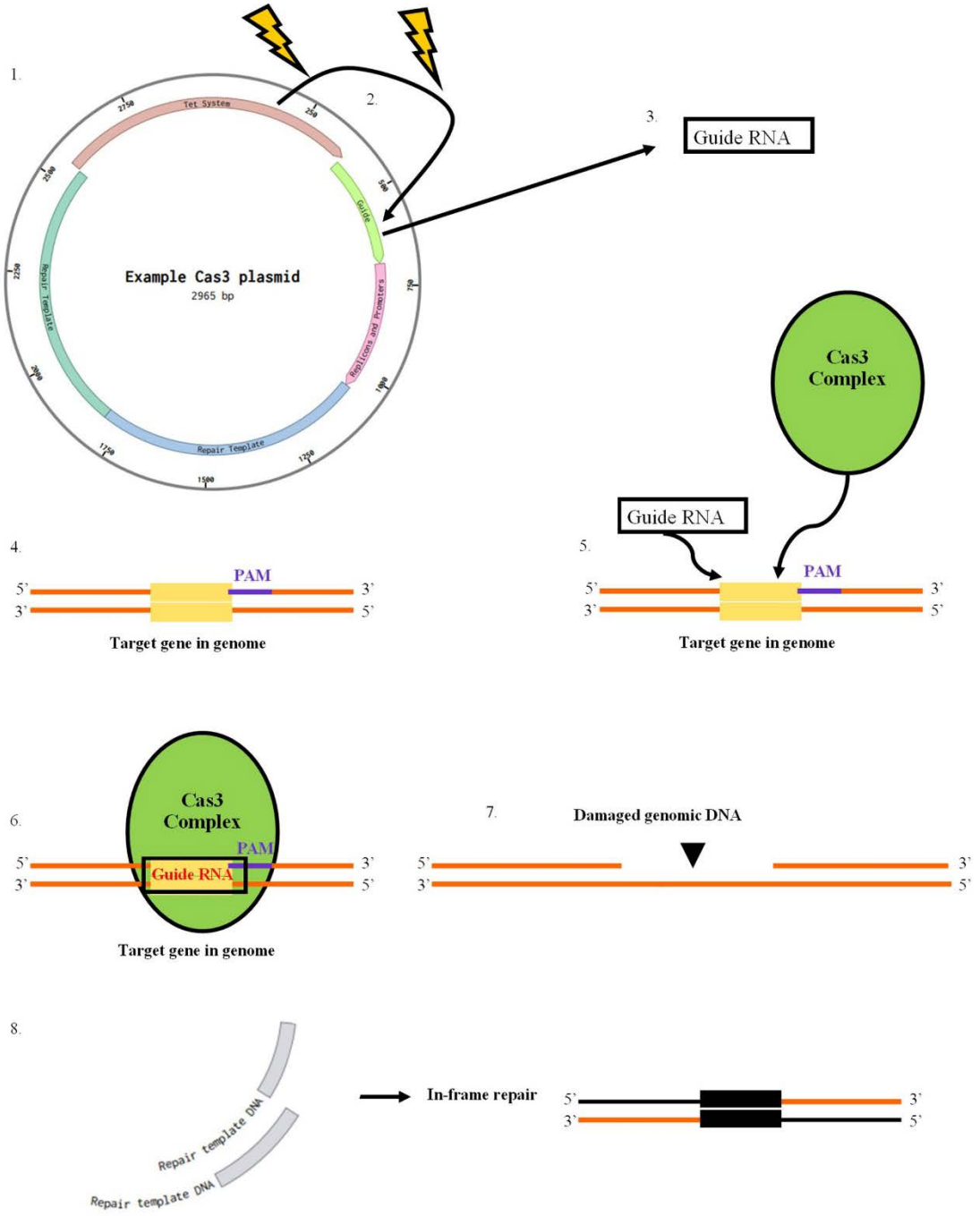

Figure 5 - Native CRISPR-Cas3 functioning with guides introduced on a plasmid (based on a figure from Hsu et al. (167)***)***

1. *A plasmid containing guides to direct Cas3, and repair template sequences is introduced into the cell.*
2. *Anhydrotetracycline is added, via supplementation of medium.*
3. *Guide is expressed from the plasmid.*
4. *The target region for modification requires the presence of a PAM.*
5. *Guide directs the Cas3 complex to site proceeding PAM.*
6. *The Cas3 complex binds*
7. *Cas3 causes DNA damage at site.*
8. *Repair template sequences on the plasmid act to allow homologous recombination and allow the repair of the single stranded DNA damage using allelic exchange, introducing the mutation.*

1.5.3.4 Transposon-Directed Insertion-Site Sequencing

There are two primary approaches to genetics: forward and reverse. Forward genetics is so termed because of the direction of inquiry, with the genetic cause of a particular phenotype being of interest. Reverse genetics is the inverse, with the direction of inquiry centring around what the phenotypical effect of a particular genotype is. The previously discussed genetic modification techniques using ClosTron, ACE, and CRISPR are examples of reverse genetics, wherein mutations are created in a directed fashion. Forward genetics works quite differently, instead relying upon random mutagenesis. Random mutagenesis can primarily be achieved in two ways, chemically or using mobile genetic elements termed transposons.

Chemical mutagenesis relies on mutagens, chemical substances, or physical effects (e.g., ionising radiation) that cause permanent changes to genetic material (223). Chemical mutagenesis as a means of experimentation predates directed mutagenesis, and while it has the distinct advantage of generating a wide array of phenotypes with relative ease, it also typically induces multiple mutations per genome, many of which may be unrelated to the observed phenotype (223).

Transposons are mobile genetic elements that cause random insertional mutations within the genome (224). Transposon-directed insertion-site sequencing, or TraDIS (152), is a technique where a large pool of insertional mutants is created. This pool of insertional mutants then undergoes high-throughput genomic sequencing. This can be used to identify essential genes and to generate and understand novel phenotypes. In the case of the generation of novel phenotypes, an insertional mutant library can have colonies isolated and examined. Placing a selective pressure upon a library of mutants can allow for those mutants that grow best under those conditions to outcompete others and survive. Cell lines can be screened, and the genetic change that caused that phenotype identified. In the case of identification of gene essentiality, an insertional library with a high degree of coverage is required. Transposon insertions within essential genes, or the essential part of an essential gene, can result in cell death. Insertion mutants for non-essential genes will survive, and so essential genes can be identified by looking for a lack of mutants in a gene; if millions of unique mutants

arise overall but none arise in some genes, then said genes are most likely essential. Insertional frequency, defined as the number of transposon insertions per base in the genome, varies based on the site-specificity of the transposon (225,226). A higher frequency is preferable to maximise coverage. For this technique to be useful in determining gene essentiality, high coverage with many unique insertional mutants is needed. If the insertion count is too low, it is difficult to differentiate between gaps in coverage due to the poor insertion count or due to the essentiality of a gene or residue.

Very high insertional mutant counts have been achieved, with over 3 million unique mutants being generated in *Escherichia coli*, with a unique insertional mutant generated approximately every 5 bases (138). If high coverage were achieved in *E. limosum*, it may allow for the identification of all essential genes. With these identified, metabolic engineering can be performed with a better baseline understanding of what is feasible, in terms of genetic modification, in the organism, preventing wasted time pursuing non-viable engineering strategies.

1.5.3.5 Type I toxin-antitoxin system

A Type I toxin-antitoxin system was developed as a counter-selection marker for making AE mutants in *E. limosum* (227). The development of this system was stimulated by previous studies in *C. difficile* (228). Toxin-antitoxin systems are widespread among bacteria in so-called “modules”, which comprise of two genes, a stable toxin gene and an unstable antitoxin gene (205,228–230). Toxin genes encode for a protein while antitoxin genes can encode for a protein or RNA, depending on the type of toxin-antitoxin module. When the toxin gene is overexpressed beyond equilibrium with the antitoxin, either cell death or an inability for cellular reproduction to occur is observed.

Natively, toxin-antitoxin systems have several functions. They can act to regulate cellular metabolism by slowing it down or putting cells into a dormant state as a stress response (231). They can also act as an important part of phage defence (232,233) and can be important in stabilization of plasmids within cellular populations due to an addiction-like effect (234–236). In addition, they are important to pathogenesis in

pathogenic bacteria (237,238). It is likely that some toxin-antitoxin systems developed as organisms evolved into the more specialist phage defence systems known as CRISPR (239).

The ability of the toxin-antitoxin module to significantly affect host cells is key to their use as a genetic modification system. The *E. limosum* plasmid-based system developed (227) made use of a toxin-antitoxin pair from *Eubacterium callanderi* strain KIST612, specifically the *relE/pare* toxin and the *relB* antitoxin. The plasmid used was a suicide plasmid, which was done to ensure rapid plasmid loss upon confirmation of mutants. The expression of the system was done using the same TET system as used for CRISPR-Cas3 (1.5.3.3.2).

1.5.4 Genome-scale models

Genome-scale models are in silico representation of cells, with genes linked to produced proteins, and proteins linked to the metabolic reactions and pathways that occur in the living cell (124,240). Genome-scale models are often crude at first and refined over time, with gap-filling reactions added to allow the model to function where essential reactions are unknown (241–244). If, however, the full genome has been sequenced and tentative functions assigned to those genes it can be possible to identify missing reactions (245). With the use of a genome-scale model, reactions can be inserted into the model to test their viability before any lab work is undertaken, ensuring time and money are not wasted on fruitless experiments (246).

To make the best use of powerful genetic modification techniques, genomic targets must be selected. An accurate genome-scale model can be very useful in this endeavour. Genome-scale models have been used to aid in the understanding of the genomic-to-metabolomic relationship in many organisms (247). An accurate genome-scale model could have the ability to accurately predict metabolic changes (248,249). For instance, if 1-pentanol was the desired product, metabolic pathways could be inserted into the model to see if they would work in terms of sufficient free ATP, carbon, etc. If pathways worked well in the genome-scale model the genes required

for that pathway could be inserted into the organism, and, if the genome-scale model is accurate enough, the expected phenotype is likely to be observed.

As previously stated, the literature concerning *E. limosum* genetics is quite limited. By using a genome-scale model, genes of potential interest might be identified and relevant deletions/knockouts/knockdowns (as appropriate) performed (250). Ideally, should the modifications be successful, mutant phenotypes can be characterised, potentially shedding light on the function of some genes and so feeding data back into the genome-scale model.

1.6 Fermentation

To make use of acetogens on an industrial scale, large-scale fermentation must be achieved (40,48,55,82,251). This requires an organism that can tolerate such growth conditions and produce a useful product. In addition, a growth medium that will support the organism and drive production of the desired product that well-defined and commercially viable is needed. Finally a fermentation process that allows for good cell growth and maximised product yields is needed, ideally one that also allows for the extraction of the product in-situ (40,48,55,82,251). This presents many challenges, some of which are outside the scope of this research, such as the scale-up of fermentation and the development of downstream product processing methods. This project is concerned with strain development, if not achieving a strain ready for commercialisation at least getting *E. limosum* closer to that goal.

As *E. limosum* is an obligate anaerobe, it is killed by oxygen (108,252–254). Fermentation and growth in laboratory conditions must therefore be free from oxygen. While a significant challenge, this apparent disadvantage has a silver lining, primarily the lack of a need to supply oxygen to the system. Some aerobic organisms require oxygen supply to allow for their fermentation (255,256). Oxygen levels for such organisms must be tightly controlled. Too little and product yield suffers, too much and reactive oxygen species kill cells (257). Aerobic fermentation processes suffer from the gas-liquid mass transfer limitations described previously (1.2.1). This can be overcome by agitation, disrupting the state of the liquid and ensuring the liquid in the

fermenter is thoroughly mixed, ensuring better gas-liquid transfer and a more even distribution of cells (258). This can be addressed by using pressurised oxygen, as solubility increases as pressure increases (259), but this is expensive and potentially dangerous as pressurised oxygen is an explosive risk (260). Furthermore there is the issue of temperature, as temperature increases solubility decreases (259), which presents a problem for fermenting aerobic thermophiles which can grow in excess of 60°C (261,262). While acetogens growing on CO or CO_2/H_2 need to worry about gas-liquid mass transfer, this is not an issue when using methanol. Additionally, acetogens have the advantage of being able to use feedstocks that other fermenters cannot such as C1 chemicals (48,73,107,117).

Fermentation of *E. limosum* is essential not only for the end goal of the project but also as a useful tool for collecting data and developing strains. Continuous, or chemostat, fermentation results in a condition of perpetual exponential phase, the ideal condition for assessing the physiology of the cell. By measuring gases in and out via gas chromatography and performing HPLC analysis of cellular samples taken at regular intervals, valuable data can be acquired (263,264). Said data can identify the baseline “wild type” phenotype, in terms of specific productivity of products of interest, gas consumption, and growth rates. This baseline can be used as a reference for comparison with mutants. Furthermore, continuous fermentation can be used as a form of directed evolution (265). Over the course of as little as a month, thousands of generations of *E. limosum* can reproduce. The conditions of the reactor provide a selective pressure in of themselves, and experiments can be designed to select for a desired phenotype such as an increased resistance to a chemical. This type of fermentation, to select for a desirable phenotype, has been understood for many decades and allowed for improvements in industrial strains prior to genetic modification and the understanding of DNA (26).

1.7 Aims of the project

The wild type ATCC 8486 strain of *E. limosum* had several disadvantages for industrial applications. Firstly, its excessive adhesive properties limited the use of lab scale techniques, such as streaking to single colonies and resuspension, and would cause significant cleaning problems in fermenters. Secondly, its dependence of yeast extract supplementation of growth medium would necessitate the use of an expensive growth medium which would limit commercial viability. Therefore, a method for strain improvement which could significantly alter the organism without the need for the mechanism of changes to be understood would need to be selected and developed (Chapter 3). It would also be of interest to investigate any genetic alterations which occur to that of a similar work done with an unknown isolate of *E. limosum* which acted as the parent for the NG-6849 strain (Professor Philippe Soucaille, INSA, University of Toulouse, personal communication). In addition, the established genetic modification tools had significant drawbacks, being slow, inefficient, and unreliable for scarless modification. Improving existing tools and aiding in the development and testing of new tools could resolve these issues. Once established these tools could be exploited to better understand the role of various genes in methanol metabolism and butyrate metabolism and how the function of these genes compared to their homologues in other organisms (Chapter 4). These are worth investigation as methanol usage of one of the points of interest for *E. limosum* as compared with other acetogens and butyrate is both a potential product of interest and would be a competing product for engineered products. In addition, developed tools could allow for the introduction of non-native products (Chapter 5) to *E. limosum* which would address whether the organism and the molecular toolset are robust enough to warrant further research and development towards industrial use.

1.8 Collaborations

As part of this PhD, I was part of an international collaboration called EraCoBiotech BioMetChem. The goal of the project was “Sustainable Production of Added Value Chemicals from Syngas-derived Methanol Through Systems and Synthetic Biology Approaches”. The originally stated goals included production of the target chemicals γ-aminobutyric acid, GABA, and 1,4-butanediol.

In addition to the University of Nottingham, collaborators included Ulm University, Institute of Microbiology and Biotechnology, HES, Germany, Goethe University Frankfurt Molecular Microbiology & Bioenergetics, HES, Germany, INSA, University of Toulouse, LISBP, Biochemical Engineering, HES, France.

As part of this project, I contributed data towards the end goal of the project. The project’s goals were aligned with that of my PhD project, and the valuable insight gained by working with researchers familiar with *E. limosum* was useful for this work.

Chapter 2

Materials and Methods

Materials and Methods

All bacterial species and strains used in this study are listed in Table 1.

Table 1 - List of all strains used.

Organism	**Strain**	**Source**
Escherichia coli	NEB® 10-beta	New England Biolabs
Escherichia coli	HB101	Promega
Eubacterium limosum	ATCC 8486	Leibniz Institute DSMZ
Eubacterium limosum	NG-6849	National Institute of Applied Sciences of Toulouse
Eubacterium limosum	ASP	This study
Eubacterium limosum	ASF01	This study
Eubacterium limosum	ASF02	This study
Eubacterium limosum	ASF03	This study

All plasmids used and created in this study are listed in Table 2.

Table 2 – List of all plasmids used.

Name	**Source**	**Description**
pMTL83151	SBRC Nottingham Culture Collection, Heap *et al.* (266)	RepH, ColE1, *traJ* and *catP* containing modular plasmid
pMTL87151	SBRC Nottingham Culture Collection, Heap *et al.* (266)	p19r, ColE1, *traJ* and *catP* containing modular plasmid
pSBRCas9/ pmtl83151-Cas9-2456-BM	SBRC Nottingham Culture	A modified pMTL83151 vector containing a *Streptococcus pyogenes* Cas9 gene under the control of P_{thl},

	Collection, Tom Bailey	an P_{araE} promoter driving Cas9 guide expression and an fdx terminator region, before which Cas9 guides are to be placed and after which homology arms are to be placed
pSBRCas9-cpf	This Study	A modified pSBRCas9 vector with *C. sporogenes* P_{fdx} driving guide expression
pSBRCas9-Cpf-*act*	This Study	A modified pSBRCas9_cpf containing guides and homology arms to allow for the knockout of *act*
pSBRCas9-Cpf-*cop*	This Study	A modified pSBRCas9_cpf containing guides and homology arms to allow for the knockout of *cop*
pSBRCas9-Cpf-*metF1*	This Study	A modified pSBRCas9_cpf containing guides and homology arms to allow for the knockout of *metF*
pSBRCas9-Cpf-*hisI*	This Study	A modified pSBRCas9_cpf containing guides and homology arms to allow for the knockout of *hisI*
pSBRCas9-Cpf-*metVmetF*	This Study	A modified pSBRCas9_cpf containing guides and homology arms to allow for the knockout of *metVmetF*
pSBRCas9-Cpf-*mti*	This Study	A modified pSBRCas9_cpf containing guides and

		homology arms to allow for the knockout of *mti*
pSBRCas9-Cpf-*mtiI*	This Study	A modified pSBRCas9_cpf containing guides and homology arms to allow for the knockout of *mtiI*
pSBRCas9-Cpf-*pbt1*	This Study	A modified pSBRCas9_cpf containing guides and homology arms to allow for the knockout of *pbt1*
pSBRCas9-Cpf-*pbt2*	This Study	A modified pSBRCas9_cpf containing guides and homology arms to allow for the knockout of *pbt2*
TAKOS	SBRC Nottingham Culture Collection, Dr. James Millard	A modified pMTL83151 containing a Type I toxin-antitoxin system made up from a RelE toxin and a RelB antitoxin driven by a tetracycline-controlled inducible promoter system
TAKOS-*act*	This Study	A modified TAKOS containing homology arms to allow for the knockout of *act*
TAKOS-*cop*	This Study	A modified TAKOS containing homology arms to allow for the knockout of *cop*
TAKOS-*cop*-short	This Study	A modified version of TAKOS-*cop* with shorter homology arms
TAKOS-*bcact*	This Study	A modified TAKOS containing homology arms to

		allow for the knockout of *bcact*
TAKOS-*metF1*	This Study	A modified TAKOS containing homology arms to allow for the knockout of *metF*
TAKOS-*metVmetF*	This Study	A modified TAKOS containing homology arms to allow for the knockout of *metVmetF*
TAKOS-*mti*	This Study	A modified TAKOS containing homology arms to allow for the knockout of *mti*
TAKOS-*mtiI*	This Study	A modified TAKOS containing homology arms to allow for the knockout of *mtiI*
TAKOS-*mtiI*-short	This Study	A modified version of TAKOS-*mtiI* with shorter homology arms
TAKOS-*pbt1*	This Study	A modified TAKOS containing homology arms to allow for the knockout of *pbt1*
TAKOS-*pbt2*	This Study	A modified TAKOS containing homology arms to allow for the knockout of *pbt2*
TAKOS-*pbt2*-short	This Study	A modified TAKOS-*pbt2* containing short homology arms (500 bp) to try to resolve issues of antibiotic

		resistance retention in double knockouts with *pbt2*
TAKOS-*IlvE*	This Study	A modified TAKOS containing homology arms to allow for the knockout of *IlvE*
TAKOS-*glyA*	This Study	A modified TAKOS containing homology arms to allow for the knockout of *glyA*
TAKIS-*cop*	This Study	A modified version of TAKOS-*cop*-short with the wild type *cop* gene present between the homology arms to knock in (hence TAKIS) *cop* and compliment the gene into a Δ*cop* mutant
TAKIS-*mti*	This Study	A modified TAKOS-*mti* with the wild type *mti* gene present between the homology arms to knock in *mti* and compliment the gene into a Δ*mti* mutant
TAKIS-*mtiI*	This Study	A modified TAKOS-*mtiI* with the wild type *mtiI* gene present between the homology arms to knock in *mtiI* and compliment the gene into a Δ*mtiI* mutant
pMTL84151-acetone-thl (24-1)	SBRC Nottingham Culture	A plasmid used in *A. woodii* containing the *C. acetobutylicum* acetone production genes

	Collection, Dr. Jonathan Baker	
pMTL87171_CA_acetone_Pfdx	This Study	A modified pMTL87151 plasmid maintaining a synthetically derived codon optimised acetone production genes based on those from *C. acetobutylicum* driven by *C. sporogeneses* P_{fdx}
pSBR-CpfAcetone	This Study	A modified pSBRCas9-cpf with the Cas9 and CRISPR related components removed with the *C. sporogeneses* P_{fdx} driving a codon optimised and synthetically derived *C. acetobutylicum* acetone production genes
pMTL8J15 Acetone 77 thl	SBRC Nottingham Culture Collection, Dr. Robert Mansfield (117)	A pMTL87151 derived vector containing the native *C. acetobutylicum* acetone production genes under the control of *C. sporogeneses* P_{fdx}
pMTL8J15 Isopropanol 77 thl	SBRC Nottingham Culture Collection, Dr. Robert Mansfield (117)	A pMTL87151 derived vector containing the native *C. acetobutylicum* acetone and an additional secondary alcohol dehydrogenase operon under the control of *C. sporogeneses* P_{fdx}
ACE_*hisI*_Acetone_KI	This Study	A pMTL83151 derived vector containing a long and short homology arm to allow

		for ACE to occur, with a cargo of the native *C. acetobutylicum* acetone production genes under the control of *C. sporogeneses* P_{fdx}

All primers used in this study are listed in Table 26 in the appendices.

2.1 Fermentation and culturing

2.1.1 Growth media

Different growth media were used throughout this work. Growth medium varied as the project developed, but also varied with the different organisms that were needed as part of this work.

2.1.1.1 LB medium

LB liquid medium was used for cultivating competent *E. coli* cells, either to produce chemically competent cells, to recover them or to grow them for vector production. LB agar plates were used to recover and select for colonies with inserted vector.

For 1 litre of liquid medium; Tryptone, 10 g; Yeast Extract 5 g; NaCl 10 g. For 1 litre of solid medium as for liquid but for the addition of 15 g of agar. Both liquid and solid media were autoclaved, though liquid medium could be filter sterilised if needed immediately.

2.1.1.2 Pacaud medium

Pacaud medium is based upon work by Pacaud et al. (254). It is a fully defined growth medium and does not require the addition of yeast extract. Yeast extract is included as an optional component in the recipe, this is for use with strains that cannot grow without yeast extract.

For 1 litre of liquid Pacaud medium: NH_4Cl, 1.0 g; NaCl, 1.5 g; resazurin (0.02%), 1 mL; cysteine hydrochloride, 0.5 g; trace element solution (Table 4), 10 mL; macromineral solution (Table 5), 50 mL; vitamin solution (Table 3), 10 mL. The carbon source for growth was either methanol, fructose, or glucose, typically concentrations were up to 200 mM Methanol, 10 mM fructose and 10 mM glucose. Addition of yeast extract is optional, when added was up to 2 g per litre.

An anaerobic 50 mM Na_2S solution was added after inoculation to a final concentration of ~3 mM. The use of Na_2S as a reducing agent was limited to serum bottle work, to avoid damage to anaerobic cabinet catalysts. For use in an anaerobic cabinet L-cysteine was used to a final concentration of 1.5 mM.

Table 3 - Pacaud medium vitamin solution 1 litre

Biotin	2 mg
Lipoic acid	5 mg
Pantothenic Acid	5 mg

Table 4 - Pacaud medium trace element solution 1 litre

$FeSO_4 \cdot 7H_2O$	0.1 g
$MnCl_2 \cdot 4H_2O$	0.1 g
$CaCl_2 \cdot H_2O$	0.17 g
$CoCl_2 \cdot 2H_2O$	0.1 g
$ZnCl_2$	0.1 g
$CuCl_2$	0.02 g
H_3BO_3	0.01 g

Na_2MoO_4	0.01 g
NaCl	1 g
Na_2SeO_3	0.017 g
$NiSO_4 \cdot 6H_2O$	0.026 g
Nitrilotriacetic acid	12.8 g
pH	6.5

Table 5 - Pacaud medium macromineral solution 1 litre

KH_2PO_4	6 g
$NaCl_2$	12 g
$MgSO_4.7H_2O$	2.4 g
$CaCl_2 \cdot 2H_2O$	1.6 g

For 1 litre of solid medium as for liquid but for the addition of 15 g of agar. For serum bottle growth initially $KHCO_3$, at 6.0 g/L, was added to provide CO_2 which was found to be highly consumed in methanol growth conditions. This proved insufficient as methanol was not fully consumed, so this was increased to 10 g/L. Liquid medium was filter sterilised where plate medium was autoclaved. Before inoculation medium was left in an anaerobic environment for at least 24 hours. Pacaud medium plates used for induction work were identical but for the addition of 100 ng/mL anhydrotetracycline. Antibiotic selection plates were made up as prior but with antibiotic added as detailed in 2.1.1.3.

2.1.1.3 Antibiotic concentrations

Antibiotics were made up to a working stock to allow for easy dilution into media, to a stock concentration (Table 6) to allow simple scaling to calculate required volume (e.g., 50 µL of antibiotic stock in 50 mL of liquid medium). Chloramphenicol was used in *E. coli* where thiamphenicol was used in *E. limosum*.

Table 6 - Antibiotic concentrations

Antibiotic	**Stock Concentration**	**Working Concentration**	**Solvent**
Ampicillin	50 mg/mL	50 µg/mL	100% dH_2O
Chloramphenicol	25 mg/mL	25 µg/mL	100% Ethanol
Thiamphenicol	7.5 mg/mL	7.5 µg/mL	50% Ethanol

2.1.2 Culturing

2.1.2.1 Serum bottles

While growth conditions varied based on the purpose of the experiment, the general conditions were as follows: 250 mL gas serum bottles with a rubber stopper and aluminium crimp were used to contain the growth medium, the growing cells, and the overhead gas. Typically, 30 or 60 mL of growth medium was used. Due to the addition of bicarbonate to the media, an overhead of anaerobic cabinet gas (~8% hydrogen, ~77% N_2, ~15% CO_2) was sufficient. The feedstock varied based on the experiment. Growth was carried out at 37°C in a shaking incubator at 225 rpm. Inoculation was generally performed with sufficient biomass for a starting OD_{620} of ~0.05.

2.1.2.2 Batch fermentation

Batch fermentations were conducted in a New Brunswick BioFlo® 1.5 L reactor with a BioFlo® 115 controller. Conditions for batch fermentation varied based on the experiment however, the general conditions were as follows. An initial starter culture of 30 mL was set up from a single colony. This starter culture was then used to create a fermenter starter culture of 100 mL in serum bottles (2.1.2.1), this was grown to a sufficient biomass to allow for a starting OD_{620} of 0.05 in the fermenter. The reactor vessel was filled with liquid medium to a maximum of 1 L, with the rest being gas headspace. The reactor was sparged continually for 48 hours prior with N_2 gas to ensure anaerobic conditions. A N_2 and CO_2 gas feed, generally in a mix of 10% CO_2 and 90% N_2, was maintained throughout experimentation though the ratio of gases, and addition of other gases, would vary based on the experiment. The pH was measured by a pH probe in the lid of the reactor and controlled by the BioFlo® 115, which had two pumps set to control pH by the addition of acid or base. The pH range was generally 6.5-7.5. Temperature was maintained at 37°C throughout experiments.

2.1.2.3 Continuous fermentation

Continuous fermentation experiments were conducted like the batch fermentations (2.1.2.2), but with additional equipment that enabled the continuous addition of fresh medium and withdrawal of spent medium. In addition to the liquid medium inside reactor vessel, an external reservoir of medium was utilised. This medium was pumped into the reactor via a peristaltic pump calibrated to determine the feed rate. Concurrently, medium removal occurred through an out tube at the set maximal liquid volume of 1 L to prevent overflow and ensure waste medium removal. A flow rate of between 100 mL per day to 200 mL per day was employed, which would vary as required to maintain biomass.

2.1.2.3.1 Initial adaptive fermentation

Starter cultures and inoculation was as detailed previously (2.1.2.2) from single colonies. Some of this fermenter starter culture was saved to act as the “before” point for later genome sequencing. A 20 L vessel containing Pacaud liquid medium without yeast extract and with 100 mM methanol was used as the medium reservoir for the bioreactor. The fermentation was conducted as described previously (2.1.2.3) with some minor modifications. The growth medium used was supplemented with 2 g of yeast extract to allow the ATCC 8486 cells in the inoculum to gain sufficient biomass for continuous fermentation.

Throughout the fermentation, a mixed N_2 and CO_2 gas feed was maintained, consisting of a N_2 feed rate of 242 mL/minute and a CO_2 feed rate of 7.5 mL/minute, resulting in a ratio of 32.3:1 or approximately 97% N_2 and 3% CO_2. CO_2 addition began 1 hour prior to inoculation, with N_2 being supplied for several days beforehand. This ensured anaerobic conditions within the bioreactor while minimising CO_2 wastage. Live samples of cells were taken throughout the fermentation. Liquid was spread onto plates to assess the phenotype but also to identify any contaminants which may have been present.

2.1.2.4 Fermentation sample analysis

Samples were taken to assess the biomass (by optical density) and to assess the chemical composition of the medium (by HPLC). The frequency of sampling varied depending on the experiment and conditions.

In serum bottles, a sample was extracted using a needle and syringe through the rubber stopper that was crimped into its top. Initial samples of 1 mL were taken to measure biomass. In the fermenter setup, samples were taken daily. Liquid was drawn out of the fermenter, and the first 15 mL was discarded, as the sampling system retained ~5 mL. Three millilitres of sample were measured using a spectrophotometer to assess optical density, and the rest was used for HPLC sample preparation.

2.1.2.4.1 Optical density

To measure biomass in a culture a spectrophotometer was used with light scattering at a specific wavelength being used to give a value, with the wavelengths varying based upon the medium used, for Pacaud medium it was 620 nm, and for LB it was 600 nm. After reaching an optical density of 1, cell samples were diluted with water, and the recorded value was multiplied by the dilution factor. This step was necessary due to the limitations of spectrophotometers at higher cell densities. Increased cell mass can make it challenging to differentiate between actual increases in mass and optical density readings. By performing a dilution, the cell density is brought within a range that allows for accurate assessment.

2.1.2.4.2 HPLC

HPLC was used for the analysis of growth supernatant. A Thermo Scientific Dionex UltiMate 3000 HPLC machine was used for the HPLC work performed as part of this research.

A mobile phase of 5 mM H_2SO_4 was used throughout this work. Standards for the chemicals to be examined were prepared in ELGA water. These standards covered the expected ranges of concentrations; any detections outside of this range could only be quantified relative to the nearest calibration (e.g., >100 mM or <2 mM). A broad range of standards could be used, allowing 100-fold variation to be detected. An internal standard of valeric acid was used for samples and standards to account for any differences in injection volume.

Supernatant samples were stored at -20℃ to minimise degradation of the sample state. Both standards and samples were diluted in a 1 in 2 dilution (or 50:50) with the diluent (the mobile phase with internal standard). This was vortexed thoroughly to ensure as consistent a mixture as possible. 300 µl of this solution was passed through a 0.2 µm syringe filter directly into HPLC vials. This was done to remove any residual particulates from the solution as this can negatively impact the column and affect readings. Chromeleon™ chromatography data system software (Thermo

Scientific™) was used for instrument control, for running the standards and samples in sequence, and for initial data processing.

Most of this HPLC was done using a Bio-Rad Aminex HPX-87H 300 mm x 7.8 mm x 9 µm column, using a flow rate of 0.5 mL/min, an injection size of 20 µl, at 35°C for 55 minutes per sample, with a refractive index (35°C) and diode array (210 nm) detectors. A Rezex ROA-Organic Acid H+ (8%) 150 mM x 4.6mm x 8µm column was also available; its parameters were nearly the same with the difference being a shorter run time per sample, at 35 minutes. Data was processed to give concentrations in millimolar units.

2.1.2.5 *E. limosum* strains and strain storage

ATCC 8486 was supplied by Leibniz Institute DSMZ of Germany. This strain had significant adhesive properties once the stationary phase of growth was reached. The only way to mitigate the adhesive properties was to perform several washes with 2M NaCl. Cells would generally be killed in this process, so this was only performed for processes like genome extraction. Further this strain required the supplementation of growth medium with yeast extract to grow well.

The NG-6849 strain was supplied by Philippe Soucaille of the Institut National des Sciences Appliquées de Toulouse (INSA). This strain did not have significant adhesive properties, and it also was able to grow well without the need for yeast extract, allowing for the use of a fully defined medium. This strain was superior in all ways except for its sensitivity to oxygen. This issue was mitigated by the use of highly pure gases and procedures to ensure an anaerobic environment was maintained at all stages of cell growth.

Strains were stocked at -80°C, with DMSO acting as a cryoprotectant, added as 10% of the total volume. When a culture was defrosted, it was never refrozen; instead, a starter culture was grown, and multiple tubes restocked to replace it. This was done because to *E. limosum* cells were not being amenable to recovery after refreezing.

2.2 Cloning and associated processes

2.2.1 Cloning strains and culturing

2.2.1.1 NEB® 10-beta

NEB® 10-beta *E. coli* cells were bought premade as competent cells. Gene transfer into NEB® 10-beta *E. coli* was done per manufacturers instruction. NEB® 10-beta *E. coli* cells were typically used to transform ligations, especially triple ligations, their higher efficiency being important in such situations.

2.2.1.2 HB101

HB101 cells were typically used for tasks that did not require as high an efficiency as NEB® 10-beta cells, such as amplification of a complete and confirmed vector. Chemically competent *E. coli* HB101 cells were made in the lab.

Cells were grown from an overnight starter culture in LB liquid medium until OD_{600} reached ~0.5. Then, the cells were washed in a 100 mM $CaCl_2$ with 20% glycerol by volume mixture several times before aliquoting ~100 µL and freezing at -80°C. Batches of competent cells were tested before use by transforming a control plasmid, such as pMTL87151. If the efficiency was satisfactory, the cells were approved for work.

To transform vectors into *E. coli* HB101, a vector was added to an aliquot of cells and mixed well. The cells were left on ice for 30 minutes before being heat-shocked at 42°C for 45 seconds. Then, cells were left to recover on ice for 2 minutes before adding of 1 mL of LB liquid medium. The cells were then recovered for 1 hour at 37°C in a 225-rpm shaking incubator. After that, the cells were then concentrated by centrifugation, and 850 µL of LB was removed. The cells were resuspended in the remaining volume and spread onto LB plates with antibiotic selection. The cells were left for 24 hours, and the colonies were carried forward for further use.

2.2.1.3 *E. coli* strain storage

E. coli strains were stored in Pro-Lab Diagnostics Microbank™ vials. Storage and retrieval of strains were as per the manufacturer's instructions.

2.2.2 Plasmid extraction

For plasmid extraction from *E. coli*, a single colony was inoculated in LB liquid medium and grown overnight at 37°C and 225 rpm. The plasmid was extracted using the Monarch® Plasmid Miniprep Kit following the manufacturer's protocol. The same kit was used for plasmid extraction from *E. limosum*, but the yield was ~10 times lower than for *E. coli*.

2.2.3 Restriction digests and ligation

All restriction digest enzymes and T4 ligase were purchased from New England Biolabs (NEB®), with all digestions performing using the CutSmart® Buffer. High fidelity versions of enzymes were always used when possible or necessary. Restriction digests were carried out according to the manufacturer's instructions (Table 7). Digestion reactions deviated from manufacturer's instructions only when poor results were observed, with a longer incubation time typically being sufficient to remediate issues.

Table 7 - Typical digestion reaction

Reaction Component	**Quantity**
Enzyme(s)	1 µL
CutSmart® Buffer	5 µL
Target DNA (~1000 ng)	~5 µL
dH2O	50 µl

Ligation of both blunt-end and overhanging DNA fragments with T4 ligase was performed according to the manufacturer's instructions. Both triple and double ligations were conducted, with two insert fragments and one backbone and one insert

fragment and one backbone, respectively. Only DNA with sticky ends were used for a triple ligation. If a ligation required the use of blunt-ended products only, a double ligation was performed. Whenever possible, NEBioCalculator®'s ligation calculator was used to ensure an optimal ratio of vector to insert for both double and triple ligations. Ligations were usually left overnight on ice before being introduced into a cloning strain, typically NEB® 10-beta. However, some ligations were carried out according to the manufacturer's instructions.

2.2.4 Gel electrophoresis

A 1% agarose gel was typically used in this work. However, for products less than 200 bases long (e.g., CRISPR guides), 2% agarose gels were used, and for larger products (>5000 bp), 0.8% agarose gels were used. Agarose gels were made using TAE buffer with SYBR™ Safe DNA Gel Stain (Invitrogen™) to stain the DNA.

Gel Loading Dye, Purple (6X) from NEB® was used as the loading dye in this work. Several DNA ladders were used including GeneRuler 1 kb DNA Ladder, GeneRuler 1 kb Plus DNA Ladder, Thermo Scientific™ O'GeneRuler 1 kb Plus DNA Ladder and 2-Log DNA Ladder from New England Biolabs. All electrophoresis runs were conducted at 400 mA and 120 V. Gels were initially visualized on a blue light box before visualisation on a LI-COR machine.

2.2.5 DNA extraction from agarose gels

A blue light box was used to guide agarose cutting during the DNA extraction process from agarose gels. Monarch® DNA Gel Extraction Kits were used to extract bands from agarose gels, and the standard protocol was followed as per the manufacturer's instructions. Extracts were either used immediately or stored at -20°C. If immediate extraction was not possible then extracted bands were stored at 4°C with DNA extraction performed as soon as possible.

2.2.6 DNA quantification and quality control

To assess DNA integrity, samples were run on a 1% agarose gel (2.2.4). This was done for all genomic DNA extractions prior to sequencing to ensure no smearing. For the quantification of plasmids, a NanoDrop™ spectrophotometer was used to determine the concentration and approximate quality by the provided 260/280 nm value. The NanoDrop™ spectrophotometer was blanked with 1 μL of dH2O, and 1 μL of sample was loaded and measured. The concentration in ng/μL of DNA was recorded along with the 260/280 nm value.

Qubit™ fluorometric quantification was used for the assessment of the quality of gDNA due to its increased precision relative to NanoDrop™ and its ability to give accurate quantification of DNA in the presence of phenol unlike NanoDrop™. The Qubit™ fluorometric quantification was performed as per the manufacturer's specifications. To use the Qubit™ fluorometric quantification, a pair of blanks needed to be created as per the instructions in Qubit® dsDNA BR/HS Assay Kits. Samples were prepared following the instructions detailed in Qubit® dsDNA BR/HS Assay Kits, with 1 μL of sample being added to 199 μL of working stock solution. The Qubit™ fluorometer was blanked with the two blanks, the samples were inserted, and the data was displayed and recorded.

2.2.7 Vector confirmation and sequencing

Vectors were first extracted (2.2.2) and then assessed for quality and quantified (2.2.6). Analytical digests (2.2.3) were performed to extract the vector insert. For the minipreps that had the correct banding pattern, such as an insert of the correct size, sequencing was performed.

Sanger sequencing was conducted by either Eurofins Genomics or Source Bioscience. Samples and primers were prepared following the instructions of the sequencing service. Primers were designed to bind in regions of the plasmid that would not vary with the insert, ensuring a consistent read length across multiple inserts. This was typically done with the aid of Eurofins Genomics' sequencing primer design tool

(https://eurofinsgenomics.eu/en/ecom/tools/sequencing-primer-design/), with a manual search against the target being performed to ensure no binding elsewhere in the vector. It is worth noting that Sanger sequencing reads diminish in fidelity near the end of a read ~1000 bases. Larger sequencings therefore required the use of multiple primers which would overlap. Once the data was received, it was aligned against the theoretical in silico vector using the online software tool Benchling. If the sequences aligned well, then the vector was confirmed and used for its intended purpose. For PCR products, the process was similar, with gel-extracted products (2.2.5) sent for Sanger sequencing.

2.2.8 Plasmid storage

All plasmids used were stocked to allow for future use in this project and to act as both a resource for other researchers and a record of the work done. Plasmids were stored at -20°C, either in water or Monarch® DNA Elution Buffer (New England Biolabs®) for short term storage and once confirmed were transformed again into NEB® 10-beta and stored at -80°C (2.2.1.3).

2.2.9 NEBuilder® HiFi DNA assembly

In this work, the NEBuilder® HiFi DNA Assembly was used to assemble some plasmids. This assembly ligates several sequences together without the need for said sequences to be restriction digested beforehand by exploiting overlapping regions created by the prior amplification of the backbone and the insert sequences with primers designed to create said overlaps. Primers were created using the NEBuilder® tool (https://nebuilder.neb.com).

The NEBuilder HiFi DNA Assembly Master Mix (NEB #E2621) was used in this work, and the protocol was followed as per the manufacturer's recommendations, with the 60-minute protocol being used.

2.2.10 Creation of synthetic genes

Genes that could not readily be obtained or required significant modification were created synthetically. Designed in silico, the genes were synthesised by Integrated DNA Technologies (IDT) of Coralville, Iowa, USA. Synthesis required the genes to meet complexity requirements, including a lack of secondary structure, which often required codon optimisation or substitution. Genes were designed with additional regions at endpoints to make use of restriction digestion and ligation (2.2.3) or NEB HiFi (2.2.9).

2.3 Gene-transfer into *E. limosum*

Gene transfer into *E. limosum* was accomplished using electroporation. Electro-competent cells were made in the lab. Starter cultures were grown in Pacaud liquid medium with glucose as a carbon source. From this starter culture a subculture into ~300 mL Pacaud liquid medium was made, with an OD_{620} ~0.05. This was grown overnight ~12-16 hours. Cells were centrifuged and washed several times in SMP buffer (270 mM sucrose, 7mM sodium phosphate, 1mM $MgCl_2$, pH 7.2). Cells were then stored in SMP buffer at -80°C in 100 µL aliquots. Defrosted aliquots were used in their entirety.

A BioRad Gene Pulser Xcell electroporator was used for transformation of *E. limosum.* The standard programme was as follows: exponential decay wave, voltage 1.5 kV, resistance 200 Ω, capacitance 50 µF. Electroporation was performed using a sterile 2 mm electroporation cuvette (117). An additional programme was created for use in case of arcing events was a time constant based protocol, again using a sterile 2 mm electroporation cuvette, with a voltage of 1.7 kV was maintained for 6 ms. with resistance and capacitance set automatically by the electroporator. This was used for lower quality or older competent cells.

Plasmids to be transformed were typically transformed in 500 - 1000 ng quantities. Typically, a control of pMTL87151 was transformed into an aliquot of cells alongside that of the intended plasmids. The control served to ensure that any failure with

transformed cell growth was not due to low quality competent cells or an issue in the electroporation procedure.

After electroporation cells were recovered in anaerobic cabinet at 37℃ for 1 to 2 hours in 2 mL Pacaud medium liquid medium with glucose as carbon source. These were then plated onto Pacaud medium plates. Generally, colonies could be observed at around 2 to 3 days, depending on plasmid size and function.

2.4 Genomic DNA extraction

For all strains described in this study, genomic DNA extraction was performed using the GenElute Bacterial Genomic DNA kit by Sigma Aldrich, according to the manufacturer's instructions for gram-positive and gram-negative organisms, as appropriate, with cells grown overnight in suitable liquid medium as needed.

Genomic DNA samples were stored at -20℃, with freeze-thawing minimised to avoid shearing. The gDNA was assessed for quality before being used as either a template for PCR or for next-generation sequencing. Samples were run on a 1% agarose gel to look for smearing, which would indicate damage to the gDNA. A Qubit Fluorometric Quantification was then performed on samples to assess DNA quantity and quality (2.2.6).

2.5 Polymerase chain reaction

PCR was used extensively throughout this work, from creating components for vector to testing colonies to determine whether deletions were successful (267).

2.5.1 Polymerases

Three polymerases were used in this work: Q5® High-Fidelity DNA Polymerase, DreamTaq DNA Polymerase, and OneTaq® DNA Polymerase. The polymerase enzymes used were used in the form of a 2X Master Mix, and reactions were performed according to the manufacturer's instructions.

Q5® High-Fidelity DNA Polymerase was used for cloning during this project when high fidelity was required. DreamTaq DNA polymerase and OneTaq® DNA Polymerase (from Thermo Scientific™ and New England Biolabs®, respectively) were used for colony PCR.

2.5.2 PCR methods

2.5.2.1 Standard Q5 PCR

A standard Q5® High-Fidelity DNA Polymerase reaction was as follows; 25 µL of Q5® High-Fidelity 2X Master Mix, 1.25 µL of 10 µM Forward Primer, 1.25 µL of 10 µM Reverse Primer, 1 µL of template, and dH_2O to top volume up to 50 µL.

PCR programmes were designed to per NEB's recommendations (Table 8), with slight modification; 2.5.2.2 and 2.5.2.3 being modifications of this standard PCR protocol for Q5® High-Fidelity DNA Polymerase. Annealing temperatures were estimated using the NEB Tm Calculator (https://tmcalculator.neb.com/#!/main) (268).

Table 8 - Standard Q5® High-Fidelity DNA Polymerase thermocycler program

Step	**Temperature**	**Time**	
Initial Denaturation	98°C	30 seconds	
Denaturation	98°C	10 seconds	25 - 35 Cycles
Annealing	Varied up to 72°C	30 seconds	
Extension	72°C	40 seconds/kb	
Final Extension	72°C	2 minutes	
Hold	10°C	∞	

30 Cycles were typically done, though more or less were done when optimisation required it, such as for amplification of guides (2.6.1.2) initially 25 cycles were done, whereas colony PCRs tended to do 35 cycles.

2.5.2.2 Overlap Extension PCR

Overlap extension PCR (as shown in Table 9) was used to assemble homology arms for the early CRISPR-Cas9 vectors. Overlap extension PCR typically requires two PCR amplifications. An initial PCR amplifies the two regions to be combined using primers with overlapping ends. The samples are then run on agarose gels (2.2.4) and extracted (2.2.5) before being used as templated in the second, overlap extension, PCR.

The reaction for the first PCR was as standard, with genomic DNA from *E. limosum* typically used as the template. The reaction mixture used for the overlap extension PCR was also standard (2.5.2.1). For the second PCR 1 µL of each gel-extracted PCR product was added as a template.

Table 9 - Overlap extension Q5® High-Fidelity DNA Polymerase thermocycler program.

Step	Temperature	Time	
Initial Denaturation	98°C	30 seconds	
Denaturation	98°C	10 seconds	30 cycles
Annealing	Varied, though typically around 60°C	30 seconds	
Extension	72°C	40 seconds/kb	
Final Extension	72°C	2 minutes	
Hold	10°C	∞	

2.5.2.3 Colony PCR

A single colony was resuspended in 50 µL of sterile dH_2O using a 1 µL sterile loop. If the colonies were to be retained, the sterile loop used to pick and resuspend the colony was streaked onto an agar plate before being inserted into 1 mL of liquid medium in a cryogenic tube. This allowed for successful colonies to be grown for stocking at -80°C. The mixture of cells and dH_2O is then vortexed vigorously, and 0.5 µL is used as a template in a 25 µL PCR reaction. Primers were designed to bind in the vicinity of the locus of interest. When bands of the appropriate size were observed, they were extracted and sent for sequencing as described above.

2.5.2.3.1 Q5® colony PCR

Initial colony PCRs (colony PCRs) were performed with Q5® High-Fidelity DNA Polymerase (Table 10) but due to the increased cost DreamTaq was used for most colony PCRs.

Table 10 - Colony PCR Q5® High-Fidelity DNA Polymerase thermocycler program

Step	Temperature	Time	
Initial Denaturation	98°C	10 minutes	
Denaturation	98°C	1 minute	35 cycles
Annealing	Varied, though typically around 60°C	30 seconds	
Extension	72°C	1 minute/kb	
Final Extension	72°C	2 minutes	
Hold	10°C	∞	

2.5.2.3.2 DreamTaq colony PCR

While initial colony PCRs were performed (Table 11) with Q5® High-Fidelity DNA Polymerase, DreamTaq was used for most colony PCRs due to the lower cost.

Table 11 - Colony PCR DreamTaq DNA Polymerase thermocycler program

Step	Temperature	Time	
Initial Denaturation	95°C	10 minutes	
Denaturation	95°C	1 minute	35 cycles
Annealing	Varied, though typically around 60°C	30 seconds	
Extension	72°C	1 minute/kb	
Final Extension	72°C	2 minutes	
Hold	10°C	∞	

2.5.2.3.3 OneTaq® colony PCR

Due to a combination of supply chain issues that resulted from the Covid-19 pandemic OneTaq® (Table 12) it was used in many of the later colony PCRs performed in this work.

Table 12 - Colony PCR OneTaq® DNA Polymerase thermocycler program

Step	Temperature	Time	
Initial Denaturation	94°C	10 minutes	
Denaturation	94°C	30 seconds	35 cycles
Annealing	Varied, though typically around 50°C	30 seconds	
Extension	68°C	1 minute/kb	
Final Extension	68°C	5 minutes	
Hold	10°C	∞	

2.5.3 Sequencing PCR products

PCR products were sequenced similarly to plasmids (2.2.7). PCR products were run on agarose gels (2.2.4) and extracted (2.2.5), eluting into 15 μL of dH_2O.

2.6 Genetic modification

Three genetic modification techniques were successfully employed as part of this work, allele-coupled exchange, CRISPR-Cas9, and a Type I toxin-antitoxin system. A CRISPR-Cas3 system was also employed but it was not possible to replicate the success of other researchers.

2.6.1 CRISPR-Cas9 vector assembly

2.6.1.1 Homology arm production

Homology arms are pairs of sequences located on either side of a target gene. These homology arms are an essential part of many genetic modification vectors as they allow homologous recombination to occur. This enables bacterial cells to survive a double-strand break or other genetic damage and allows selection via the cessation of

auxotrophy or the use of other such counter-selectable systems (269–271). Homology arms were designed in silico using the online molecular biology software tool Benchling (benchling.com), as were amplification primers. Initially, the designed primers contained sites that allowed for digestion with enzymes, and the choice of enzymes was such that they would not cut elsewhere in the homology arms. The left homology arm reverse primer and the right homology arm forward primer had an overhang region to allow for overlap extension PCR (2.5.2.2).

As an alternative to the PCR amplification method of homology arm creation, some homology arms were synthesised. Typically, this was done for sequences for which it was difficult to design good amplification primers. Synthetic inserts were synthesised by Integrated DNA Technologies (IDT; Coralville, IA, USA). Synthetic inserts were designed to contain both homology arms and guide as part of a single unit. The homology arms were generally designed shorter in length than those created via PCR to keep the synthesis economical. Guides were designed as in 2.6.1.2, and the whole insert was assembled in silico using Benchling. Digestion sites at the ends of each section, between the guide and homology arms, and the ends of the insert itself, were added to allow for insertion into a vector. In silico analysis ensured no additional digestion sites for the enzymes of interest were present in the fully assembled vector.

2.6.1.2 Guide production

Guides are necessary to direct Cas3 and Cas9 to the target gene (168,169,200,272) for the CRISPR-Cas systems to function. For constructs making that use *Streptococcus pyogenes* Cas9, guides were designed using the PAM sequence NGG with Benchling. For the those making use of the endogenous Cas3, a PAM of TCA was used, with guides of 36 bases being created manually. Notably, the endogenous Cas3 works in a 3’ to 5’ direction, which is the reverse direction to the PAM and protospacer. Initially, guide sequences were created using PCR with primers designed in silico. The primers were designed to facilitate overlap extension PCR (2.5.2.2) (273) and had ends that would allow for digestion by enzymes to enable insertion into the vector.

2.6.1.3 Restriction digests and ligation

Homology arms and guide RNAs, produced through PCR amplification, were digested using two pairs of enzymes (2.2.3). One enzyme from each pair was the same, creating compatible ends, allowing the homology arms and CRISPR guides to form a continuous unit. The other enzyme in each pair differed for each component, ensuring their correct placement within the vector backbone. This facilitated a single-step triple ligation (2.2.3) process. When using a synthetic insert containing both the guide and homology arm pair as a single unit, vector assembly resembled that of the PCR amplification-derived insert. However, a single enzyme pair was used to position the synthetic insert in the vector, and a double ligation (2.2.3) was performed on the digestion products.

2.6.2 Type I toxin-antitoxin system vector assembly

A base vector containing the type I toxin-antitoxin system was used as the foundation of the work with this system (4.6). The NEBuilder® HiFi DNA Assembly system was used for the assembly of all type I toxin-antitoxin system vectors. This system offered both reliability and simplicity relative to the restriction digestion-based cloning methods used for CRISPR-Cas. Homology arms were designed in silico, and the New England Biolabs NEBuilder® tool was used to add the correct overhangs to primers for amplification of the homology arms and the vector.

Both the type I toxin-antitoxin system-containing vector and the homology arms were amplified using PCR (2.5.2.1), and upon confirmation of amplification (2.2.4) and extraction of products (2.2.5), made use of the NEBuilder® HiFi DNA Assembly master mix (2.2.9) to connect the two homology arm halves and vector into a single continuous vector.

2.6.3 Allele-coupled exchange

The use of ACE required an auxotrophic background with Δ*hisI* used as the auxotrophic mutant used for the ACE work in this study. These mutants required a partial deletion of the gene (approximately two-thirds) rendering it non-functional but allowing for a region against which homologous recombination could occur.

pMTL83151 was used as the basis for the creation of ACE vectors. Vectors were created using NEBuilder® HiFi DNA Assembly (2.2.9). Vectors consisted of left and right repair regions, which were effectively homology arms for the auxotrophic deletion (Δ*hisI*) site, containing as part of the repair region the native gene (*hisI*). Between these two regions was the cargo to be inserted. Vectors were confirmed (2.6.4) and introduced into Δ*hisI E. limosum* (2.3) and allowed to recover on rich medium. Cells were then plated onto a histidine-deficient medium, and all colonies that arose were mutants.

2.6.4 Vector transformation and quality assessment

Prior to transformation of ligations into *E. coli*, the ligations underwent gel electrophoresis on a 1% agarose gel (2.2.4) to assess their quality. Analytical digests were conducted on extracted plasmids using the same pair of enzymes employed for locating the insert in the vector backbone, ensuring successful integration of the insert. Subsequently, samples were sent for sequencing (2.2.7) once a satisfactory analytical digest had been obtained. For vectors containing Type I toxin-antitoxin systems, digestion was not performed; instead, PCR amplification of the insert region using sequencing primers was carried out (2.5.2.3.1). If the amplification was successful, the vector was subjected to Sanger sequencing (2.2.7).

Sequencing data were checked against an in silico vector map using Benchling. Generally, multiple minipreps from a single transformant plate were sent, and the best of those that had their sequences confirmed were carried forward for transformation into *E. limosum*.

2.6.5 Generating gene knockouts with CRISPR and toxin-antitoxin systems

To generate knockouts using the systems employed in this study, namely CRISPR-Cas9 and a Type I toxin-antitoxin system, a system-specific vector needed to be assembled, its sequence confirmed, and then the correctly assembled vector was transferred into *E. limosum* (2.3). The method for generating knockouts varied depending on the system used. Due to the failure of the endogenous CRISPR-Cas3 system to generate mutants, it seems imprudent to disseminate a method that does not work.

2.6.5.1 CRISPR-Cas9

After assembling CRISPR-Cas9 vectors (2.6.1) and confirming their sequences (2.2.7), they were introduced into E. limosum (2.3). The transformed cells were plated onto selective plates (2.1.1.2) and allowed to grow. Once colonies had formed, screening was performed using colony PCR (2.5.2.3.2). If mixed colonies were detected instead of pure knockouts, the mixed colonies were replated onto new selection plates and incubated for an additional week before retesting. This process was repeated until a pure knockout mutant was obtained. If no mixed colonies were observed in the initial colony PCR, random colonies were selected, replated onto selection plates, and allowed to grow further. If no mixed colonies emerged even after additional growth, it was assumed that the vector failed to induce the desired mutation.

2.6.5.2 Type I toxin-antitoxin system

Once the toxin-antitoxin system vectors were assembled (2.6.2) and their sequences confirmed (2.2.7), they were introduced into *E. limosum* (2.3). The cells were placed onto selective plates to grow. Once colonies arose, they were plated onto inductive medium (2.1.1.2). The cells were then left to recover from transformation and perform homologous recombination before screening using colony PCR with external primers (2.5.2.3.2).

2.6.6 Mutant screening

To test for successful genetic alteration many colonies were screened by colony PCR (2.5.2.3), with cells mass from colonies as the template for the PCR (267). To assess knockouts, a primer pair binding outside the homology arm region was utilised. This primer pair would generate a smaller product, corresponding to the size of the deleted gene, in the case of a successful deletion mutant. However, it would also produce a product in the case of a non-mutant cell. A control sample of wild type gDNA served as a positive control for colony PCR reactions.

For testing a purported knock-in, a primer pair, which bound within the inserted sequence and outside the homology arm region, was used. Additionally, an external primer pair was used in tandem. The external primers would yield a product in both genetic states, while the internal primers would only produce products if a successful knock-in occurred. Consequently, the external primers were employed first. Only upon observing a band larger than the wild-type size, as expected from the insertion of a gene, were the internal primers used.

When a deletion or insertion-sized band was observed it was extracted (2.2.5) and sent for sequencing (2.2.7) with primers that bind to the theoretical product. These primers were designed to do this without binding to the homology arms present in the plasmid. The sequence data was compared to an in silico version of the theoretical knockout/knock-in region of the genome using Benchling. If no additional wild-type band is observed, and the product of the colony PCR matches the expected product of a pure knockout/knock-in then the mutant was considered to be confirmed.

2.6.7 Plasmid loss

Vectors that can replicate within a cell after transformation need to be lost after their purpose is fulfilled. This was the case for the CRISPR-Cas9 vectors used in this work. Loss of the plasmid was needed to allow for additional knockouts to be created in the same background, and to perform phenotype assessments as plasmid retention may affect phenotype. Plasmid loss was essential to ensure the only variable between

assessed cell lines is the presence or lack thereof of a gene of interest. This was not an issue for the type I toxin-antitoxin system which made use of suicide vectors.

To encourage plasmid loss, mutant cells were grown without selection on plates and streaked to single colonies. This would result in the loss of plasmid as there was no selection encouraging the retention of the plasmid. Once colonies grew, they were plated onto a selection and non-selection plate pair, with a single colony being spread onto each plate type. Cells that grew on both plates had retained the plasmid, with those that only grow on the non-selective plate had lost the plasmid. Colonies that had lost the plasmid would have an advantage over those which retained it and would come to dominate in the population. For colonies that appeared to lose the plasmid, another round of colony PCRs was performed (2.5.2.3) to ensure they were pure mutants. If all colonies failed to lose plasmids, single colonies were taken from the non-selective plate, replated onto selection and non-selection, and screened again using colony PCR (2.5.2.3). Successful colonies were then stored, as described previously (2.1.2.5).

2.7 Genome sequencing and analysis

2.7.1 Sequencing

2.7.1.1 Sample preparation

Samples were prepared using a standard GenElute Bacterial Genomic DNA kit (2.4). DNA quality assessment was done using Qubit (2.2.6) and gel electrophoresis (2.2.4), for concentration and integrity, respectively. Samples were sent eluted in a solvent per the requirements of the sequencing company (Table 13) and at the concentration they required, typically 10 mM Tris-HCl at a DNA concentration of 50-100 ng/µL.

Table 13 - Genomic DNA quality control standards.

DNA concentration (Qubit)	**OD 260/280**	**OD 260/230**	**Yield**
100ng/µL	~1.8	2.0-2.2	At least 2 ug, average fragment length > 30Kb

2.7.1.2 Illumina sequencing with MiSeq

Illumina genome sequencing was done by Deep Seq (Nottingham, UK) using an Illumina MiSeq 300 cycle v2 kit. This would theoretically deliver a coverage of ~ 60x for each sample using 150 PE reads.

2.7.1.3 Oxford Nanopore sequencing with MinION

Oxford Nanopore genome sequencing was done by Deep Seq (Nottingham, UK) using a MinION Mk1B. This would theoretically give reads more than 4000 bases with a coverage of at least 20 times.

2.7.2 Assembly

Genomic sequencing data in the form of fastq files were received and processed using the QIAGEN Bioinformatics CLC Genomics Workbench 21/22. A previously sequenced genome for the type strain was used as a reference genome (117). Read pairs were imported into the program and mapped against the reference genome. A consensus sequence could be extracted from mapping when needed.

2.7.3 Single nucleotide polymorphisms

To identify SNPs, the variant detection tools of the QIAGEN Bioinformatics CLC Genomics Workbench were used. A fixed ploidy variant detection analysis was performed with a required variant probability of 90% and a minimum frequency of 80%, while all other settings were left as default. This ensured that SNPs that were not common to almost all reads of an area were ignored. A 90% threshold was chosen instead of 100% as with read depth in some parts exceeding 200 read, was not unlikely that one read across a region may not be perfect and could result in a genuine SNP being missed. By setting a slightly lower threshold, the chances of a false negative were dramatically reduced.

2.7.4 Insertion deletions (Indel) and structural variations

The detection of indels and structural variations was performed using the QIAGEN Bioinformatics CLC Genomics Workbench's dedicated tool, using the default settings. It is important to note that this tool has inherent limitations, as stated in the manual for the version of the program used (CLC Genomics Workbench 21.0.2). The manual specifies that, in its current version, the indel and structural variants tool is limited to detecting only intra-chromosomal structural variants. Consequently, there is a possibility that this tool may overlook certain indels associated with more substantial structural variations. To address this issue, manual assessment of the sequence alignments was conducted to identify structural variations.

2.7.5 Large variations

QIAGEN Bioinformatics CLC Genomics Workbench 22 allowed for rapid identification of misalignment on standalone read mappings using the “Find Low Coverage” tool in the navigation pane. Threshold levels for coverage could be set to allow differentiation between regions that aligned but with low read counts and those which did not align at all.

Chapter 3

Directed Evolution of *E. limosum* towards an improved phenotype

3.1 Introduction

The ATCC 8486 strain of *E. limosum* produces an extracellular polysaccharide with a similar phenotype to that of some *Clostridia* (274). This extracellular polysaccharide not only serves as a carbon sink, reducing the theoretical yield of target chemicals, but also hinders laboratory tasks due to its adhesive properties. This results in difficulty in pelleting cells using centrifugation, which can lead to carryover of medium when resuspending in a new growth medium and makes single colony isolation difficult.

These adhesive properties also caused problems in the transformation of plasmids into *E. limosum*. While cells could be pelleted using centrifugation, it became very difficult to resuspend them, which could cause arching in electroporation if competent cells were not fully resuspended when made. This was also an issue most kit-based methods, such as genomic DNA extraction (2.4), where proper resuspension of cells is essential. Washing cells with a 2 M solution of sodium chloride alleviated much of the issue and was used when performing genome extraction of the ATCC 8486 strain. However, this solution was not practical with living cells, where the salt carryover would not be readily removed and would impede or perturb growth. In the case of transformation of *E. limosum*, this would result in additional stress on the cells, in addition to electroporation, antibiotic selection, and the presence and maintenance of a plasmid-based system acting upon the cells.

In addition to the issues mentioned above, the ATCC 8486 strain was unable to grow without yeast extract present in the growth media. The lack of a fully defined medium made the creation and validation of accurate genome-scale models from the data generated in this project much more difficult (275,276). In addition, the presence of yeast extract casts doubts on whether growth on methanol is truly autotrophic in ATCC 8486. The requirement of yeast extract limits the use of some genetic modification techniques, such as ACE (2.6.3). As discussed previously (2.6.3), ACE is performed in two steps, the second of which is limited by this phenotype. A counter-selectable chromosomal gene target (164), which in the case of ACE as employed in this work confers autotrophy to an amino acid is needed to select for mutants. Antibiotic selection for the removal of wild-type cells, and the “repairing” of

the counter-selectable chromosomal gene target ensures a return to prototrophy (159). However, if the organism cannot grow in a defined medium, it is very difficult to create a deficient medium, limiting the practical usage of this technique. While some work had previously been done using casamino acid (117), this was still not fully defined, with casamino acid typically containing all essential amino acids.

Carbon monoxide gas growth is a common feature of many industrially used acetogens (48,277,278). *E. limosum* would, therefore, benefit from being able to perform this task like its peers, allowing its integration into existing systems that are already commercially viable. Furthermore, being able to grow on both methanol and CO may offer an advantage over other similar organisms. For instance, *C. acetobutylicum* is unable to grow on methanol. If *E. limosum* were engineered to produce a similar product spectrum, it could allow for *E. limosum* to compete with *C. acetobutylicum*. Having the ability to use multiple feedstocks allows for fermentation conditions to be altered to maximise profit and production. Moreover, it provides resilience to unforeseen events that affect global supply chains and prices of feedstocks.

Collaborators on the Era CoBiotech BioMetChem project had previously adapted *E. limosum* using continuous fermentation as a form of directed evolution to lose the extracellular polysaccharide and to allow for growth without yeast extract. This strain, NG-6849, was sent to Nottingham and used for most of the later work in this project and the Era CoBiotech BioMetChem project.

Questions arose surrounding the genetic causes of the altered phenotype and whether the changes observed were due to smaller or larger genetic alterations. Unfortunately, the progenitor strain of NG-6849 before its adaptation had been lost in the collaborating laboratory. This meant that it was not possible to pinpoint the changes in the adapted NG-6849 strain that was responsible for its new phenotype. To answer these questions, directed evolution to adapt the extracellular polysaccharide phenotype of ATCC 8486 was undertaken.

3.2 Aims

This chapter of work aimed to create an improved strain with a phenotype more amenable to laboratory handling. Such a strain would be able to grow in a fully defined growth media, without the need for yeast extract supplementation, and would no longer possess excessive adhesive properties. Once such a strain had been produced it was of interest to attempt to understand the genetic cause of the phenotypical changes which occurred. This would not only increase the understanding of the metabolism of *E. limosum* but may also serve as a foundation for further, directed, improvements to genetic regions of interest.

3.3 Results

A chemostat, or continuous fermentation, aims to keep cells in a perpetual exponential phase growth (or metabolic steady state), by removing biomass and supplying new feedstock. Whilst maintaining such conditions it is possible to allow for directed evolution, or adaptive fermentation, to occur. By enforcing a selective pressure over hundreds or thousands of generations of cells genotypical changes occur and can push cells, albeit slowly, towards a more desirable phenotype (279,280).

The adhesive properties of *E. limosum* ATCC 8486 were not desirable. The NG-6849 strain lacked this undesirable phenotype. Its altered phenotype had arisen after long periods of adaptive fermentation. In addition, this adaptation had also conferred a lack of requirement for yeast extract supplementation in growth media. It was thought that the phenotype could be replicated in the ATCC 8486 strain by an adaptive continuous fermentation where the reactor was brought towards a point of containing no yeast extract (3.3.1).

3.3.1 Initial pilot adaptive fermentation

Chemostat fermentation was performed as detailed previously (2.1.2.3.1). Figure 6 displays the optical density and methanol concentration throughout the first continuous fermentation undertaken in this study. Between 192- and 237-hours growth appeared to stagnate. This was due to the use of growth medium as a blank for spectrophotometry measurements. This blank consisted of Pacaud medium supplemented with 2 g/L yeast extract and had become gradually contaminated. This was likely due to the presence of yeast extract in the otherwise inert medium. A blank containing yeast extract was used as the initial growth medium contained yeast extract and as the rate of yeast extract consumption was unknown it was used throughout the fermentation. Contamination of the blank was not noticed until the fourth reading, where the blank was swapped out for an uncontaminated blank and OD_{620} returned to expected levels.

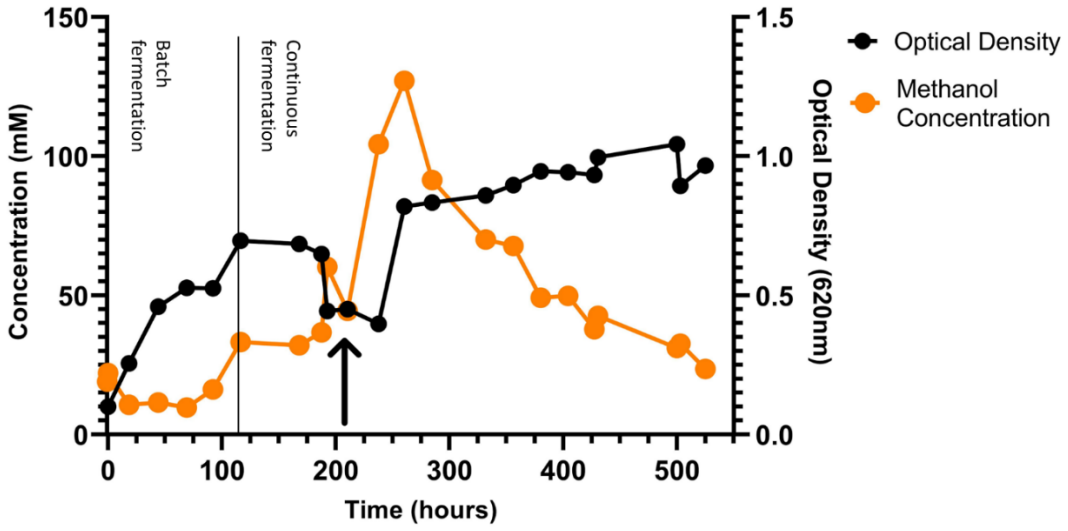

Figure 6 - E. limosum ATCC 8486 growth and methanol concentration throughout adaptive fermentation

Wild-type E. limosum ATCC 8486 was grown on methanol in batch, for 116 hours, then continuous fermentation in a 1.5 L fermenter, with 1 L of Pacaud growth medium and 500 mL of headspace for gas. Methanol was used as the substrate and a continuous gas flow of 249.5 mL/min with 97% N_2 and 3% CO_2. Pacaud growth medium containing no yeast extract was added from a 20 L feed tank (continually sparged with N_2) at a rate of 300 mL per day. Spectrophotometer readings for absorbance at 620 nm and samples for HLPC were taken every 24 hours; cell samples were taken every 5 days for growth on agar plates to assess phenotype adaptation. The arrow indicates the point at which additional methanol was added to the 20 L feed tank.

Additional methanol was added to the 20 L feed tank at ~210 hours to attempt to recover the perceived loss of biomass. This combined with a decreased feed rate was the likely cause of the spike in methanol concentration seen ~250 hours (Figure 6).

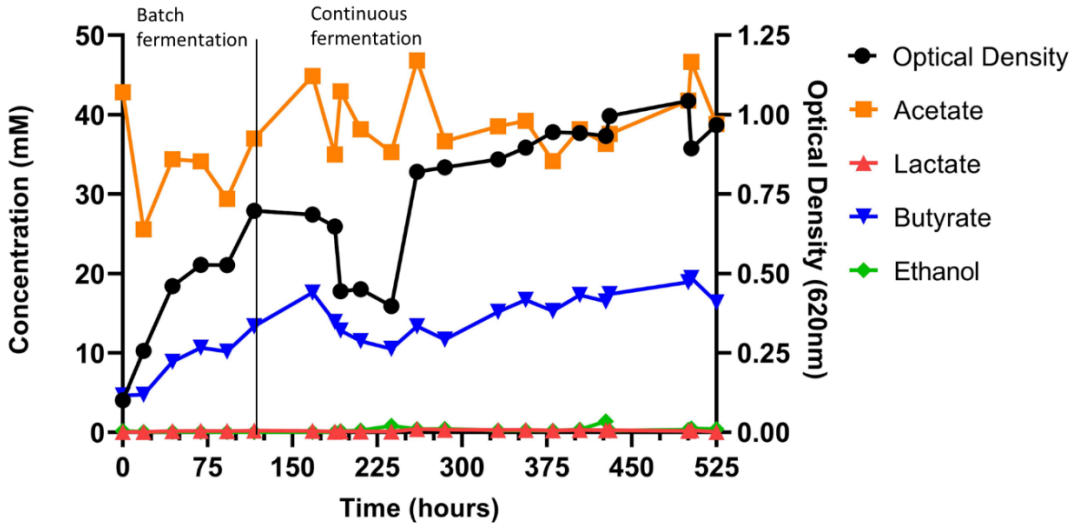

Figure 7 - E. limosum ATCC 8486 growth and metabolite concentration throughout adaptive fermentation

Wild-type E. limosum ATCC 8486 was grown on methanol in batch, for 116 hours, then continuous fermentation in a 1.5 L fermenter, with 1 L of Pacaud growth medium and 500 mL of headspace for gas. Methanol was used as the substrate and a continuous gas flow of 249.5 mL/min with 97% N_2 and 3% CO_2. Pacaud growth medium containing no yeast extract was added from a 20 L feed tank (continually sparged with N_2) at a rate of 300 mL per day. Spectrophotometer readings for absorbance at 620 nm and samples for HLPC were taken every 24 hours.

Ethanol and lactate levels (Figure 7) were low, as was expected for an *E. limosum* ATCC 8486 fermentation on methanol and CO_2. Methanol levels (Figure 6) varied throughout growth despite the intention of continual maintenance around 100 mM. Addition of methanol directly to the 20 L feed tank may have resulted in an inconsistent methanol feed as it is likely that methanol was being stripped from the 20 L media reservoir with the continual nitrogen sparging. A continual feed of methanol to the main reactor vessel using a peristaltic pump under the control of the BioFlo® 115 fermenter control system was adopted for subsequent fermentations to ensure a consistent feed of methanol to the reactor.

After ~20 days in the reactor, cells appeared lack the adhesive properties of the wild-type ATCC 8486, more closely matching that of the NG-6849. The cells were easy to spread on plates and easy to resuspend in liquid. Evolved cells were tested extensively

to ensure the retention of phenotype, with multiple rounds of plating on to solid medium including after stocking at -80°C. Given that the similar adaptive fermentation which resulted in NG-6849 also resulted in a loss of the requirement for yeast extract alongside the loss of extracellular polysaccharide, growth without yeast extract on the adapted cells was performed.

The adapted cells, however, had not lost their requirement for yeast extract supplementation in the growth medium. This resulted in a distinct phenotype, with reduced adhesive properties compared to the parent strain but a need for yeast extract supplementation in growth media. This cell line is referred to as adapted strain pilot or ASP. While this was initially perplexing, given growth in the reactor without yeast extract in the continuous stage, there are several factors that may explain this apparent inverse of previously observed data. Firstly, it is possible that a very small amount of yeast extract remained in the reactor. Medium was added and removed through the top of the reactor. This combined with continuous agitation maintained at 250 rpm a concentration gradient of yeast extract should not have been possible. However, it may have been that the time frame was insufficient to completely wash out all retaining yeast extract. The continuous stage of growth had a medium flow rate of 100 mL/day. It would therefore take 10 days to fully wash out the reactor, which should have occurred over one and a half times in the 409 hours of continuous fermentation. Secondly, it is possible that the cells were growing without yeast extract but could only do so under the conditions of a reactor, as in with continuous feeding of methanol and CO_2. The plate assay was done using glucose plates, so it did not replicate reactor conditions. It is possible that the adaptation for growth without yeast extract, in the case of this adapted strain, required a high flux of CO_2 down both the methyl and carbonyl branches of the Wood-Ljungdahl pathway. Therefore, while this adaption did work to a degree, further work was conducted in line with the original adaptation performed to create the NG-6849 strain from its natural isolate, as an adaptation for growth on CO (Professor Philippe Soucaille, INSA, University of Toulouse, personal communication).

3.3.2 Sequencing of adapted and non-adapted *E. limosum* ATCC 8486

To determine the genetic cause of the altered phenotype, genomic sequencing was required. Genomic DNA extracted from the fermentation inoculation culture was also sent for sequencing, acting as a reference point to the genome before adaptation, as described in the methods section above (2.7). Before the extraction of genomic DNA, a 16S rRNA assessment was performed from fermenter cell mass to ensure a pure culture. A colony PCR (2.5.2.3) using primers 16S 8F and 16S 1492R (Table 26) was then performed. These primers would amplify the 16S ribosomal RNA gene sequence, allowing for the identification of the species using NCBI's BLAST (281). Genomic DNA from a known *E. limosum* sample was used as a control and amplified alongside cell mass from the adaptive fermentation. Bands were extracted (2.2.5) and sent for Sanger sequencing (2.4.4) with 16S 8F and 16S 1492R. Sequences were inputted into NCBI's BLAST to look for alignments with organisms. Upon examination of the initial sequencing data, it became apparent that the sample was contaminated with non-*E. limosum* DNA, with 16S RNA gene sequencing suggesting *Stenotrophomonas maltophilia*. Given that this species is an obligate aerobe (282), contamination is likely to have occurred when cells were removed from the bioreactor. This was done in a crude manner using the sampling port, and in retrospect, it would have been better to integrate a fully contained system into the lid of the bioreactor.

Samples were plated out, and colonies that appeared to be *E. limosum* were selected. A colony PCR (2.5.2.3) using 16S 8F and 16S 1492R was then performed with gDNA from a known *E. limosum* sample acting as a control. Amplification products were again sequenced (2.4.4). Sequences were inputted into NCBI's BLAST, and only those that met the standard demonstrated by the gDNA control were trusted to in fact be *E. limosum*. Colonies that were confirmed to be *E. limosum* were grown on plates (2.1.1.2) to ensure that the phenotype change initially observed was retained. From these colonies, genomic DNA was extracted (2.7.1.1) and assessed for quality (2.4). A salt wash proved useful in this procedure for the wild-type ATCC 8486 cells but proved unnecessary for the adapted cells, which aligns with the phenotype described above. Sequencing was done both with an Illumina system and with Oxford Nanopore

(2.7.1), which allowed for structural comparisons to be made. Structural analysis using *de novo* assembly of Oxford Nanopore data and whole genome alignment was not performed until 2021, two years after the data was received, with the "Long Read Support (beta)" plugin, provided by Qiagen, not being available with support materials until July of 2021.

3.3.3 Oxford Nanopore data and structural changes

CLC genomics 21 was used to analysis of NGS data. The "Whole Genome Alignment" and "*De novo* Assemble Long Reads (beta) tool" plugins were downloaded to increase the utility of the base program. Oxford Nanopore data was imported, and a *de novo* contig list was created using the "*De novo* Assemble Long Reads (beta) tool". Once a sequence was created the "Whole Genome Alignment" tool was used to compare large structural changes between the contig list and the SBRC ATCC 8486 reference genome (Figure 8).

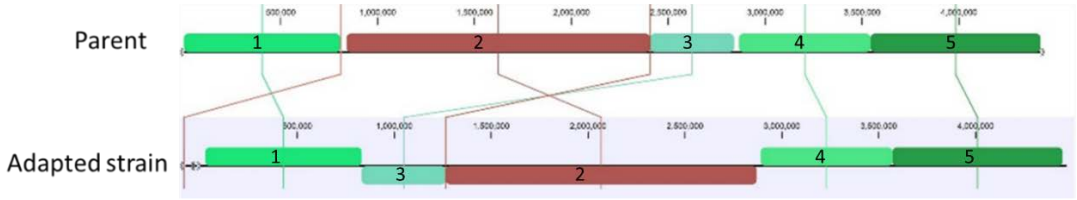

Figure 8 - Output of whole genome alignment between ASP ATCC 8486 and SBRC ATCC 8486

The adapted ASP shows structural differences relative to the reference sequence SBRC ATCC 8486 (on top). Different colours represent different regions and serve to distinguish regions which have moved. There are 8 regions in total, the 5 easily discernible blocks (1 – 5), but also two smaller regions between regions 1 and 2 (both of which moved to before region 1) and a small block between regions 2 and 3.

Analysis revealed an apparent rearrangement, resulting in two large regions (regions 2 and 3 in Figure 8) swapping their genomic location. The region 2 is inverted relative to the reference strain; however, region 3, retains directional parity with the reference strain despite changing genomic location. These genomic rearrangements may have affected rates of transcription, with some genes moved closer to the origin of replication. As it was possible that genetic disruption had occurred at break-point sites, breakpoint regions were examined to determine if any genes were affected. It was possible that disruption could cause affected genes to become non-functional or cause a fusion of two proteins. However, this did not appear to be the case, as most

breakpoints were located between genes or in non-coding regions. Where genes were affected, their function was not related to any of the observed phenotypes.

This likely means that if rearrangements play a role in the altered phenotype, they do so by altering transcription. This data was used as the rationale for sequencing the CO-adapted strains ASF01, ASF02 and ASF03 using Nanopore (3.3.4), as these strains displayed a similar adhesive phenotype and so may have undergone similar genetic rearrangements. Similar regions may have moved in this strain, which could be identified and studied further.

3.3.4 Initial adaptive fermentation on CO

To identify the genetic causes of the altered adhesive phenotype, triplicate adaptive fermentations were performed and NGS was carried out on the resulting cell lines. Cells were adapted to growth on CO, by slowly reducing methanol feed and gradually increasing CO feed, similar to the work done to create the NG-6849 strain. This experiment involved growing cells initially on methanol, then maximising dilution rate and slowly transitioning away from methanol towards carbon monoxide (Figure 9). It was thought that these fermentation conditions would lead to a strain that did not require yeast extract supplementation of growth medium.

The parent strain selected for these repeat fermentations was different to that for the initial pilot fermentation (3.3.1) due to issues in recovering the stock. The parent was taken as a single colony from an ATCC 8486 *E. limosum* stock streaked to single colonies on plates and again stocked (2.1.2.5) to allow for replicates to be performed. This colony displayed the native adhesive phenotype of ATCC 8486 *E. limosum*. Due to the expectation of adhesive properties and given the necessity of the isolation of single colonies the plate medium used (2.1.1.2) was fortified with additional agar (to ~20 g/L), allowing more force to be used to remove the colony without resulting in the agar plate breaking and contaminating the liquid growth culture.

The fermentations were set up similarly to the previous fermentation (2.1.2.3.1), using wild-type *E. limosum* ATCC 8486 in Pacaud medium (2.1.1.2), initially supplemented

with yeast extract to 1 g per litre and an initial carbon source of methanol, with a gas mix of 85% N_2 and 15% CO_2. Inoculated cells were grown in batch culture until an OD_{620} of ~1.5 was reached. Dilution with Pacaud medium, without carbon source or yeast extract, was gradually increased to a maximum of ~200 mL/day. Methanol was fed into the reactor via a peristaltic pump controlled by a BioFlo® 115, with the methanol feed rate varying according to the stage of the experiment, starting with ~3 mL/day of methanol and decreasing by the equivalent of 0.3 mL/day.

The methanol feed rate was decreased in tandem with an increase in the CO feed rate. This was done gradually with CO feed increased at a rate of 0.001 standard litres per minute (SLPM) per day, to ensure that toxicity issues were detected long before cell washout had occurred. Over 2 weeks, the CO feed was increased to a maximum of 0.030 SLPM. Due to operational issues, the gas supply was interrupted at around 875 hours for 24 hours, resulting in a decrease in biomass. The cell density of the culture was restored by reintroducing of methanol to increase biomass, before gradually removing it once more. Live cells were extracted from the fermenter every 7 days (and every few days after 900 hours) and spread onto plates to assess phenotype. After ~2 weeks of this adaptation, an altered phenotype was observed, where the adhesive properties of the wild-type ATCC 8486 was no longer present.

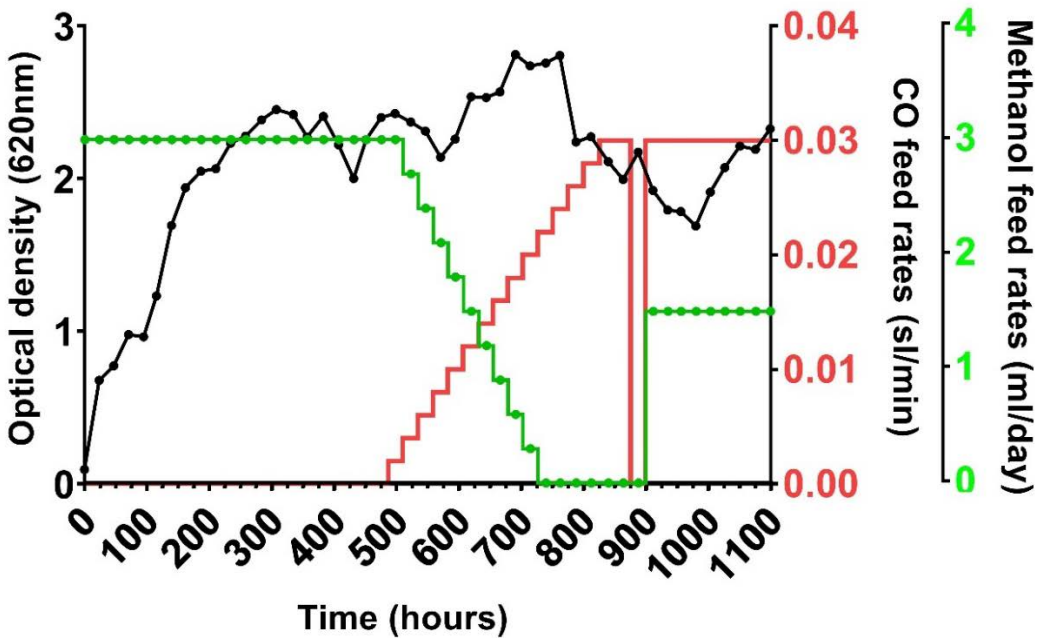

Figure 9 - Growth curve of E. limosum ATCC 8486 in continuous fermentation during CO adaptation

The OD is displayed on the black lines with the circles representing individual readings. Methanol feed is similarly represented by the green lines and circles. The red line represents CO feed rate. Growth curve of wild-type E. limosum ATCC 8486 fermentation in Pacaud medium. Cells were initially grown in batch in 1 g/L yeast extract-supplemented medium with a gas supply of 85% N_2 and 15% CO_2 until an OD of ~1.7 was reached. After biomass had reached this point continuous fermentation began at ~140 hours. Continuous fermentation was performed at a feed rate of 200 mL/day of medium without any yeast extract supplementation. Cells initially grew on ~3 mL/day of methanol, with a transition to Carbon monoxide beginning at ~500 hours, with carbon monoxide increasing 0.002 SLPM per day up to a maximum of 0.03 SLPM. After this growth cells lost the adhesive phenotype associated with the starting strain. Cell mass was removed from the reactor every 7 days to assess phenotype. Of note, at ~875 hours gas supply was interrupted.

Adapted cells were screened using 16S primers (16S 8F and 16S 1492R) via colony PCR, and Sanger sequencing was used to ensure that the cells were *E. limosum*. Adapted cell lines that were confirmed to be *E. limosum* were stocked (2.1.2.5) to await future work. The cell line isolated in this fermentation being referred to elsewhere in this study as adapted strain fermentation 1 or ASF01. From this initial repeat, several improvements were made to the fermentation protocol. Cellular tolerance to CO was higher than expected, as such cells were to be exposed to CO throughout all the repeat fermentations. The feed rate of CO was further increased to 0.002 SLPM per day until a feed of 0.030 SLPM was reached, with methanol feed being decreased in tandem at the same rate. Additional repeats were performed under this improved regime to provide triplicate cell lines for further analysis.

3.3.4.1 First repeat of the adaptive CO fermentation

A repletion of adaptive fermentation (Figure 10) was set up using the same methods described in 3.3.4 with the modifications described at the end of the previous section, which included continuous exposure of *E. limosum* to CO and an increased rate increase of CO feed.

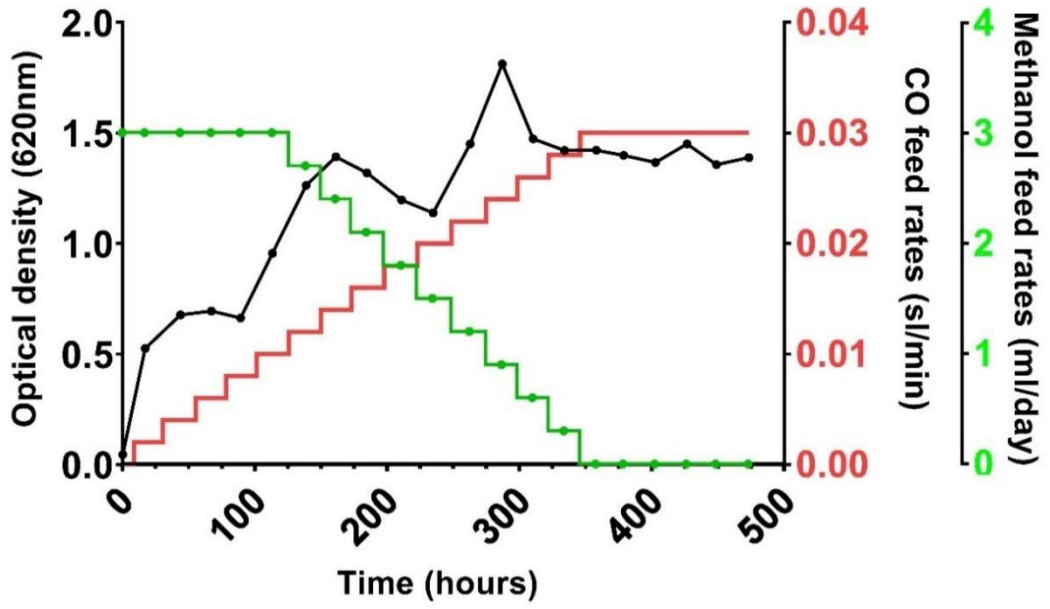

Figure 10 - Growth curve of E. limosum ATCC 8486 in continuous fermentation during CO adaptation

The OD is displayed on the black lines with the circles representing individual readings. Methanol feed is similarly represented by the green lines and circles. The red line represents CO feed rate. Growth curve of wild-type E. limosum ATCC 8486 fermentation in Pacaud medium. Cells were initially grown in batch in 1 g/L yeast extract-supplemented medium with a gas supply of 85% N_2 and 15% CO_2 until an OD of ~1 was reached. After biomass had reached this point continuous fermentation began at ~120 hours. Continuous fermentation was performed at a feed rate of 100 mL/day of medium without any yeast extract supplementation. Cells initially grew on ~3 mL/day of methanol, with a transition to CO beginning at ~120 hours, with CO feed increasing 0.002 SLPM per day up to a maximum of 0.03 SLPM. Cells were removed from the reactor every 4 days to assess phenotype. After 400 hours cells had lost the adhesive phenotype associated with the starting strain.

Cells were screened using 16S primers via colony PCR, and Sanger sequencing of the obtained products, to confirm that the cells were *E. limosum*. Cells displaying a decreased adhesive phenotype were stocked (2.1.2.5) to await future work . The cell line isolated in this fermentation being referred to elsewhere in this study as adapted strain fermentation 2 or ASF02. The adaptation occurred much faster than the previous one, taking around 2 weeks instead of 4 weeks. This is likely due to the

addition of CO much earlier into the fermentation, and the decreased time before the methanol feed was reduced.

3.3.4.2 Second repeat of the adaptive CO fermentation

A repletion of adaptive fermentation (Figure 11) was set up per the same methods described in 3.3.4.1.

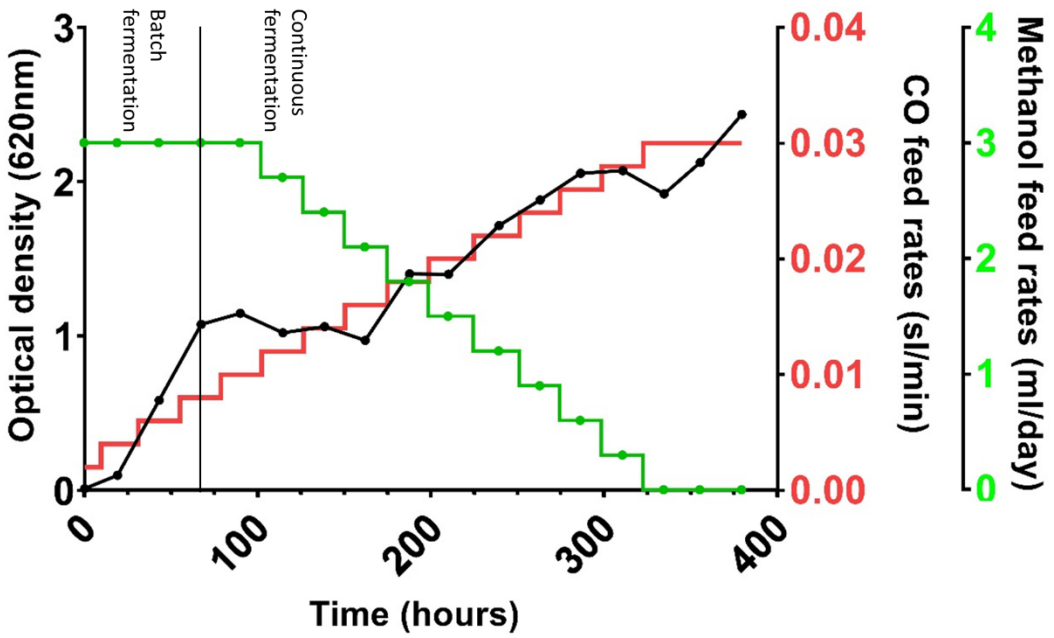

Figure 11 - Growth curve of E. limosum ATCC 8486 in continuous fermentation during CO adaptation

The OD is displayed on the black lines with the circles representing individual readings. Methanol feed is similarly represented by the green lines and circles. The red line represents CO feed rate. Growth curve of wild-type E. limosum ATCC 8486 fermentation in Pacaud medium. Cells were initially grown in batch in 1 g/L yeast extract-supplemented medium with a gas supply of 85% N_2 and 15% CO_2 until an OD of ~1 was reached. After biomass had reached this point continuous fermentation began at ~96 hours. Continuous fermentation was performed at a feed rate of 100 mL/day of medium without any yeast extract supplementation. Cells initially grew on ~3 mL/day of methanol, with a transition to CO beginning at ~96 hours, with CO feed increasing 0.002 SLPM per day up to a maximum of 0.03 SLPM. After this growth cells had lost the adhesive phenotype associated with the starting strain. Cells were removed from the reactor every 4 days to assess phenotype.

Due to emergency maintenance work in the lab all gas fermentation had to be shut off. This prevented the continuation of gas fermentation and so a sample was taken prior to gas shut off and plated onto solid medium. It was observed that only a minority of cells had non-adhesive properties like that of the NG-6849 strain, those cells were taken forward for further assessment. The cell line isolated in this fermentation being

referred to elsewhere in this study as adapted strain fermentation 3 or ASF03. This suggests that by 400 hours of growth a minority of cells have undergone adaption and in the subsequent 100 hours of growth become the majority of the population due to the advantage of their altered phenotype.

3.3.5 Phenotype assessment

Colonies from all repeat adaptations were taken and sub-cultured on plates to ensure the maintenance of the non-adhesive phenotype. Cells were grown in both liquid and solid medium. The changes incurred during adaptation appeared stable, strains did not revert to the adhesive phenotype of the ATCC 8486 wildtype and retained this phenotype on plates for weeks, with colonies transferred to fresh plates not reverting to the wild-type adhesive phenotype. Cells from all three adaptations were able to grow without the medium being supplemented with yeast extract. The phenotypes observed after adaptation closely matched those of the *E. limosum* NG-6849, strain suggesting that some genetic features may be shared.

3.3.6 Genomic sequencing and analysis

Single colonies were isolated after several re-platings for each repeat adaption experiment to ensure phenotype retention and clonal purity. Amplification of 16S RNA gene sequences from biomass was performed using colony PCR (2.5.2.3.1), and the resulting products sent for Sanger sequencing (2.5.3) to confirm that the cells were *E. limosum*. Upon species confirmation, genomic DNA extractions were performed (2.5) and the genomic DNA was sent for Illumina-based next generation sequencing (2.7.1).

Before analysis of sequence data could begin, it was necessary to ensure that the reference genome sequence used for this study, SBRC ATCC 8486, a sequence generated at the University of Nottingham (117), was of sufficient quality. Further given the unpublished nature of the reference strain sequence it seemed prudent to compare with a published sequence. A published sequence would be of a sufficiently high standard and by using the reference tags of a published sequence it would be

possible for researchers without access to the unpublished sequence to investigate any genes of interest themselves. This published strain was therefore used as a means of verification of the existing reference strain sequence. A published ATCC 8486 reference sequence was obtained published by the Korea Advanced Institute of Science and Technology (KAIST) (283). The sequence has the GenBank accession number of NZ_CP019962. Using the "Whole Genome Alignment" plugin for CLC Genomics a genomic alignment was performed between the SBRC and KAIST (283) sequences, which showed that they had over 99.9% identity. Single-base analysis allows for a comparison of each base in a genome compared to another. The discrepancy observed between the two strains is likely due to a ~1.5-kilobases size sequence that is present in KAIST but not in the SBRC sequence. *E. limosum* phage regions may also contribute to this discrepancy because phage regions in bacteria have a tendency to move (284). Besides this however there were no other indels or SNPs identified. This was observed extensively in the following genomic comparisons, and all phage regions were excluded from consideration as being candidates for the observed phenotypes. The main differences between the genomes were not sequence-related but in the annotation quality of the KAIST sequence, due to manual curation of the KAIST sequence. The KAIST genome was not used for any comparisons in this work beyond the verification of the circularised internal SBRC ATCC 8486 *E. limosum* sequence.

Once the reference sequence was confirmed to be of sufficient quality, the data from the adapted strains' Illumina reads had to be verified. CLC Genomics Workbench can produce reports for mapping. These reports provide several data points that can be useful in determining the quality of Illumina data. A low number of reads is indicative of poor coverage. The percentage of reads that align to the reference sequence can indicate contamination of the genomic sample. The average read length will vary dependent on the sequencing technique used, in the case this study reads lengths of up to 300 were expected (2.7.1.2). Sequence length distribution can serve as an indication of quality of reads, with read lengths outside the expected range indicating quality issues with either the prepared sample or the sequencing kit (130).

Illumina sequencing data for the parent strain and adapted strains ASF01, ASF02 and ASF03 consisted of over one million reads. These reads were all mapped to the SBRC *E. limosum* ATCC 8486 reference sequence with over 99% of reads mapping to the reference sequence. Once reads from the parent strain and the adapted strains were mapped to the reference SBRC ATCC 8486 sequence analysis could be performed.

3.3.6.1 Large misalignments

Several regions of low coverage were identified in both the adapted strains and the parent strain using the CLC genomics "Find Low Coverage" tool. Most regions of low coverage were phage related regions. CLC genomics may have struggled to align these regions due to the presence of multiple copies within the genome. This suggests multiple copies of phage genes throughout the genome, either from multiple copies of a single phage or of similar genes from multiple phages. More striking was a low coverage region in the genome of the parental strain between ~3,173,100 and ~3,174,250 (Figure 12).

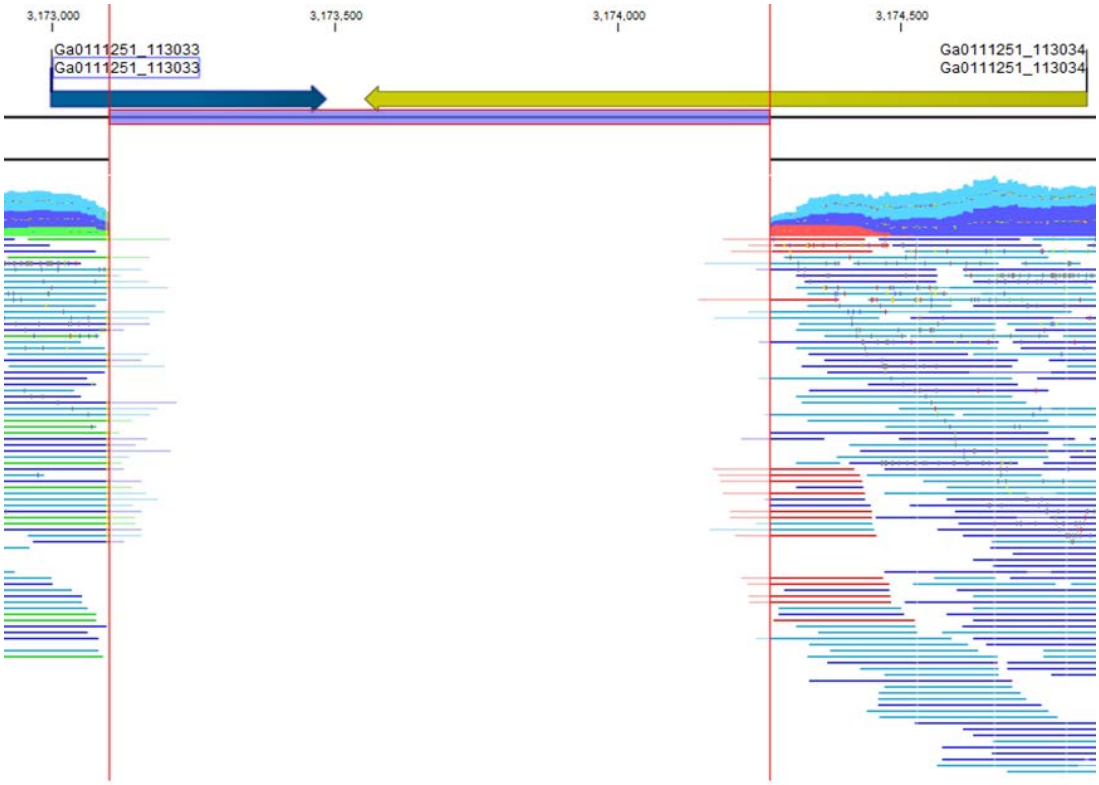

Figure 12 - Alignment of parent strain ATCC 8486 against the reference sequence SBRC ATCC 8486 displaying a misalignment at a tryptophan-rich sensory protein and M23 Metallopeptidase.

Output from CLC genomics workbench displaying a lack of reads across two genes, purportedly a tryptophan-rich sensory protein and M23 Metallopeptidase.

However, this region was complete and present in all the adapted strains, suggesting that the apparent gap is a sequencing artifact rather than a missing region. To corroborate, this a PCR was performed across this region (2.5.2.1, using the primers 3,173,000 SCR F and 3,173,000 SCR R). The resulting amplicon was sent for Sanger sequencing (2.5.3), which confirmed that the region was intact and present in the parent strain.

3.3.6.1 Single Nucleotide polymorphisms and indels

Single nucleotide polymorphisms (SNPs) are single base changes in a DNA sequence. A single base change can have significant effects, such as altering the amino acid translation, thereby affecting protein structure (285), or causing a premature stop codon to occur resulting in a truncated protein that was non-functional or significantly altered in function. Insertions or deletions of bases in a gene or genome (indels) can cause frameshift mutations, altering the entire amino acid sequence following the indel, which can significantly affect protein function. In the *E. limosum* literature indels appear to be a more reported cause of genetic changes than SNPs (286). Therefore, while both SNPs and indels were investigated the working hypothesis was that indels were more likely to be responsible for the phenotypical changes observed.

CLC Genomic Workbench has dedicated SNP and indel detection tools, which were used as detailed previously (2.7.3 and 2.7.4, respectively). Initially, analysis was performed between the SBRC ATCC 8486 reference sequence, and the parent strain used for adaptive fermentation (Table 14). This was done to exclude these SNPs and indels from the analysis; if they were common to the parent, then they are irrelevant to the altered adhesive and yeast extract growth phenotype.

Table 14 - Single nucleotide polymorphisms (SNPs) in parent strain E. limosum ATCC 8486 relative to the reference sequence SBRC ATCC 8486

Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Amino acid change
774978	SNP	G	A	71	73	97.26027	
1047857	SNP	G	A	87	89	97.75281	Arg174Cys
1272483	SNP	G	A	46	49	93.87755	
3173096	SNP	A	G	27	27	100	
4400072	SNP	G	A	100	100	100	Leu632Phe

These SNPs were all verified by reviewing the sequence alignment manually to ensure the regions were both of good read depth and alignment quality. Once verified they were eliminated from adapted strain SNP results, as they were present in the parent strain and thus not responsible from the altered phenotype observed. SNP analysis was then performed on the three adapted strains, ASF01, ASF02 and ASF03, using a fixed ploidy variant analysis as described previously (2.7.3). The initial SNP data set contained many SNPs which upon manual curation were found to be unreliable, primarily many SNPs in the region 3,721,700 - 3,721,800. These areas were of poor read coverage and manual verification of SNPs determined they were present in end of read regions. Once these low confidence SNPs were removed those which were of higher confidence, were assembled into a table and underwent further analysis (Table 15); rows highlighted in yellow were not within open reading frames, the rest were.

Table 15 - Manually verified single nucleotide polymorphisms and insertion/deletions (InDels) common to all adapted strains of E. limosum ATCC 8486 relative to the parent sequence SBRC ATCC 8486

Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Amino acid change
644544	SNP	A	G	163	163	100	Arg63Gly
906908	Deletion	T	-	65	81	80	Ile65 then frame shift
980564	Insertion	-	GGC	112	119	94	Gly insertion between Gly847 and Arg848
1227240	SNP	G	A	145	146	99	Gly335Glu
1247875	SNP	G	T	135	137	98	Thr336Asn
1540554	SNP	A	C	146	147	99	
2086124	Deletion	T	-	145	148	97	
2124539	SNP	G	T	132	133	99	Met78Ile
3306729	SNP	C	T	120	120	100	Gly117Asp
3306732	SNP	A	G	125	125	100	Ile116Thr
3857175	SNP	A	G	156	156	100	Tyr172His
3892157	SNP	C	T	171	171	100	
4139799	Deletion	A	-	184	187	98	

Several SNPs were initially identified, some of which resulted in coding changes. For all regions where coding change occurred, BLASTp and BLASTn searches were performed to look for similar genes/proteins in other organisms that may have been

characterised. This allowed for the elucidation of their significance regarding the phenotype change observed (Table 16).

Table 16 - SNP genes which underwent amino acid changes, the change, and the putative gene function.

Reference Position	**KAIST Annotation**	**Amino acid change**	**Gene**
644544	WP_038352518.1	Arg63Gly	Response regulator
906908	WP_013382081.1	Ile65 then frame shift	Hypothetical Protein
980564	WP_038352995.1	Gly insertion between Gly847 and Arg848	Alanine--tRNA ligase
1227240	WP_038352818.1	Gly335Glu	Alanine--tRNA ligase
1247875	WP_038352802.1	Thr336Asn	ABC transporter permease
2124539	WP_110060352.1	Met78Ile	Mobilization protein
3306729	WP_038351469.1	Gly117Asp	Purine permease
3306732	WP_038351469.1	Ile116Thr	Purine permease
3857175	WP_038350871.1	Tyr172His	Thioesterase

Upon review and based on the known annotations and BLAST data, it is unlikely that the genes affected by the SNPs have any effect on the observed phenotype, whether it is the adhesive properties of cells or the ability to grow without yeast extract. In the case of the adhesive phenotype, genes related to polysaccharide production, transport, or presentation were those expected to be modified by SNPs if they were the cause of the phenotypical changes observed.

3.3.6.2 Structural variations

The CLC genomics workbench indel and structural variants tool can also produce data related to structural variants. Structural variants in this context denote differences between genomes resulting in a different structure of the DNA where the bases may, or may not, be the same. This can include inversion of regions, where the sequence remains identical but the 5’ end is now where the 3’ end used to be, translocation, where a region’s position within the genome changes, and can also include copy number variants where genes are present at higher number than previously. When a

deletion is large enough, typically in excess of 50 based it can also be considered a structural variant (287). Structural variant analysis was performed in parallel with the indel analysis. Three structural variants (Table 17) were observed in the parent strain compared to SBRC ATCC 8486. However, the significant variance observed did not appear to result in a distinct phenotype, as evidenced by the adhesive properties of colonies on plates and the necessity for yeast extract supplementation of plates for colonies to grow, relative to the SBRC ATCC 8486 sequenced strain (3.3.4), so said variants in the adapted strains can be eliminated. It is surprising that such variation is observed, given that the parent strain was supposedly a sibling of the SBRC ATCC 8486. It is difficult to assess such variants manually, and verification would likely only possible though Oxford Nanopore sequencing. Variants in adapted strains were assessed similarly, and variants identical to the parent strain were eliminated (Table 18).

Table 17 - Structural variants between the parent strain and the SBRC ATCC 8486 reference strain.

Region	Name	Evidence	Length	Variant ratio	# Signatures	Left breakpoints	Right breakpoints	Mapping scores fraction	# Reads	Sequence complexity	Split group
1684858..>1684957	Inversion	Cross mapped breakpoints	2032041	0.505051	2	1684857,3716898		0.93,1.0	44	0.210395	1
3173096..3174269	Deletion	Cross mapped breakpoints	1174	1	2	3174269	3173095	0.98,1.0	33	0.321703	
<3716799..3716898	Inversion	Cross mapped breakpoints	2032041	0.505051	2	1684857,3716898		0.93,1.0	44	0.210395	1

Table 18 - Structural variants common to all of the adapted strains relative to the SBRC ATCC 8486 reference strain.

Region	Name	Evidence	Length	Variant ratio	# Signatures	Left breakpoints	Right breakpoints	Mapping scores fraction	# Reads	Sequence complexity	Split group
1684855..>1684954	Insertion	Tandem duplication	745309	1	2	1684854	2430163	1.0,1.0	23	0.127223	1
1720578..>1720677	Inversion	Cross mapped breakpoints	2035333	0.211268	2		1720577,3755910	0.98,0.98	10	0.140275	3
<2430064..2430163	Insertion	Tandem duplication	745309	1	2	1684854	2430163	1.0,1.0	23	0.127223	1
2430161^2430164	Insertion	Close breakpoints	0	1	2	2430161	2430163		46	0.193631	

2927735..2928893	Complex	Multiple breakpoints	1159	0.168808	4	2927734	2927751,2928893,2928893	1.0,1.0,1.0,1.0	21	0.03316	
3283512..3283886	Complex	Multiple breakpoints	375	0.878044	8	3283511,3283511,3283871,3283871	3283526,3283526,3283886,3283886	1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0	246	0.001232	
<3716801..3716900	Complex	Multiple breakpoints	2032043	0.981481	8	1684857,1684857,3716898,3716898	1684859,1684859,3716900,3716900	0.86,0.86,1.0,1.0,1.0,1.0,1.0,1.0	202	0.004826	2
<3755811..3755910	Inversion	Cross mapped breakpoints	2035333	0.211268	2		1720577,3755910	0.98,0.98	10	0.140275	3
4097199..4097384	Replacement	Paired breakpoint	186	0.53913	2	4097251	4097198		26	0.268226	
4071944..4072042	Inversion	Paired breakpoint	99	0.308696	2	4072042	4071943		13	0.152638	
4397747^4397886	Insertion	Close breakpoints	0	0.818182	2	4397885	4397747		65	0.248201	

CLC genomics identified many structural variants common to all three adapted strains (Table 18). However, due to the high read depth of the Illumina reads, those identified structural variants with low numbers of reads attributed to them are unlikely to be real. It seems somewhat unlikely that such significant arrangements could occur in three independent adaptive fermentations unless these were important to the adaptation. For most of the identified variants only a minority of the total coverage of their respective region (generally less than 10% of coverage of a region) suggested the change. This suggests that these variations were erroneously identified due to these regions producing ambiguous reads when sequenced with Illumina technology, greater read length may allow for a higher degree of certainty regarding these variants. However, two complex multiple break points (present in >200 reads) occurred. These may match up similarly to the structural variation observed in the original adaptive fermentation (3.1.4).

Given the uncertainty of the variants described by the Illumina data and given the import these variants may have if common to all three adapted strains resequencing both the parent and the adapted strains using Oxford Nanopore sequencing was imperative. Such sequencing would generate long reads allowing for *de novo* assembly, which was likely to have maintained the structural variations present relative to the reference sequence.

3.3.7 Oxford Nanopore sequencing of adapted cells

Oxford Nanopore sequencing was performed by Deep Seq (Nottingham, UK) using a MinION Mk1B on the parent strain, all adapted variants, and an independent strain of ATCC 8486 obtained from a stock as that purported to be the same used for the original SBRC reference sequence (2.7.1.3). Data was received in fastq format and the fast5 format. The fast5 format contains methylation data and could not be processed on the hardware available for the genomics portion of this work. As such methylation data was processed by DeepSeq.

All data was imported into CLC Genomics Workbench, and the Beta Long Read Support tools were used to assemble the reads into contig lists using *de novo*

assembly. For both the adapted strains and the parent two contigs lists were generated, one ~4.4 megabases (the full length of the *E. limosum* genome) and one <30 kb, the latter of which aligned fully with a region within the larger contig. As Oxford Nanopore data has a higher rate of errors relative to Illumina data (288), contig lists were then polished with the prior Illumina reads. These contig lists were then aligned using the whole genome alignment plugin with the SBRC ATCC 8486 reference sequence. This was primarily done to compare the parent ATCC 8486 to the published SBRC reference, though it provided useful data for all adapted strains. Alignment was then performed against the parent strain contig list (Figure 13).

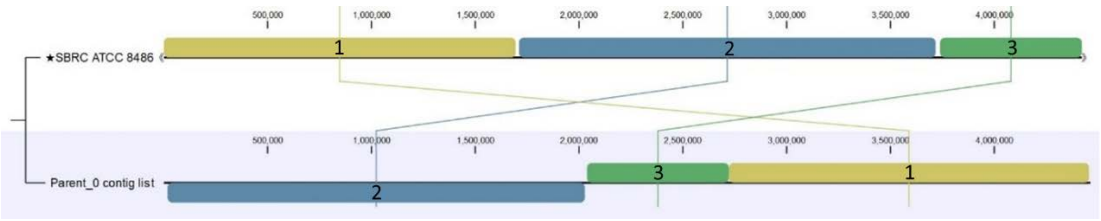

Figure 13 - Output of whole genome alignment between parent ATCC 8486 and SBRC ATCC 8486

The Parent ATCC 8486 (on the bottom) aligned against the SBRC ATCC 8486 (on top). The blocks of colour represent different regions and serve to distinguish regions which may have moved. The parent blue region being below that of the black line to which regions are placed along signifies an inversion in directionality relative to the genome it was aligned against.

While it initially appears that changes have occurred, with the order of regions effectively remaining the same, with region 1 (yellow) running onto region 2 (blue) and then region 3 (green), region 2 was however inverted relative to the parental reference. This displays one of the key limitations of the alignment tool: the inability to align around a set point of origin. Rather it uses the start of each sequence as the starting point. The parent strain was phenotypically identical to its sequenced progenitor so the structural changes may have been unexpected.

To verify that the parent strain was genetically indicative of the ATCC 8486 colony used to create the SBRC ATCC 8486 reference strain, the oldest stock of *E. limosum* in the lab was taken and sequenced. This stock was a subculture of the original ATCC 8486 stock that was acquired from DMZ and was expected to be chronologically similar to that of the reference colony. Both Oxford Nanopore and Illumina data were generated and aligned against the SBRC ATCC 8486 reference genome using the

whole-genome alignment plugin (Figure 14). As was expected for a colony of cells closely related to the reference sequence, no genomic rearrangements were apparent.

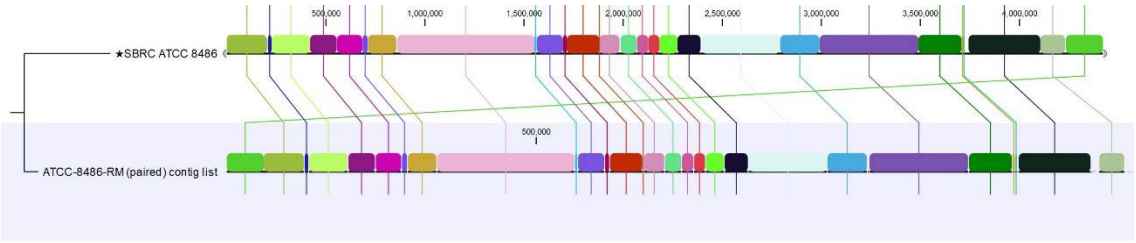

Figure 14 - Output of whole genome alignment between daughter colony ATCC 8486 and SBRC ATCC 8486

The ATCC 8486 daughter colony (on the bottom) aligned against the SBRC ATCC 8486 (on top). The blocks of colour represent different regions and serve to distinguish regions which may have moved. No changes occurred, the apparent repositing of the far right most block of the SBRC ATCC 8486 to the far left of the ATCC 8486 daughter colony is mostly likely due to a difference in the start of the de novo sequence and would not account for a rearrangement once the circular nature of the chromosome is accounted for.

3.3.7.1 Adaptation Repeats Genomic Rearrangements Relative to Parent ATCC 8486

For all repeats of adaptive fermentation on CO (3.3.1), ASF01, ASF02 and ASF03, genomic alignments were performed as previously described (3.3.7). Figures for these alignments are in the appendix (8.2). While it initially appeared that genomic rearrangements had occurred (Figure 103, Figure 104, Figure 105), the changes in genomic order could be entirely accounted for by the differences in the start of the *de novo* sequences produced by CLC genomics. Once the circular nature of the genome was accounted for, no change occurred.

3.3.7.2 Alterative Assembly and Analysis Algorithms

While methylation was investigated next chronologically (3.3.8), alterative assembly and analysis algorithms were also investigated to try to see if the issues of the CLC genomics approach could be resolved. These issues primarily include an inability to compare from a set start point and gaps in comparison. The raw nanopore data was processed from scratch using Flye *de novo* assembler for single-molecule sequencing reads (289) via the open-source, web-based platform Galaxy (290). This resulted in the generation of complete circular genomes. Genomic comparison was performed using the Mauve genomic comparison tool (291) within the software sandbox

Geneious Prime 2022 (https://www.geneious.com). Comparison was performed with all three adapted *E. limosum* ATCC 8486 strains' *de novo* genomes against that of the parent *E. limosum* ATCC 8486 (Figure 15, Figure 16, Figure 17).

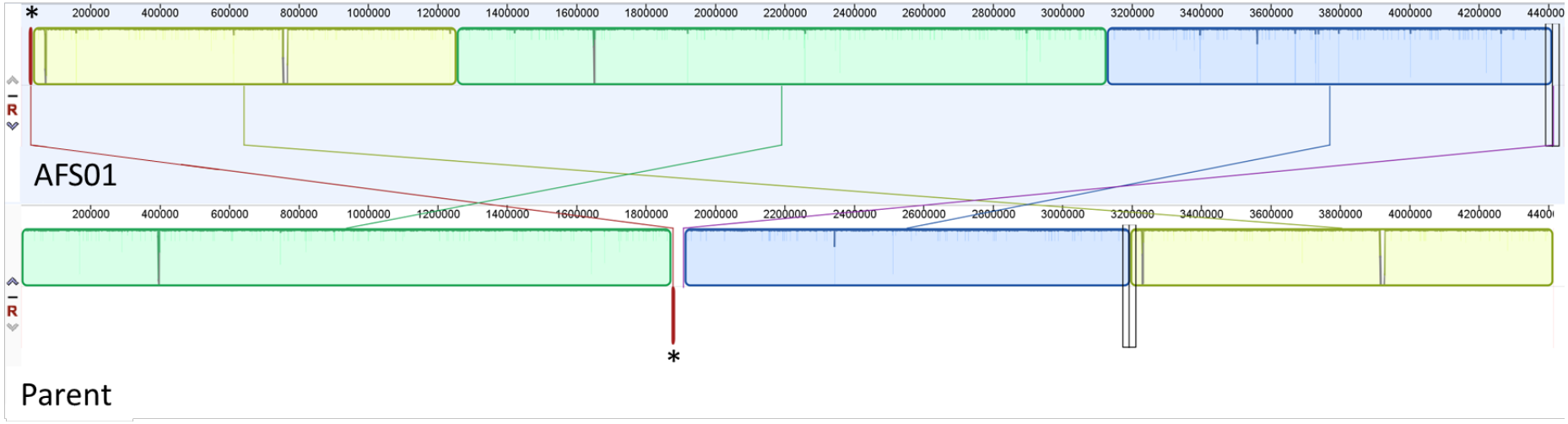

Figure 15 - Mauve genomic alignment of ASF01 ATCC 8486 and Parent ATCC 8486

The ATCC 8486 ASF01 (on top) aligned against the parent ATCC 8486 (on the bottom). The blocks of colour represent different regions and serve to distinguish regions which may have moved. These major regions do not move relative to each other due to the circular nature of the complete circularised genome. The minor region in red (*) does appear to move.

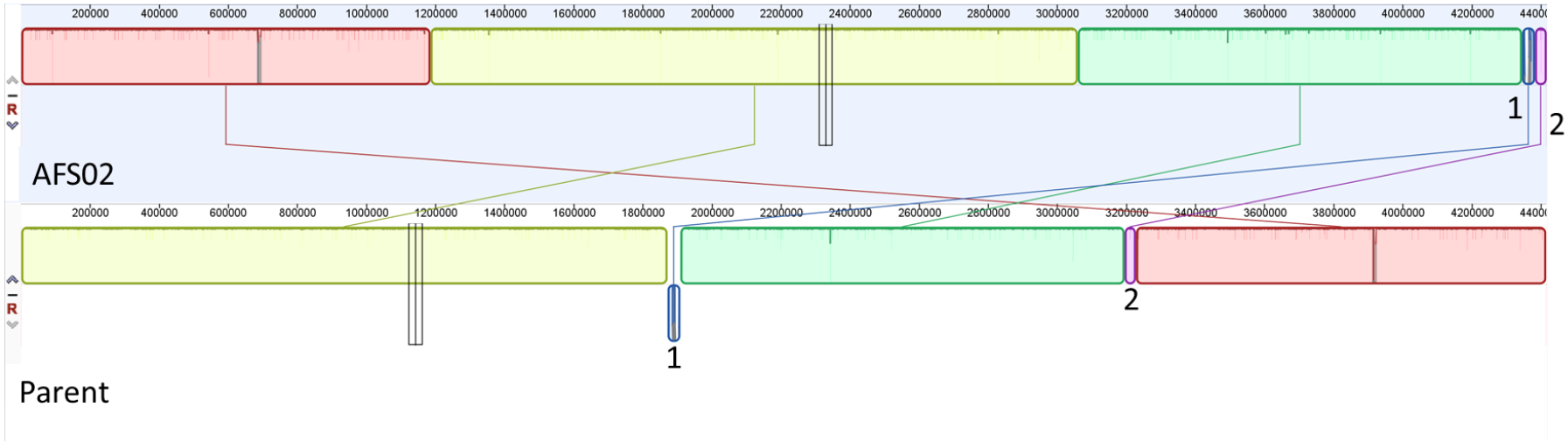

Figure 16 - Mauve genomic alignment of ASF02 ATCC 8486 and Parent ATCC 8486

The ATCC 8486 ASF02 (on top) aligned against the parent ATCC 8486 (on the bottom). The blocks of colour represent different regions and serve to distinguish regions which may have moved. These major regions do not move relative to each other due to the circular nature of the complete circularised genome. The minor regions in blue (1) and purple (2), however do appear to move.

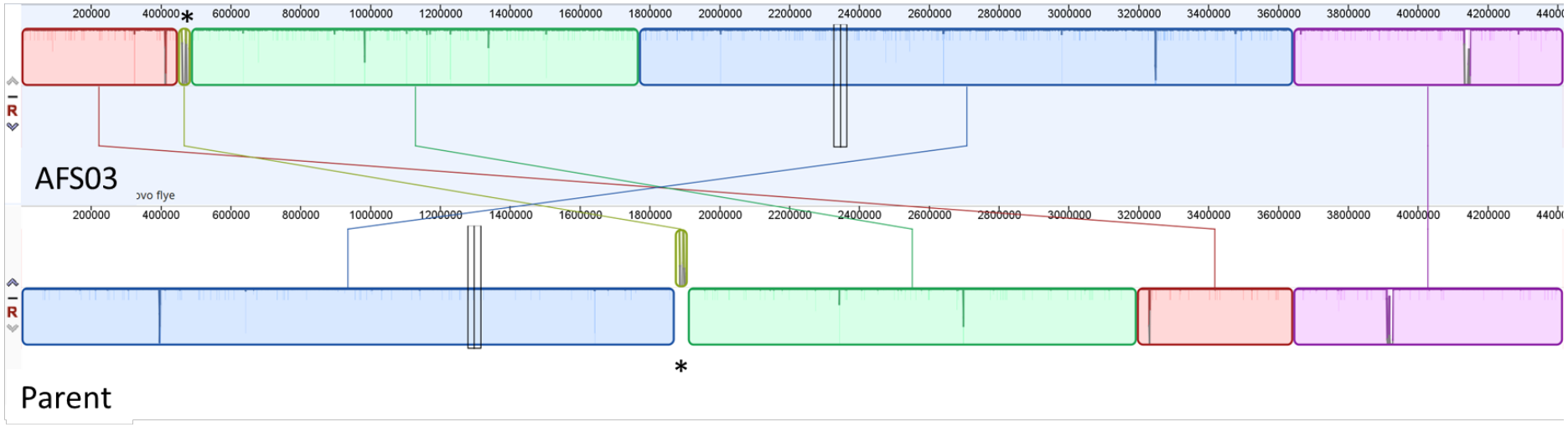

Figure 17 - Mauve genomic alignment of ASF03 ATCC 8486 and Parent ATCC 8486

The ATCC 8486 ASF03 (on top) aligned against the parent ATCC 8486 (on the bottom). The blocks of colour represent different regions and serve to distinguish regions which may have moved. These major regions do not move relative to each other due to the circular nature of the complete circularised genome. The minor region in yellow (*), between regions blue and green on the parent, at between 1,800,000 and 2,000,000 bases on the parent however does appear to move.

All alignments showed movement of a region located on the parent between 1,800,000 and 2,000,000 bases. The sizes of the region varied by the adapted strain, with the longest being ~50 kb and the shortest being ~12 kb. If this movement were to have any connection to the observed phenotype, then it is likely that the smaller 12 kb, which moved from the parent position 1873248 – 1885522, region would contain the genes responsible. As such, this short sequence, and flanking genes, was taken and analysed (Table 27). The region contained 18 genes none of which were overtly linked to the phenotype with many hypothetical proteins and some definitive phage-associated genes. The rearrangement also involved an inversion of orientation, with the 5' end of the reference sequence becoming the 3' end in the adapted strains. The rearrangement did not appear to disturb any surrounding genes by truncation or fusion, meaning that if this region had any effect on the observed phenotype, it is likely due to transcriptional level variance due to the altered location.

The lack of any significant structural variation and the lack of any apparent coding change difference between the adapted cell lines and the parent cell line, which could explain the clear phenotypical differences between the cell lines, suggested that some other form of change had occurred in the genome. Besides a coding change, an epigenetic change was the next logical avenue to investigate.

3.3.8 Methylation data

The methylation status of a codon is a form of epigenetic regulation in prokaryotic organisms (292). This epigenetic regulation is heritable to daughter cells, allowing an altered phenotype to be conveyed across generations without an altered genotype (293,294). This may mean that the methylation status of genes, which may affect transcription and expression of the gene, is responsible for the adapted phenotype observed with the independent CO adaptation replicates, where a consistent phenotype was observed but no clear genetic cause identified.

Upon comparing the methylation patterns of the parent strain and the CO adapted cell lines, potential regions in the methylome that might have changed due to adaptive fermentation were examined. These data could allow genes within regions to be

investigated further, and their function compared with homologues in other organisms. While it was likely that methylation patterns varied between adapted cell lines and the parent, it is also probable that patterns would vary between the differing adapted cells. Furthermore, it is also probable that different methylation pattern changes may produce a similar phenotypical effect. Further investigation was needed to fully understand the relationship between the methylation pattern and the altered adhesive and yeast extract requirement phenotypes. Data processing was done as a service at the DeepSeq facility in Nottingham, wherein a report was produced containing the methylation frequency (Table 19) between the parent strain, the ATCC 8486 reference strain, and the adapted strains. The methylation profile figure (Figure 18) displays the frequency of methylation, as in how many of the residues within a given motif are methylated, throughout the genome. The vast majority of sites are not methylated or at such a low rate as to be below the threshold for reliable determination as indicated by the error bars.

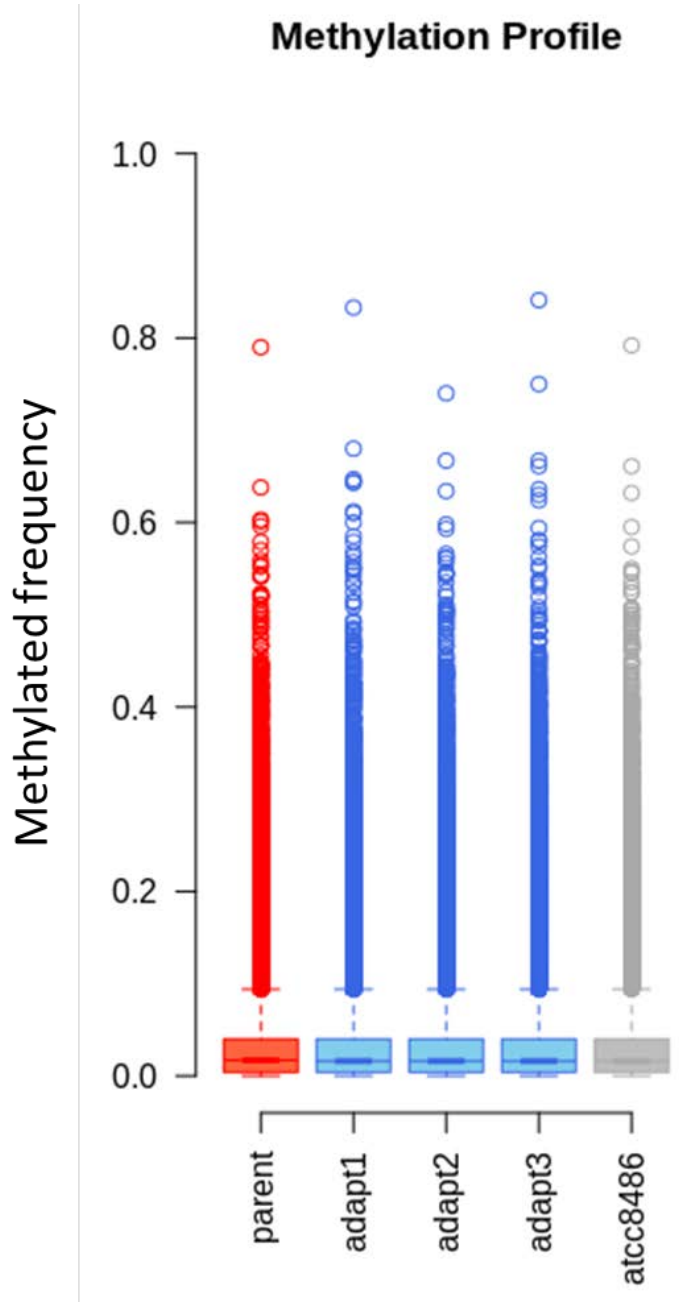

Figure 18 - Graph displaying methylation patterns between adapted, parent and reference strain E. limosum ATCC 8486

Displays relative methylation status of sites of different strains. adapt1, ASF01 ; adapt2, ASF02; adapt3, ASF03. Circles represent regions of the chromosome; most methylated regions are between 10% and 40% methylated. A smaller number of regions are methylated about 50%. Variation occurs between all strains. The bottom two bars include error bars and the vast majority of regions which are either unmethylated or methylated in a small percentage of called sites.

Table 19 - Comparative methylation frequency between strains.

Sample	Total Sites	Unmethylated sites	Ambiguous sites	Methylation frequency		
				> 0.1	> 0.3	> 0.5
Parent	131,228	119,985	11,243	6,833	279	21
Adapted 1	131,227	119,368	11,859	7,074	290	23
Adapted 2	131,227	119,218	12,009	7,350	308	23
Adapted 3	131,226	119,131	12,095	7,366	312	21
Reference ATCC 8486	131,228	119,591	11,637	7,077	280	16

As with the comparisons made so far, the differences between the parent and adapted strains were of interest. As such, all methylated genes that were not unique to the adapted strains were discarded, leaving only those regions that may contain genes of interest (Table 20).

Table 20 - Genes uniquely methylated in adapted strains.

Start point	End point	KAIST annotation	Product	Adapted Strain(s)
639457	639457	WP_038352514.1	betaine/carnitine transporter, BCCT family	ASF02
759981	759981	WP_013382081.1	ABC-2 type transport system ATP-binding protein	ASF01, ASF02, ASF03
1416792	1416792	WP_167617846.1	EAL domain, c-di-GMP-specific phosphodiesterase class I (or its enzymatically inactive variant)	ASF01, ASF02
2084006	2084006	WP_038352002.1	PTS system D-fructose-specific IIA component (F1P-forming), Frc family /PTS system D-fructose-specific	ASF02
2468741	2468741	WP_081571245.1	Acyl-CoA dehydrogenase	ASF02
2689419	2689419	WP_038353834.1	hypothetical protein	ASF03
3270790	3270790	WP_038351499.1	hypothetical protein	ASF01
3722046	3722046	WP_038354248.1	hypothetical protein	ASF01 ASF02

Demethylation may play just as important a role as methylation. Accordingly, a comparison of genes which were not methylated in the adapted strains was also undertaken (Table 21).

Table 21 - Genes uniquely demethylated in adapted strains.

Start point	End point	KAIST annotation	Product	Adapted Strain(s)
658120	658120	B2M23_RS06170/ WP_013382320.1	Na+-translocating ferredoxin:NAD+ oxidoreductase RNF, RnfC subunit	ASF01, ASF03
1786506	1786506	WP_081571206.1	hypothetical protein	ASF02
2338759	2338759	WP_038350658.1	ATP-binding cassette, subfamily B	ASF01, ASF03
2567970	2567970	WP_038353932.1	acetyl-CoA carboxylase, biotin carboxylase subunit	ASF01, ASF02, ASF03
2725020	2725020	WP_052237444.1	PAS domain S-box-containing protein/diguanylate cyclase (GGDEF) domain-containing protein	ASF01, ASF02
2737540	2737540	WP_052237443.1	menaquinone-dependent protoporphyrinogen oxidase	ASF01, ASF02
2909364	2909364	WP_038352382.1	Helix-turn-helix domain-containing protein	ASF03
3319257	3319257	WP_038351461.1	dihydropyrimidinase	ASF01
4139936	4139936	WP_167617727.1	hypothetical protein	ASF01, ASF02, ASF03
4255082	4255082	WP_038353190.1	ATP-binding cassette, subfamily B	ASF03

Further, changes in the degree of methylation of genes may also play a role in altering phenotype (Table 22).

Table 22 - Change in methylated frequency between parent and adapted strain

Start point	End point	KAIST annotation	Product	Adapted strain(s)	Change in degree of methylation
13523	13523	WP_038353321.1	DUF1638 domain-containing protein	ASF01	0.042
13523	13523	WP_038353321.1	DUF1638 domain-containing protein	ASF02	-0.045
13523	13523	WP_038353321.1	DUF1638 domain-containing protein	ASF03	-0.002
658120	658120	B2M23_RS06170/ WP_013382320.1	Na+-translocating ferredoxin:NAD+	ASF02	0.008

			oxidoreductase RNF, RnfC subunit		
1523875	1523875	WP_038352596.1	Lectin like domain-containing protein	ASF01	0.048
1523875	1523875	WP_038352596.1	Lectin like domain-containing protein	ASF02	0.002
1523875	1523875	WP_038352596.1	Lectin like domain-containing protein	ASF03	-0.046
1647448	1647448	WP_038353585.1	PTS sugar transporter subunit IIA	ASF01	0.038
1647448	1647448	WP_038353585.1	PTS sugar transporter subunit IIA	ASF02	-0.016
1647448	1647448	WP_038353585.1	PTS sugar transporter subunit IIA	ASF03	0.019
2338759	2338759	WP_038350658.1	ATP-binding cassette, subfamily B	ASF02	-0.037
2725020	2725020	WP_052237444.1	PAS domain S-box-containing protein/ diguanylate cyclase (GGDEF) domain-containing protein	ASF03	0.003
2737540	2737540	WP_052237443.1	menaquinone-dependent protoporphyrinogen oxidase	ASF03	0.015
2836173	2836173	WP_052237238.1	Adenine-specific DNA methylase, N12 class	ASF01	0.049
2836173	2836173	WP_052237238.1	Adenine-specific DNA methylase, N12 class	ASF02	0.001
2836173	2836173	WP_052237238.1	Adenine-specific DNA methylase, N12 class	ASF03	0.011
2909364	2909364	WP_038352382.1	Helix-turn-helix domain-containing protein	ASF01	0.099
2909364	2909364	WP_038352382.1	Helix-turn-helix domain-containing protein	ASF02	0.123
3292648	3292648	WP_038351484.1	V/A-type H+-transporting ATPase subunit I	ASF01	0.028
3292648	3292648	WP_038351484.1	V/A-type H+-transporting ATPase subunit I	ASF02	0.043
3292648	3292648	WP_038351484.1	V/A-type H+-transporting ATPase subunit I	ASF03	0.057
3294071	3294071	WP_038351482.1	hypothetical protein	ASF01	-0.011
3294071	3294071	WP_038351482.1	hypothetical protein	ASF02	0.025
3294071	3294071	WP_038351482.1	hypothetical protein	ASF03	0.042

3319257	3319257	WP_038351461.1	dihydropyrimidinase	ASF02	0.003
3319257	3319257	WP_038351461.1	dihydropyrimidinase	ASF03	0.088
3753999	3753999	WP_110060263.1	Host cell surface-exposed lipoprotein	ASF01	-0.07
3753999	3753999	WP_110060263.1	Host cell surface-exposed lipoprotein	ASF02	-0.058
3753999	3753999	WP_110060263.1	Host cell surface-exposed lipoprotein	ASF03	-0.069
4000767	4000767	WP_038351169.1	two-component system, OmpR family, sensor kinase Ihk	ASF01	0.043
4000767	4000767	WP_038351169.1	two-component system, OmpR family, sensor kinase Ihk	ASF02	-0.05
4000767	4000767	WP_038351169.1	two-component system, OmpR family, sensor kinase Ihk	ASF03	0.051
4255082	4255082	WP_038353190.1	ATP-binding cassette, subfamily B	ASF01	-0.036
4255082	4255082	WP_038353190.1	ATP-binding cassette, subfamily B	ASF02	-0.054
4407570	4407570	rpsR/WP_013382824.1	30S ribosomal protein S18	ASF01	0.033
4407570	4407570	rpsR/WP_013382824.1	30S ribosomal protein S18	ASF02	-0.047
4407570	4407570	rpsR/WP_013382824.1	30S ribosomal protein S18	ASF03	0.072

The hypothetical proteins were discarded from further analysis because, while none of them were identified as phage proteins, their lack of characterised homologues and short protein length make it is unlikely that they are linked to the phenotype. Several genes of interest were initially identified and investigated in silico.

The ABC-2 type transport system ATP-binding protein, WP_013382081.1, appeared to have an increased degree of methylation in all three of the separately adapted cell lines (Table 20). BLASTp revealed a high degree of conservation among the *Eubacterium* genus. It is possible that this gene is involved in the transport of the exopolysaccharide to the cell wall (295–297). Given the lack of coding change within the produced protein, it is likely that if this gene were related at all to the observed phenotype, decreased expression is likely the cause. If the transporter responsible for allowing the produced adhesive polysaccharide to exit the cell membrane to allow for presentation is not produced, then said polysaccharide will not be presented. The acetyl-CoA carboxylase, biotin carboxylase subunit, WP_038353932.1, was

demethylated in all adapted strains. However, its link to the observed phenotype observed is unclear. Due to time constraints on the project, and these data not being received until the weeks before the thesis submission deadline, laboratory investigation of neither of these genes was performed.

3.4 Discussion

Directed evolution through adaptive fermentation successfully created improved strains for both laboratory work and any industrial application of *E. limosum*. Adapted cells displayed two unique phenotypical changes from the parent strain: a significant diminishment of the adhesive properties of cells, and the ability to grow in liquid and on solid growth medium without the need for yeast extract supplementation. While beyond the scope of this work, it seems prudent to engineer organisms in a manner that will make them useful to other future researchers and to industry, if possible. Yeast extract composition is not defined, and there are variations between batches and manufacturers, which can cause inconsistencies between fermentations (273). The elimination of yeast extract as an essential medium component not only made medium cheaper but allowed for much finer control of the medium. With the use of a fully defined medium, nitrogen and carbon balancing can be performed to maximise the cost efficiency of growth medium for commercial application. The lack of excessive adhesive properties of adapted strains made laboratory handling easier, and it would make the maintenance of fermenters easier by no longer forming biofilm-like structures lining surfaces such as that of the baffles. The ability to grow on CO made for a more robust organism, able to feed on multiple feedstocks with a combination of gas and methanol feed being possible. Data gained from initial continuous fermentation resulted in improved protocols for working with *E. limosum*, with CO being able to be introduced immediately to freshly inoculated cells and a range of reliable feed rates for methanol/CO/CO_2 feeds being discovered, at around 100 to 200 mL of growth per day per litre of reactor volume, or ~ 10 – 20% volume per day.

Interesting results were noted, most significantly in the third repeated fermentation (3.3.4.2), where only a minority of the screened population displayed the reduced adhesiveness phenotype. This suggests that this adaptation is somehow advantageous

to the reactor conditions and outcompetes the wild type, rather than it being essential for growth on CO. This may indicate that the altered adhesive phenotype arises because of continuous fermentation alone (which aligns with the initial adaptation that occurred without CO), and CO adaptation is responsible for the negation of the requirement to supplement with yeast extract. This apparent link between CO adaptation and the alteration of yeast extract requirement by adapted cells suggests a link between CO and yeast extract in terms of cell biochemistry. CO is involved in both the methyl and carbonyl branches of the Wood-Ljungdahl pathway, so it is possible that the components of yeast extract that are essential are involved in these pathways, although given how wide a net that casts it is not overtly clear as to what the link is.

Adapted strains were confirmed as *E. limosum* with 16S rRNA gene sequencing of single colonies and both Illumina and Oxford Nanopore genomic sequencing. The phenotypes of adapted strains were retained across many generations, with no reversion of phenotype to parent ATCC 8486 observed even under any tested conditions, including glucose-containing medium supplemented with yeast extract. Genetic analysis was hoped to ascertain the causes of these changes as they had not been determined for the NG-6849 strain when it was created (Professor Philippe Soucaille, INSA, University of Toulouse, personal communication). However, SNP and indel analysis displayed no mutations that could explain the observed phenotypes. Structural variations can cause significant variations within a genome beyond a single base change (298). Structural differences between the parent ATCC 8486 and the ATCC 8486 used for the creation of the SBRC *E. limosum* (117) reference sequence were noted. This suggests that *E. limosum* has a degree of plasticity regarding its genetic arrangement. This is likely an advantage for a species as it would allow for a minority of a population to contain more potentially useful mutations than otherwise; this hypothesis is supported by the general literature (299). Of note, when whole-genome alignments of adapted strains are performed against the ATCC 8486 reference sequence, there is even more variation than with the parent strain, which further reinforces this hypothesis. This also highlights, however, that genetic rearrangements do not necessarily lead to radically different phenotypes. All the non-CO adapted ATCC 8486 displayed the same phenotype as regarding their adhesive

properties and their need for yeast extract to be present in growth medium despite the apparent structural differences.

Regarding sequence changes resulting from alternation in structure, it does not appear that any of the structural variations observed between the parent strain and the adapted strains can explain the changes observed. This was perhaps unsurprising, as *E. limosum* contains the native pathways to utilise CO (108,280), and negating the need for yeast extract supplementation would likely require entire amino acid synthesis pathways. This suggested that the genes to allow for this phenotype were already present in *E. limosum* and supported the hypothesis that the changes observed may have been due to a change in the epigenome rather than the genome.

A similar study was conducted by the KAIST group (286), where they identified changes in the CO dehydrogenase/acetyl-CoA synthase (CODH/ACS) complex coding genes, *acsA*, and *cooC*. However, despite using ATCC 8486 as the parent strain, similar mutations were not observed as part of this study. This may have been due to the different fermentation conditions used, which may select for different mutations. The KAIST study used a different medium (modified DSMZ 135 medium), and syngas was used (286) as opposed to a N_2:CO feed. Syngas contains a higher concentration of CO than was used in this study, and contains H_2 which acts as an additional electron donor (300). This was done as their work aimed to increase CO fixation as compared to this study which aimed to understand how adaptive fermentation led to the observed phenotypical changes. In discussion with the KAIST group (Dr. Yoseb Song, Massachusetts Institute of Technology, personal communication) it was revealed that the same phenotypical changes observed in this study, such as the lack of adhesive properties and no longer requiring medium supplementation with yeast extract, had resulted from their fermentation, and similarly no genetic cause for it was identified.

Genetic regulation is of considerable importance in prokaryotes and is a significant means of phenotypical control (301,302). If a gene is not transcribed or transcribed at too low a level, then the phenotype of the host will be unaffected. Epigenetic changes, heritable non-coding changes that affect the displayed phenotype of an organism

(303), have been well established in eukaryotes as being of significant importance to phenotype, yet in prokaryotes, it is generally poorly understood (304). DNA methylation can act as a form of epigenetic genetic regulation in prokaryotic organisms (292,304). If methylation patterns change during adaptive fermentation, then the phenotype of the cell could be altered. Methylome analysis was performed on the Oxford Nanopore data, and regions of altered methylation were identified. Although some genes may warrant further analysis, there were no genes which stood out as overtly involved in the phenotype observed.

In a collaborative study between INSA and KAIST, both world leaders for *E. limosum* research, similar phenotypes to those observed here were obtained in *E. limosum* fermentations on CO (Professor Philippe Soucaille, INSA, University of Toulouse, personal communication; Professor Byung-Kwan Cho, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea , personal communication). However, no SNPs or genetic changes that could explain these changes were identified. The most likely explanation is that such changes were absent, therefore investigation shifted towards heritable changes that were non-coding. Of note, work was conducted after the conclusion of this study, which suggested the presence of an epigenetic methylation system in *E. limosum*, supporting the epigenetic change hypothesis further (305).

While the genetic or epigenetic causes of these adaptations had not been elucidated, it was clear that these adaptations were advantageous, with the ability to grow without the supplementation of amino acids, as occurs when medium contains yeast extract, and the lack of cellular adhesive properties being selected for under CO adaptive fermentation conditions.

3.4 Future work

The adaptive fermentations created cell lines with phenotypes distinct from the parent ATCC 8486. This phenotype was like that of the NG-6849 strain, with cells no longer requiring yeast extract supplementation in medium to grow and not having significant adhesive properties.

The genes of interest identified by methylation pattern analysis include those encoding for an ABC-2 type transport system ATP-binding protein and acetyl-CoA carboxylase biotin carboxylase subunit. Initial data (Table 20) suggested that these genes where methylated in the adapted strain but unmethylated in the parent strain. As the effect of methylation on transcription in *E. limosum* is unknown, knockouts may provide useful information. If methylation was involved with increased transcription of the gene, then a deletion of the gene in an adapted background should restore the phenotype to that of the parent, at least with respect to either adhesion or yeast extract requirement. If, however, if the methylation state was involved in a decrease in synthesis, then a deletion of the gene in the parent strain should create a phenotype more like that of the adapted strains. While it is unclear how these genes are related to the observed phenotype, using them to test for how, if at all, methylation status affects expression in *E. limosum* would be valuable research. These genes may have been better served, therefore, as a target for transcription analysis, likely using RT-qPCR (306).

Adaptive evolutionary fermentation is an important means to develop industrially relevant organisms. While traditionally a slow approach, without the restrictions placed upon other industries (for example, with yeast for alcoholic drink production), genetic modification techniques can be used in tandem with fermentation to produce many useful phenotypes for *E. limosum*. Of greatest interest would be performing an adaptive fermentation under a selective pressure of interest, for example CO, using a library of transposon mutants (224) as the starter culture. Such a library would have a large amount of genetic diversity. Selective pressure during the fermentation would select for the genotypes that produced the phenotype best suited for the conditions.

This would allow for the identification of genes contributing to fitness under these conditions, by sequencing the resultant strains.

Process-level optimisations are essential for the viability of *E. limosum* as an industrial strain. This would be done with the goal of maximising product formation by either increasing biomass or by increasing the amount of product produced in a given time. This would involve improving the growth medium further, optimising the fermentation protocol, and setting up and developing improved downstream processing methods. Downstream processing methods may include the removal of products from the fermentation medium during fermentation. This would be useful as many products are growth limiting or result in a change in the growth state (307–309). Furthermore, this optimisation may, as shown for the NG-6849 strain relative to the wild-type ATCC 8486, lead to the production of an improved strain without the use of genetic modification tools such as CRISPR-Cas9. Some process-level work has been done by other researchers in *E. limosum*, using methanol in combination with formate to drive the production of butanol (310). Of some interest is the concept of co-culturing, i.e., growing two organisms together to improve a characteristic of fermentation processes. While some preliminary research has provided poor results from the point of view of *E. limosum* (311), the study's goals may not be aligned with industrially related goals. Therefore, the concept may be worth further investigation, especially as is related to the production of some non-native products, which may be growth-linked in another organism but not in *E. limosum*.

While the defined Pacaud medium (254) used as part of this work allowed for the growth of both the NG-6849 and ATCC 8486 strains, with the former and CO adapted ATCC 8486 mutants able to grow in that defined state, it is likely that the medium could be further developed and improved. It is possible that there is an excess of some vitamins or of some other macronutrients or minerals. While on a laboratory scale, these are likely to have a minimal effect, from an industrial perspective, having a medium which is as minimal and affordable as possible is desirable. Decreasing the medium cost increases profitability and, therefore, is essential to the viability of *E. limosum* as an industrial organism. Furthermore it may be possible to optimise medium for different products; for instance, *E. limosum* natively produces vitamin

B12 (312), and it may be possible to modify Pacaud medium to design a new medium to maximise native production of B12.

It is possible that methylation epigenetics are used by the organism to adapt to differing environmental conditions. However, questions arise as regards the stability of epigenetically linked phenotypes and the heritability of said phenotypes over many generations without the selective pressure which caused them. This is of interest to this work because long after CO adaptive fermentation occurred, adapted cells appeared to retain the altered phenotype despite no longer being in fermenter conditions. Based on the retention of phenotype in adapted *E. limosum*, it is likely that the status of the methylome only changes under selective pressure if there is no selective disadvantage conferred by an existing methylation pattern. This being the case would offer many advantages to the organism, primarily that pathways can be switched off in conditions where they are no longer needed, potentially allowing for increased biomass under preferential conditions. A better understanding of the system of epigenetics in *E. limosum* would not only shed light on the cause of the phenotypes created by CO adaptive fermentation but also potentially allow for novel genetic modification strategies or a better understanding of regulation within the transcriptome, both of which would be useful in the further development of *E. limosum*. Work has been conducted in *E. coli* whereby an epigenetic memory system was developed and allowed for set phenotypical alterations to occur in response to environmental stimuli (313). If there was a better understanding of the *E. limosum* epigenome and such a system could be developed for *E. limosum* a strain which is able to alter its phenotype based upon its feed stock could be developed which would confer a significant advantage over the current *E. limosum* strains available for industrial use. A system of epigenetic memory also would confer significant advantage to an organism like *E. limosum* if, as has been described elsewhere in the literature (314), it was present natively. This is because *E. limosum*, being present in animal gut rumen (315), is natively exposed to both the anaerobic conditions of the host gut but upon defecation is exposed to a more oxygen rich environment where, for instance, an extracellular polysaccharide matrix may provide some protection from aerobic conditions (316). It is worth highlighting that while the hypothesis of epigenetic memory is interesting there is presently no evidence to support its

existence in *E. limosum* beyond conjecture. The simplest explanation for the apparent stability of the altered epigenetic state is that of there being no selective disadvantage that pressures cells towards the wildtype epigenome. This is likely due to these cells being primarily maintained on growth medium, which is deficient in yeast extract, in effect selecting against the wild type.

A combined approach, utilising both directed genetic modification and adaptive fermentation, would likely produce improvements in terms of C1 substrate utilisation, which is of interest for real-world industrial applications. Native products are likely to be linked to growth, with the greater the biomass the greater the production of native products. However, introduced pathways may not be linked to growth, especially where reactions are ATP consuming. While this is also a problem in aerobic bacteria (317) it is especially an issue in acetogenic bacteria which have very limited free energy (121). Addition of ATP to cells via supplementation in the growth medium is not tenable due to both the cost of ATP and its instability. The use of pathways which are ATP neutral, or coupled to ATP regeneration (318) is therefore preferable over those which are ATP negative. Work to maximise the production of said pathway products would require much more in-depth work than for those which were growth-linked. It may also be possible to introduce additional pathways for the sole purpose of ATP regeneration (319), this would offset the cost of introduced pathways making them viable.

3.5 Chapter conclusions

Several independent adapted strains were produced as part of this work, with improved phenotypes, including no longer requiring yeast extract-supplemented growth medium and decreased cellular adhesive properties. The lack of significant extracellular adhesive properties improved its lab handling significantly, with it being comparable to that of *E. coli*. Further its lack of requirement for yeast extract supplementation, allowed for the use of a fully defined medium, that would enable the use of genetic modification systems which use auxotrophic phenotype as a counter-selection. While the genetic causes of the altered phenotype were not fully elucidated, data suggested that the observed changes may be due to regulatory changes. These regulatory changes may be caused by heritable epigenetics, an interesting research query for future work.

Chapter 4

Genetic Modification of *E. limosum* using CRISPR-Cas9 and a Type 1 toxin-antitoxin system

4.1 Introduction

To engineer *E. limosum* to produce commodity products, it is necessary to genetically modify the organism. For this purpose, genetic modification technologies which can create stable, specific changes to the genome are needed. Genetic techniques can generally be categorised as either forward or reverse techniques (320). Forward techniques attempt to find the genotypical conditions that cause a particular phenotype, while reverse techniques attempt to find the phenotypical effects of a genetic sequence. They can be summarised as forward, being from phenotype to gene, and reverse, from gene to phenotype.

Transposon mutagenesis (152) is an example of reverse genetics. It produces many random mutants, and cell survival indicates that a mutation does not result in a non-viable cell line, making it useful in identifying of essential genes or gene regions. This can allow for the creation of mutants with desirable phenotypes resulting from previously unknown genetics, leading to a better understanding of the genotype-to-phenotype relationship, and directing further research. Mutated genes that cause unexpected phenotypes may be worth investigating further.

CRISPR-Cas9 can be used as both a forward and a reverse technique (320). Genes can be knocked out to observe the phenotype. Cell lines with altered phenotype can have their genetic changes assessed by next-generation sequencing. Identified mutations can be replicated in different genetic backgrounds using CRISPR-Cas9 to determine the genetic change causing the phenotypic change.

To produce non-native products, genes must be introduced into the target organism. This can be achieved by either introducing them into the organism's chromosome or by utilising a stable plasmid that can be maintained in cells across generations without the need for antibiotic selection. To effectively produce non-native products, it may also be necessary to knockout genes associated with competing pathways or which make the production of the target molecule energetically unfavourable, especially in the low energy context of acetogenic bacteria (75,121). Ideally products would be, growth coupled such that product formation is linked to biomass, as acetate already is

for acetogens. This would aid in the production of more biomass, which would lead to the production of more product. Furthermore, to determine the function of specific genes, genes deletion and analysis of the resulting phenotype can be highly beneficial. To accomplish this, a robust genetic modification toolset is required.

Given its advantages over traditional counter-selection markers, CRISPR-Cas9 was explored as the primary tool for identifying and isolating allelic exchange mutants. Initially, attention focused on CRISPR-Cas9 from *Streptococcus pyogenes* (SpCas9). Two such systems had previously been developed at the SBRC, using either a Cas9 nickase mutant derivative (251) or a RiboCas variant (218) in which the expression of *cas9* was tightly controlled through a theophylline responsive riboswitch (1.5.3.3.1). A different CRISPR-Cas9 system has been used at the Korea Advanced Institute of Science and Technology (KAIST) (315). However, the KAIST approach did not result in scarless mutants as it relied upon the integration of an antibiotic selection marker into the position of the knockout (KO) or knock-in (KI), hence, this approach was not taken.

Parallel work within the BioMetChem project by Dr. James Millard (University of Nottingham) explored the feasibility of exploiting an endogenous CRISPR-Cas3 system present in *E. limosum.* The use of endogenous systems had several potential advantages. As Cas genes do not need to be incorporated into an introduced vector, vectors can be made smaller and therefore more efficiently transformed. Smaller vectors also make multiplexing more feasible and allow for introducing larger payloads for genomic integration. An endogenous system should work efficiently (284) with little need for optimisation of promoters. Accordingly, the potential of the system was explored here.

In addition to the various Cas-based counter-selection systems, an alternative system based on a type 1 toxin-antitoxin system from *Eubacterium callanderi* (formerly *E. limosum* KIST612) as a counter-selection marker was explored. This suicide vector-based system had been developed by Dr. James Millard (University of Nottingham) and appeared highly effective in generating knockout and knock-in mutants (227). The NG-6849 strain was used as the background for all the genetic modification

work. This was primarily due to its availability. The adapted strain developed as part of this work was not fully isolated until around 3 years into this project, which would not have provided sufficient time to complete the work of this chapter.

4.2 Aims

The aims of this chapter were twofold. First, to identity and establish a robust genetic modification system for use in *E. limosum*, with particular emphasis on generating clean in-frame mutants without the need for integrating antibiotic resistance markers into the genome. The ideal system should allow for the creation of multiple mutant backgrounds. While CRISPR-Cas9, given its prior success in *E. limosum* and extensive presence in the general literature, was the primary candidate, other counter-selection systems, notably toxin-antitoxin-based systems, were also investigated.

Once a suitable genetic modification system was identified, knockouts were created using this system to study the function of genes likely to be involved in methanol metabolism and in the production of butyrate (Figure 19). Specifically, the genes *cop, metF1, metVmetF, mti,* and *mtiI* for methanol metabolism and *act, actpbt2, actpbt1, actpbt1pbt2,* and *bcact* for butyrate production. These genes were homologues of genes in other anaerobic bacterium (primarily *A. woodii*), and it was expected that they would have a similar function in *E. limosum*.

The investigation of methanol metabolism in *E. limosum* was of significant interest, as the organism's ability to utilise methanol played a pivotal role in its selection for research over other currently industrialised acetogens. Butyrate was seen as a potential competitive product to any which may be introduced due to the significant carbon it would consume. The creation of auxotrophic backgrounds was also desirable to make use of ACE as a means of insertion of genes into the genome which would be needed for the for the introduction of non-native products.

Beyond the functional analysis of the genes, generating mutants also aimed to validate the effectiveness of the genetic modification systems and identify any potential limitations or challenges they might pose. Because of the limitations associated with

the ATCC 8486 strain, such as excessive adhesive properties, hindering single colony selection, cell resuspension, and electroporation, all directed genetic modifications were conducted using the NG-6849 strain.

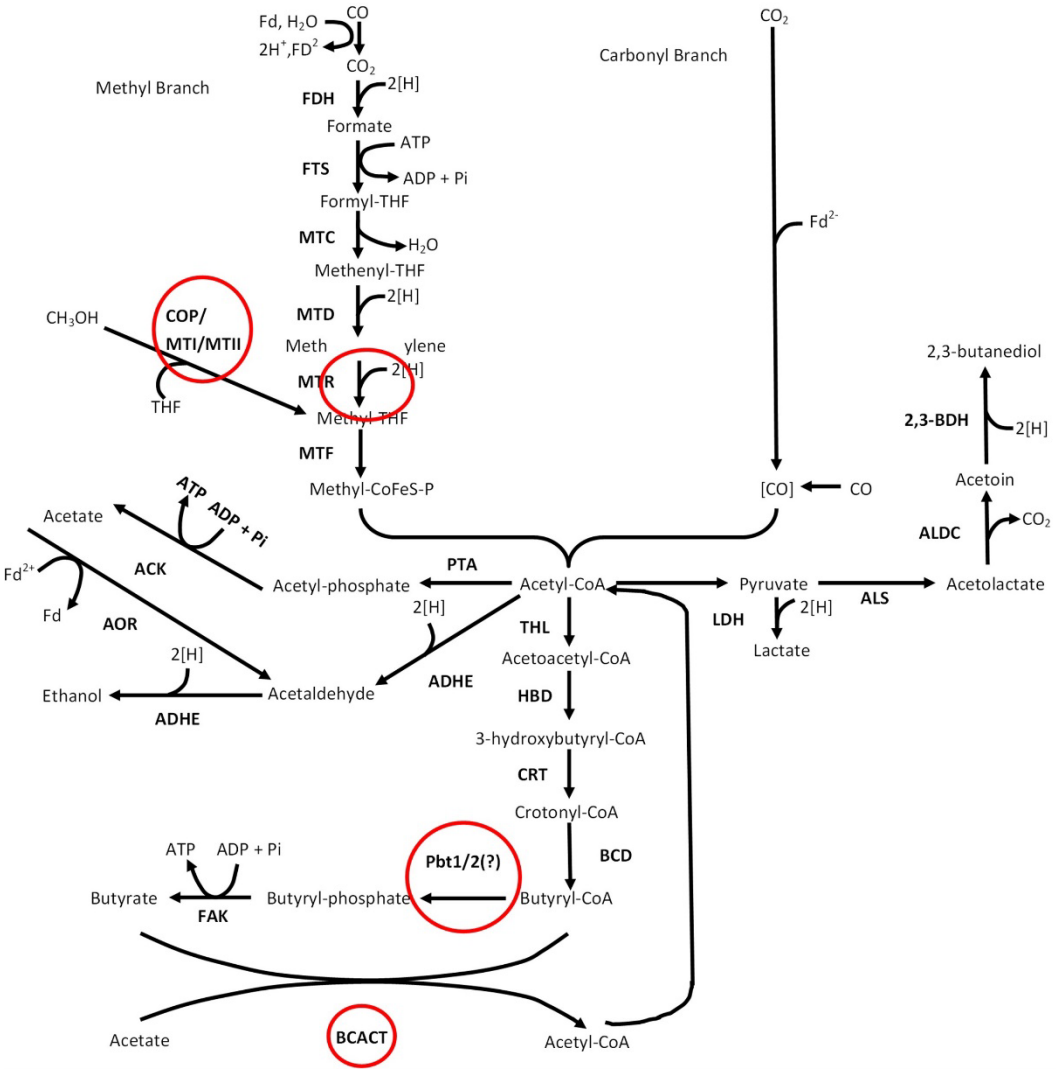

Figure 19 - Wood-Ljungdahl pathway in E. limosum displaying pathway branches/genes of interest

In the red circles are genes/groups of genes identified as homologues for those in other acetogenic bacteria. Their functional homology was to be investigated using the molecular biology tools developed. Additionally, amino acid biosynthesis pathway genes were of interest to generate auxotrophic backgrounds. These pathways are not part of the Wood-Ljungdahl pathway and so not displayed in the figure.

Both CRISPR-Cas9 and Toxin-antitoxin knockout systems were considered as candidates for genetic modifications tools in this work. Due to the prior work done in *E. limosum* and the existence of a purportedly successful scarless CRISPR-Cas9 system this was investigated first.

4.3 Results

4.3.1 Troubleshooting and re-establishing CRISPR-Cas9 in *E. limosum*

Initially, all knockout work was performed using a plasmid based constitutive CRISPR-Cas9 system. This system used the *thl* gene promoter (P_{thl}) from *C. acetobutylicum* to control expression of *cas9* and the *C. acetobutylicum araE* promoter (P_{araE}) controlling single guide RNA (sgRNA) expression. This system had previously been used to make mutants in *E. limosum* ATCC 8486 (117), including deletion of a phosphate butyryltransferase gene, *pbt1* (117), and a gene encoding the restriction subunit of a type I restriction-methylation system, WP_038353978.1, known internally as *orfD*.

This system, however, proved ineffective at creating knockouts in *E. limosum* (data not shown). In one case, *act*, making use of three different sgRNAs against the gene failing to produce a single knockout. Later work would show that this gene was not essential. It was hypothesised that this may have been caused by poor sgRNA expression. Sequencing (2.2.7) revealed the Cas9 gene not to be mutated beyond mutations which had previously been identified by other researchers at SBRC, none of which has prevented others from successfully creating mutants (Dr. James Millard, University of Nottingham, personal communication). Further homology arm and guide sequences seemed to align with the in silico maps indicating correct insertion. The presence of mutations within Cas9 indicated that constitutive expression while hosted in *E. coli* was sufficiently high to select for the mutations, therefore guide expression was seen as the most likely cause of the failures. As effective sgRNA expression is crucial for creating mutants using CRISPR-Cas9 (321) work to remediate the issues centred around increasing sgRNA expression. Previous work (117) indicated that the promoter used for sgRNA expression, P_{araE}, was relatively weak in *E. limosum*. This promoter was therefore exchanged with the higher strength promoters (P_{fdx}) of the ferredoxin gene (*fdx*) of *C. perfringens* and *C. sporogenes*.

Using NEBuilder® HiFi DNA Assembly Master Mix, P_{araE} was replaced with *C. perfringens* P_{fdx} in the existing CRISPR-Cas9-containing vector, pSBRCas9, creating the new vector pSBRCas9-Cpf (2.6.1). Guides to target the *act* gene and homology arms to allow for homologous recombination were designed and inserted into pSBRCas9-Cpf. The guide made use of the PAM NGG and bound approximately 400 bases from the 5' end of the gene and the homology arms were 800 bases in length. The fully assembled vector, now named pSBRCas9-Cpf-*act* (Figure 20), was transformed into *E. limosum* NG-6849 (2.3).

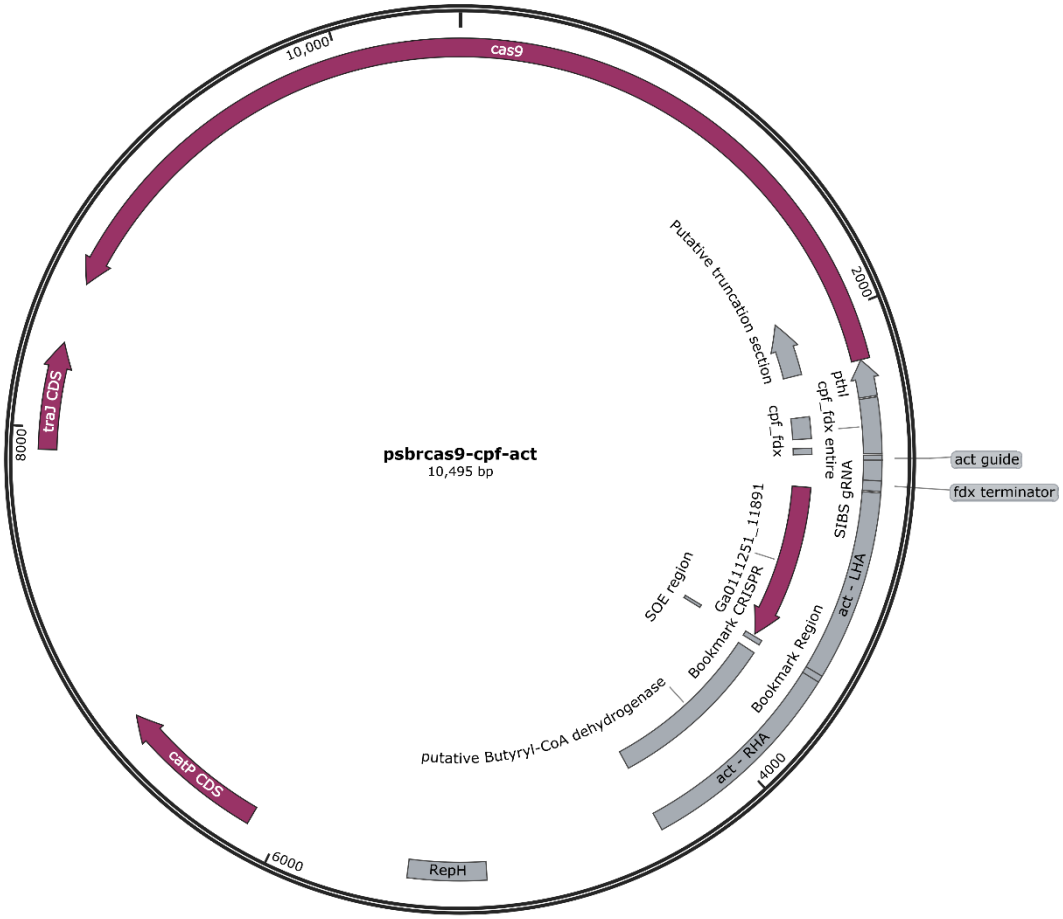

Figure 20 - Map of pSBRCas9-Cpf-act

Transformed cells were grown on plates (2.1.1.2) supplemented with 7.5 µg/mL of thiamphenicol to select for cells which were maintaining the pSBRCas9-Cpf-*act* CRISPR vector. Once colonies arose, they were screened using colony PCR (2.6.6) with the primers act LHA scr F2 and act RHA scr R2. Initial screening revealed no mutant colonies or mixed mutants (Figure 21A). Colonies were re-streaked to single

colonies on antibiotic selection plates (2.1.1.2) and screened with cPCR. The second round of screening identified pure Δ*act* mutants (Figure 21B). These pure mutant colonies were then streaked to single colonies on non-selective plates (2.1.1.2), in an attempt to lose the pSBRCas9-Cpf-*act* plasmid, before the final round of screening to ensure no contamination of the wild type (Figure 21C).

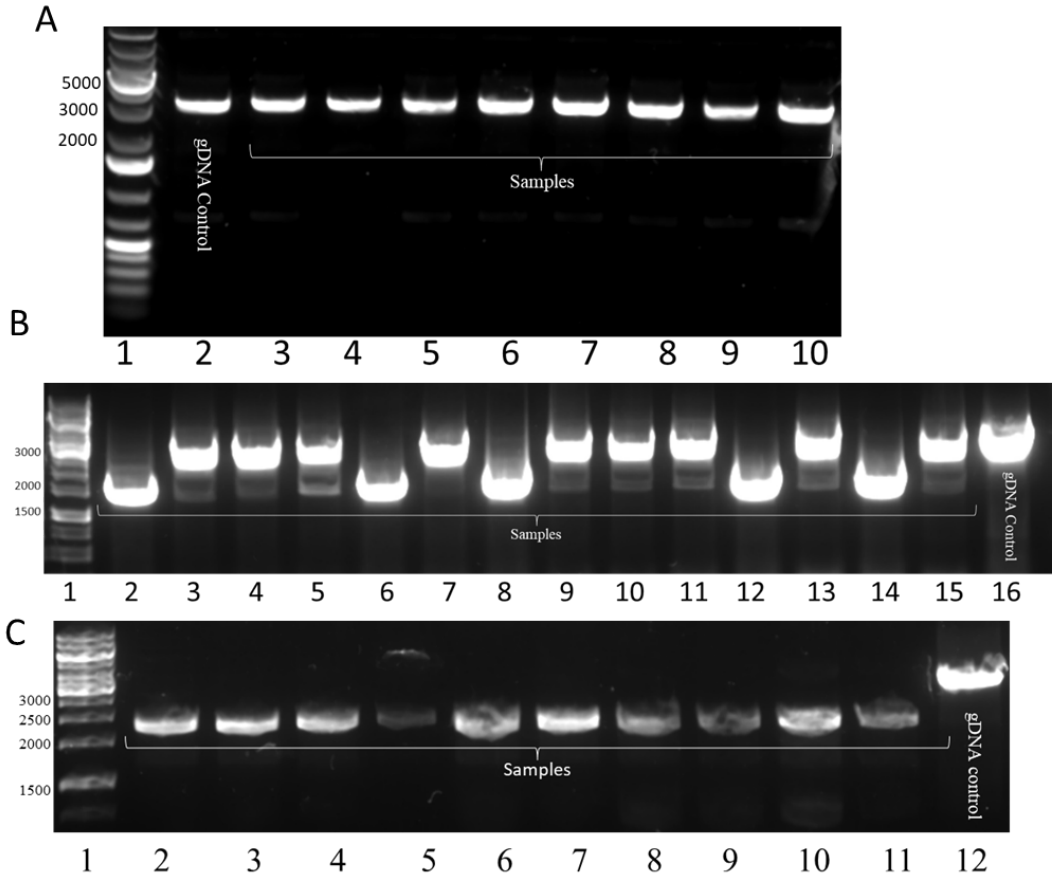

Figure 21 - PCR analysis of act mutant candidates
Three rounds of standard Q5® Colony PCR was performed on colonies of E. limosum NG-6849 transformed with pSBRCas9-Cpf-act. The first round (A) were colonies which had arisen after the initial transformation, the second round (B) are descendants of colonies which had been re-streaked from the first round of screening, with the third round being descendants of the pure mutants identified in gel B . Primers (act LHA scr F2 and act RHA scr R2) were used to screen for a chromosomal knockout of the act gene. The ladder used (in lane 1) for both PCRs was GeneRuler 1 kb Plus DNA Ladder; samples were run on a 1% agarose gel.

A. *Samples were run in lanes 3 - 10. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. All screened colonies displayed a wild-type sized bands (~3500 bases) indicating that no knockout occurred.*
B. *Samples were run in lanes 2 - 15. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 16. Successful pure deletion bands were observed in lanes 2, 6, 8, 12, and 14. None of the 5 lanes had wild type bands indicating pure mutants.*
C. *Samples were run in lanes 2 – 11. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 12. Successful pure deletions were observed in all screened cells. No wild-type bands (~3500 bases) appeared in screened colonies (2 - 11) indicating no mixed genotypes.*

Successful Δ*act* mutants were identified and stocked (2.1.2.5). pSBRCas9-Cpf-*act* was identical to the vector used previously which failed to produce knockouts (data not shown) but for the differing sgRNA promoters. It is highly likely that the relatively stronger nature of the *C. perfringens* P_{fdx} promoter is the reason for the success of this knockout. Further knockouts attempted using this constitutive system were performed using vectors whereby the *C. perfringens* P_{fdx} promoter controlled sgRNA synthesis.

4.3.2 CRISPR-Cas9 knockouts

Following successful inactivation of *act*, multiple knockouts in *E. limosum* were created using CRISPR-Cas9 using the newly created pSBRCas9-Cpf vector (Table 23). Each of the below sections contains the rational, the experiment itself and the results of said work; further characterisation of each mutant is also included.

Table 23 - Genes and encoded enzymes, knocked out using CRISPR-Cas9

Gene Name	**Locus Tag**[1]	**NCBI Accession Number**[2]	**Enzyme Name**	**EC number**
act	Ga0111251_11893	WP_038353018.1	Act; Acetate CoA-transferase	2.8.3.8
cop	Ga0111251_112775	WP_038352387.1	MtaC19; [Methylamine--corrinoid protein] Co-methyltransferase	2.1.1.248
hisI	Ga0111251_11325	WP_013378694.1	*hisI*; Phosphoribosyl-AMP cyclohydrolase	3.5.4.19
metF	Ga0111251_113114	WP_038351520.1	MetF; methylene-THF reductase	1.5.1.20
metFV	Ga0111251_11345/6	WP_081571099.1	MetFV; methylene-THF reductase	1.5.1.20
mti	Ga0111251_112774	WP_038352386.1	Mta18; Methanol-corrinoid protein Co-methyltransferase.	2.1.1.90
mtiI	Ga0111251_112776	WP_038352459.1	MtaA; methyltetrahydrofolate:cobalamin methyltransferase	2.1.1.13
pbt1	Ga0111251_111765	WP_038352195.1	*Pbt1*; Phosphate butyryltransferase	2.3.1.19
pbt2	Ga0111251_11734	WP_052237430.1	*Pbt2*; Phosphate butyryltransferase	2.3.1.19

[1] The " Locus Tag" refers to the locus tags used on the genome that was used as a reference for this work.
[2] NCBI Accession Number refers to a unique identifier that allows for rapid identification of an encoded protein and its function.

4.3.2.1 Generation of *E. limosum* NG-6849 Δ*act* mutants

The first successful knockout was that of the *act* gene (Figure 21), as described above. These cells, which had lost the plasmid (determined by an inability for colonies to grow on thiamphenicol selection) were stocked (2.1.2.5) and used for future work involving the Δ*act* strain. This was done for all CRISPR-Cas9 mutants, with their respective primers and plasmids varying as needed.

Initially, the *act* gene was believed to encode for a butyryl CoA:acetate CoA transferase (WP_038353018.1), an enzyme involved in CoA transfer. However, upon reviewing the data it became clear that due to a misannotation, the gene previously referred to as *act* was putative acyl CoA:acetate/3-ketoacid CoA transferase, which is a different CoA transferase and not the one needed to cause the expected phenotypic change upon deletion. The correct gene was identified later (4.3.4.8). However, phenotype assessment for *act* was performed with the original rationale in mind. The resulting data confirmed that *act* was not a butyryl CoA:acetate CoA transferase as butyrate production still occurs, as discussed in 4.3.2.1.1.

4.3.2.1.1 Characterisation of *E. limosum* NG-6849 Δ*act* mutants

Stocked mutant cells were confirmed by colony PCR using the primer pair *act* LHA scr F2 and *act* RHA scr R2 and used to conduct phenotypic characterisation of the mutant strain. All work was done from single colonies. Cells were grown in triplicate (2.1.3.1), with a wild-type NG-6849 strain control, using glucose and methanol as the carbon sources. Samples for HPLC analysis were taken at 24-hour intervals throughout growth. Originally, characterisation was to be done in yeast extract-deficient medium. However, coordination with BioMetChem partners required the use of a medium with yeast extract present to keep data consistent and comparable. Cultures were therefore grown with 2 g/L yeast extract. Cells were grown in serum bottles under standard conditions (2.1.2.1) with 30 mL of Pacaud medium (2.1.1.2) using 10 g/L bicarbonate. HPLC samples were processed and analysed (2.1.2.4.2). Of interest were butyrate and acetate productions levels, the former to assess the mutant phenotype and the latter as a strong indication of cell fitness; it is important to have

both as a lower butyrate level could be due to lower cell growth or fitness rather than due to the lack of *act*.

4.3.2.1.1.1 Characterisation of *E. limosum* NG-6849 Δ*act* mutants with glucose

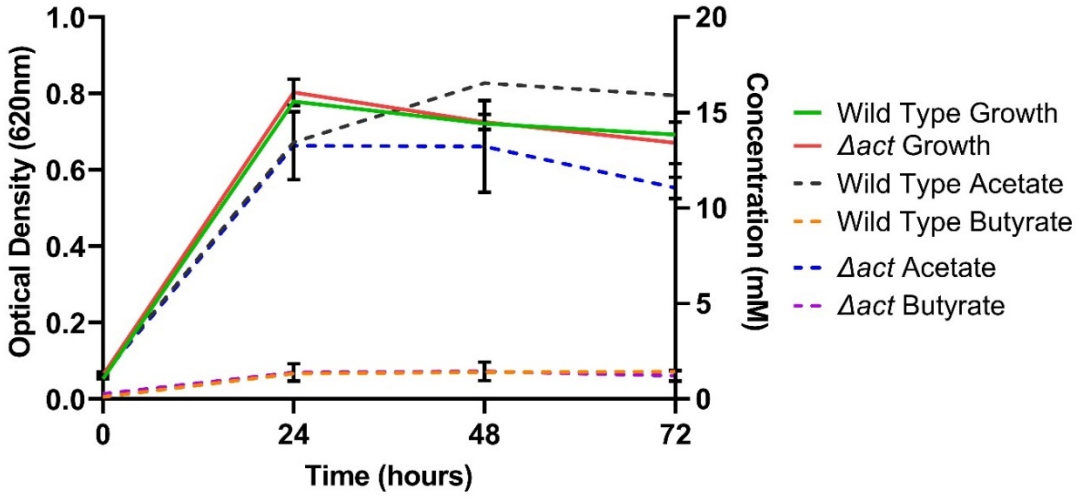

Figure 22 - Acetate and butyrate levels throughout E. limosum Δact growth on glucose

Wild-type and Δact E. limosum NG-6849 strain were grown on 50 mL Pacaud medium with 2 g/L yeast extract in batch culture in serum bottles with 15 mM glucose as the substrate and 200 mL anaerobic gas headspace (~8% hydrogen, ~77% N2, ~15% CO2). Spectrophotometer readings for absorbance at 620 nm and samples for HLPC were taken every 24 hours.

The Δ*act E. limosum* NG-6849 strain and a wild-type control were characterised using 15 mM glucose as the substrate, all other conditions as previously detailed (4.3.2.1.1). Both the wild-type and Δ*act E. limosum* NG-6849 strains grew comparably (Figure 22), reaching maximum ODs of 0.72 and 0.73, respectively, after 48 hours. The levels of acetate and butyrate produced were also comparably similar, with maximal acetate titres of 15.8 mM and 11 mM, respectively, and corresponding butyrate titres of 1.4 mM and 1.2 mM.

4.3.2.1.1.2 Characterisation of *E. limosum* NG-6849 Δ*act* with methanol

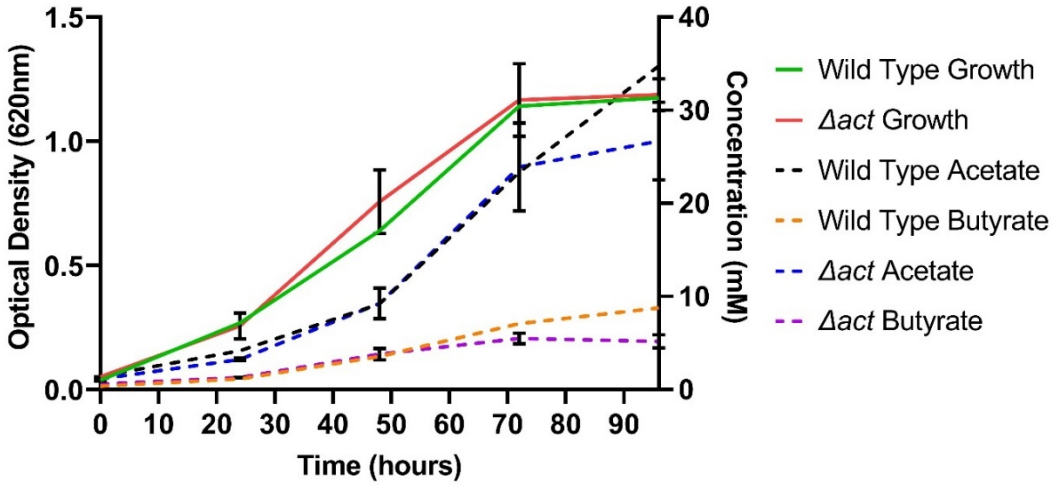

Figure 23 - Acetate and butyrate levels throughout E. limosum growth on methanol

Wild-type and Δact E. limosum NG-6849 strain were grown on grown on 50 mL Pacaud medium with 2 g/L yeast with 100 mM methanol as the substrate and 200 mL anaerobic gas headspace (~8% hydrogen, ~77% N2, ~15% CO2). Spectrophotometer readings for absorbance at 620 nm and samples for HLPC were taken every 24 hours.

The Δ*act E. limosum* NG-6849 strain was characterised using 100 mM methanol as the substrate, under the same conditions previously detailed (4.3.2.1.1). This additional substrate characterisation was performed because methanol was the feedstock of interest for this study. Both the wild-type and Δ*act E. limosum* NG-6849 strains grew similarly (Figure 23), reaching maximum ODs of 1.17 and 1.18, respectively, after 96 hours. Acetate production differed, with the wild-type strain producing 34.7 mM and the Δ*act* mutant strain producing 26.6 mM, and this difference was mirrored in the respective butyrate production levels of 8.7 mM and 5.1 mM.

Although a decrease in butyrate levels was observed at the end of the time course for the Δ*act E. limosum* NG-6849 strain using methanol as a carbon source, butyrate production on glucose was comparable to that of the wild-type strain. This suggests the possibility of an alternate pathway, mechanism, or a functionally similar gene may be present in *E. limosum*. However, acetate production on glucose was decreased in the Δ*act* mutant strain relative to the wild-type strain, but this decrease was far less significant when the strains were grown on methanol. Because butyrate was produced at detectable quantities on both methanol and glucose feedstocks, further

characterisations would only be performed on glucose. Only if butyrate production decreased would further mutants be characterised on different feedstocks.

4.3.2.2 Generation of *E. limosum* NG-6849 Δ*hisI* mutants

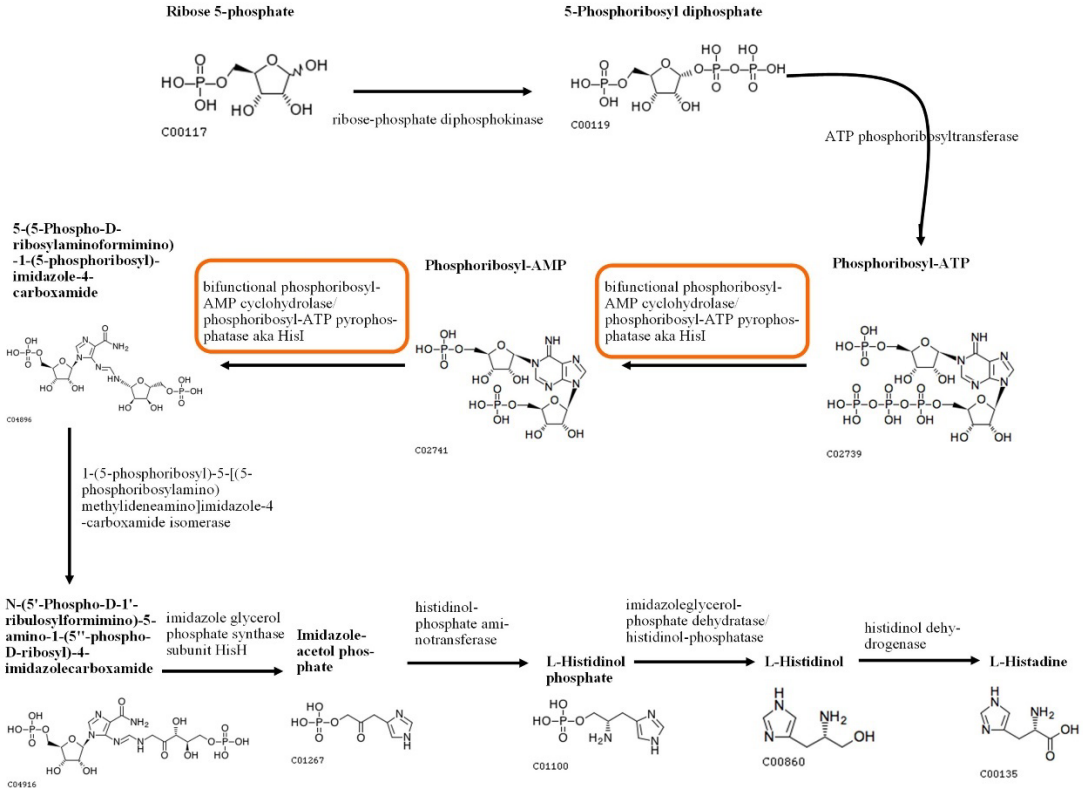

Figure 24 - Part of the histidine synthesis pathway with the protein encoded by hisI highlighted.

Figure displays histidine synthesis from ribose-5-phosphate, the hisI protein is highlighted in the orange boxes. The pathway and structural images are taken from KEGG databases (https://www.genome.jp/kegg/) (322–324).

The *hisI* gene encodes a bifunctional enzyme, phosphoribosyl-AMP cyclohydrolase/ phosphoribosyl-ATP diphosphatase, (WP_013378694.1), which catalyses the third and fourth steps of histidine biosynthesis from ribose 5-phosphate (Figure 24). Deletion of this gene would result in an inability to synthesise histidine. This histidine auxotrophy can in turn be exploited to insert heterologous genes into the chromosome using ACE (2.6.3). In this method, restoration of prototrophy is tied to the integration of a target gene downstream of the gene whose truncation caused the initial auxotrophic phenotype. Selection of successful integrations can be done by growth on histidine-deficient medium. Cells with the initial Δ*hisI* genotype would not survive, while those with *hisI* successfully restored, indicating successful genetic modification, would grow on deficient plates.

The CRISPR-Cas9 knockout vector, pSBRCas9-Cpf-*hisI*, was assembled (2.6.1) using a backbone of pSBRCas9-Cpf, a constitutive CRISPR-Cas9 vector with *C. perfringens* P_{fdx} driving guide RNA production. The plasmids were sequenced (2.4.5) and introduced into the *E. limosum* NG-6849 strain (2.3). After a week on selection plates, a colony PCR (2.5.2.3), using the primer pair hisI LHA scr F and hisI RHA scr R, was performed to assess any genetic changes (Figure 25).

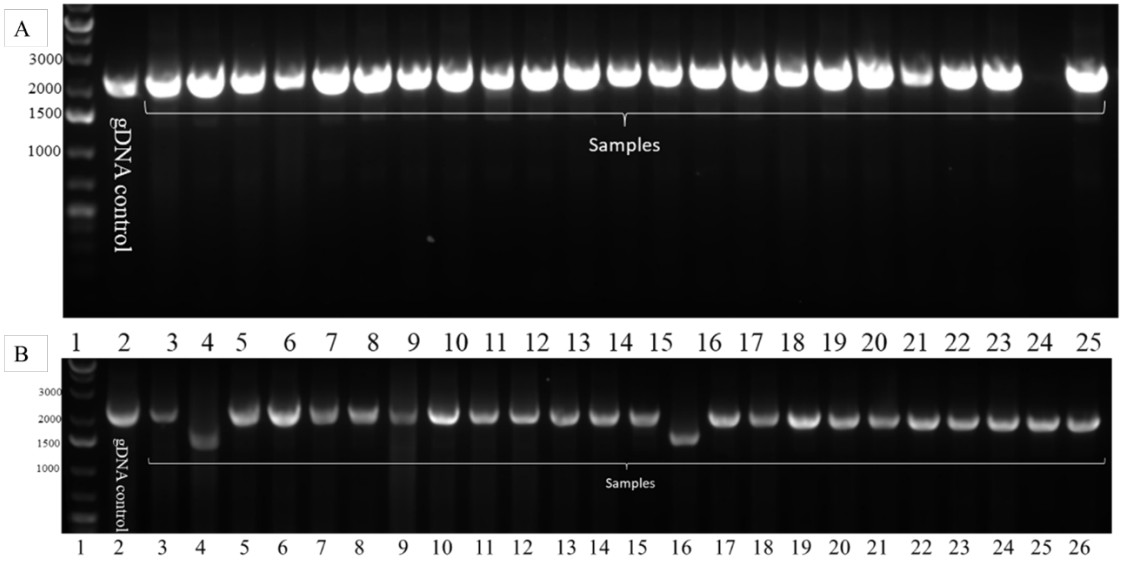

Figure 25 – Screening for E. limosum hisI deletion mutants by colony PCR

A standard DreamTaq colony PCR was performed on E. limosum NG-6849 transformed with pSBRCas9-Cpf-hisI. Primers (hisI LHA scr F and hisI RHA scr R) were used to screen for a chromosomal knockout of the hisI. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb Plus DNA Ladder, samples were run on a 1% agarose gel.

- ***A.*** *23 colonies were screened for a knockout of the hisI (lanes 3 - 25). Wild-type bands (~2200 bases) appeared in screened colonies indicating no mixed genotypes and no pure mutants. Initial colonies did not display any genetic changes after 1 week of growth on selective media. These colonies were re-plated and allowed to grow before being tested again.*
- ***B.*** *24 colonies, which were descendants of re-streaks of the colonies from A, were screened for a knockout of the hisI (lanes 3 - 26). No mixed colonies were observed, however lanes 4 and 16 displayed pure mutant bands, the hisI gene being only ~600 base pairs in size, so a band at ~1500 versus ~2200 bases pairs for the wild type was consistent with a pure mutant.*

Initial screening(Figure 25A) identified no mutants. Isolates were streaked onto solid medium and allowed to grow before being screened again (Figure 25B), identifying colonies displaying a knockout of *hisI*. Interestingly, no mixed colonies were observed at any point in this deletion, suggesting that the guide was more efficient than the *act* guide or that the CRISPR-Cas9 system used is better suited to knocking out smaller genes (*hisI* being ~600 base pairs vs the ~1600 base pair length of *act*). As

with the *act* deletion (4.3.2.1), efficiency was low, with only 2 out of 23 colonies displaying a deletion phenotype, or ~8.5%. Plasmid loss (2.6.7) was more difficult to achieve with Δ*hisI*, relative to Δ*act*, with cells maintaining antibiotic resistance over several rounds of replating. Growth in the presence of the antibiotic was not observed after three rounds of replating, and the mutant clone was stocked (2.1.2.5).

4.3.2.2.1 Characterisation of *E. limosum NG-6849* Δ*hisI* mutants

A Δ*hisI* mutant should be unable to grow on or in histidine-deficient medium, so a phenotype assay was performed where single colonies were split and grown on both histidine-deficient and histidine-supplemented medium for both the wild-type and Δ*hisI E. limosum* NG-6849 strains. Colonies that grew in the absence of histidine should not be Δ*hisI* if the gene annotation was correct. All the colonies of the Δ*hisI E. limosum* NG-6849 displayed the expected phenotype, only growing in the presence of histidine (data not shown). The control of wild-type *E. limosum* NG-6849 strain was able to grow with and without histidine supplementation of the growth medium.

This experiment demonstrated one of the advantages of the *E. limosum* NG-6849 strain over the ATCC 8486 strain of *E. limosum*, which is the ability of the NG-6849 strain to natively grow without yeast extract, allowing for amino acid deficiency phenotype analysis like this to be easily performed.

4.3.2.3 Generation of *E. limosum* NG-6849 Δ*cop* mutants

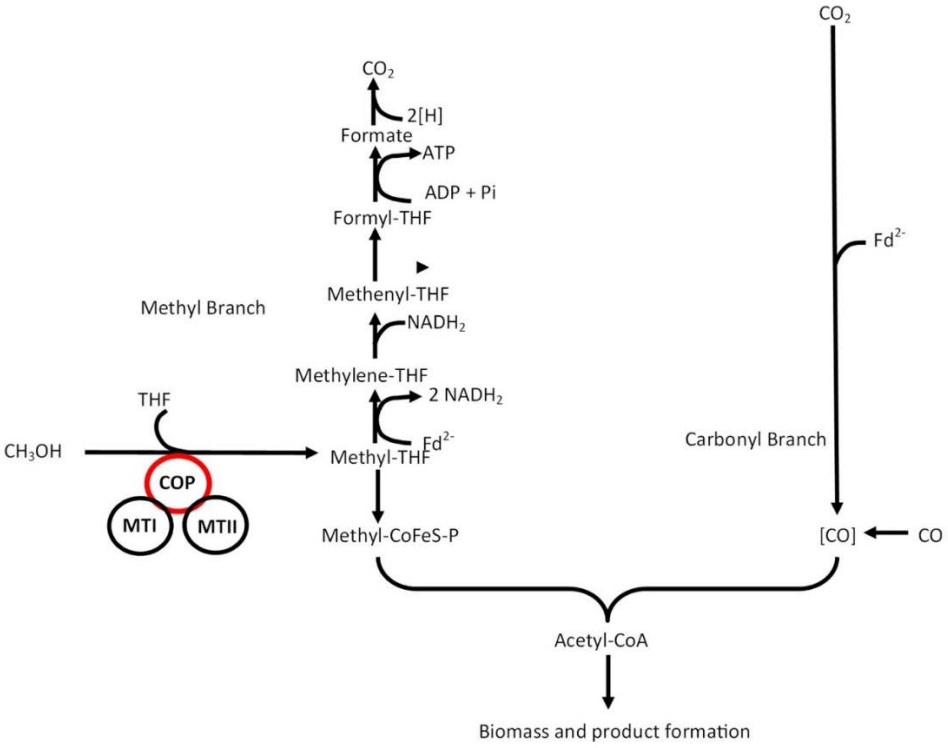

Figure 26 - Hypothetical pathway for methanol utilisation in E. limosum

Enzymes hypothesised to initiate methanol breakdown are displayed in the black and red circles and include the proteins encoded by the cop, mti and mtiI, red indicating the gene to be knocked out in the following section. Based on the hypothetical methanol utilisation pathway proposed in (325).

The gene *cop* is predicted to encode an MtaC19 (WP_038352387.1) like protein, cop, a homologue of the *A. woodii* protein MtaC2. Cop is a MtaC2 like is a corrinoid protein (106,108) that forms a complex with Mta18 like protein mti (encoded by *mti*) and an MtaA like protein mtiI (encoded by *mtiI*) which together act to metabolise methanol (105) by the forming methyl-THF, which is further metabolised (Figure 26). Deletion of this gene (and the other two in this complex) was of interest to the group of Professor Peter Dürre (University of Ulm), collaborators on the Era CoBiotech BioMetChem project under which this study was conducted. Deletion of this gene was predicted to prevent growth on methanol.

This was the first deletion performed using shortened homology arms. These homology arms were roughly half the length of the homology arms used previously (4.3.2.1), which ranged in length from ~1 to 1.5 kb. This was done to increase transformation efficiency, speed of plasmid production, and allowing for much better

Sanger sequencing coverage. Smaller homology arms were thought to be viable based on related work within the SBRC at the University of Nottingham (Dr. Jonathan Baker, University of Nottingham, personal communication) in which homology arms as small as 200 bases long were used in *A. woodii*, an organism which has a degree of relatedness and similarity to *E. limosum* (73). These smaller homology arms were synthesised by IDT (Coralville, IA, USA) in place of being created by PCR (2.4.1.3).

The CRISPR-Cas9 knockout vector, pSBRCas9-Cpf-*cop*, was produced (2.6.1) and a knockout of *cop* attempted as detailed previously (2.6.5.1). Transformed NG-6849 cells were allowed to recover on antibiotic selection plates (2.1.1.2) for a week prior to screening using colony PCR (2.5.2.3) using the primer pair cop External Forward and cop External Reverse. Initial screening identified no mixed or pure colonies (Figure 27A) so cell mass was taken and streaked to single colonies. Colonies from this second plate were screened again (Figure 27B), with three mixed colonies being identified. Mixed colonies were streaked to single colonies once more on selective medium and allowed to recover for a week before rescreening using colony PCR (Figure 28).

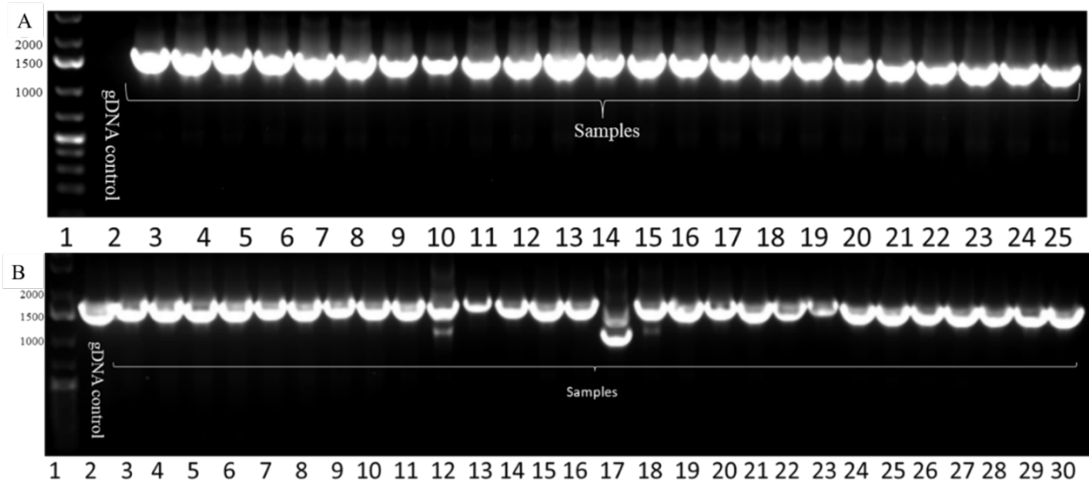

Figure 27 - Screening for E. limosum cop deletion mutants by colony PCR

A standard DreamTaq colony PCR was performed on E. limosum NG-6849 transformed with pSBRCas9-Cpf-cop. Primers (cop External Forward and cop External Reverse) were used to screen for a knockout of cop. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb Plus DNA Ladder, samples were run on a 1% agarose gel.

- ***A.*** *23 colonies were screened for a knockout of cop (lanes 3 - 25). The control failed in lane 2. All colonies displayed a band at the wild-type size of 1 kB. Colonies were streaked onto plates and allowed to grow for an additional week before being screened again via colony PCR.*

B. *27 colonies were screened for a knockout of cop (lanes 3 - 30). Lanes 12, 17 and 18 showed a deletion band at ~600 bases less than that of the wild type, with all 3 being mixed.*

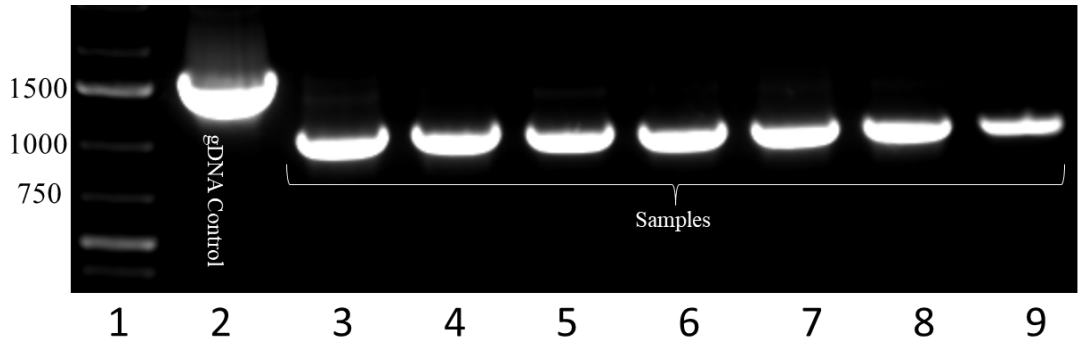

Figure 28 - Rescreening for pure E. limosum cop deletion mutants by colony PCR

A standard DreamTaq colony PCR was performed on 7 colonies of E. limosum NG-6849 transformed with pSBRCas9-Cpf-cop, these cells were previously identified as mixed Δcop mutants. Primers (cop External Forward and cop External Reverse) were used to screen for a chromosomal knockout of the cop (lanes 3 - 9). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb DNA Ladder, samples were run on a 1% agarose gel. Lanes 3, 5 and 7 showed faint wild-type band in addition to the mutant bands, indicating a mixed genotype, these lines were discarded for later growth, all other samples showed mutant bands only.

Rescreening (Figure 28) showed pure chromosomal knockouts of *cop* in all tested colonies. Plasmid loss was then performed on successful colonies (2.6.7). After one week on non-selective plates, colony PCR was performed (Figure 29) using the same primers as with the previous cPCR. Colonies that displayed the desired genotype were stocked (2.1.2.5) and grown on both selective and non-selective plate medium to confirm the loss of the plasmid. After cessation of growth on antibiotic selection plates, colony PCR was done again to test for stability, with successful cells being stocked (2.1.2.5) for phenotype assay. Screening (Figure 29) showed the retention of Δ*cop* mutant genotype in all colonies which had lost the ability to grow under antibiotic selection conditions.

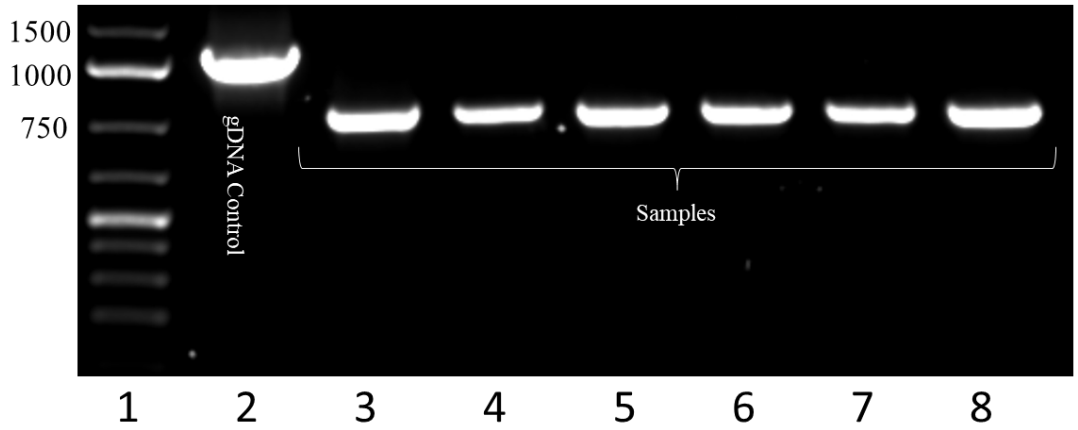

Figure 29 - Rescreening for E. limosum cop deletion mutants by colony PCR after purported plasmid loss.

A standard DreamTaq colony PCR was performed on 6 colonies of E. limosum NG-6849 transformed with pSBRCas9-Cpf-cop, these cells were previously identified as pure Δcop mutants which had purportedly undergone plasmid loss. Primers (cop External Forward and cop External Reverse) were used to screen for a chromosomal knockout of the cop (lanes 3 - 8). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb DNA Ladder, samples were run on a 1% agarose gel. Lanes 3 - 8 show mutant bands, indicating a Δcop genotype.

4.3.2.3.1 Characterisation of *E. limosum* NG-6849 Δ*cop* mutants

A Δ*cop* mutant was hypothesised to be unable or have a perturbed ability to grow on methanol due to the deletion of a corrinoid protein onto which other proteins assemble. Even if the other proteins were present, it would follow that assembly into this complex is needed to process methanol. Therefore, an inability to form said complex would prevent their function, either completely or partially. A phenotype assessment for growth on methanol was performed using single colonies of both Δ*cop* mutants and wild-type *E. limosum* NG-6849 strain. All phenotypic assays were performed using single colonies of *E. limosum* NG-6849 Δ*cop* mutants confirmed using colony PCR with the primers cop External Forward and cop External Reverse (Figure 30).

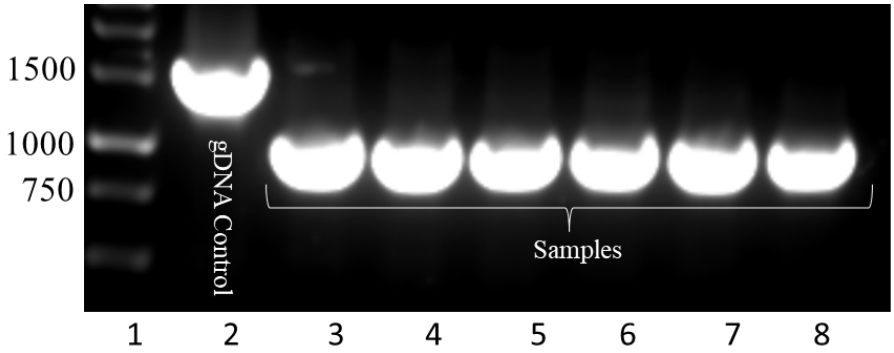

Figure 30 - Screening for E. limosum cop deletion mutants by colony PCR

A standard DreamTaq colony PCR was performed on 6 colonies of E. limosum NG-6849 previously identified as pure Δcop mutants. Primers (cop External Forward and cop External Reverse) were used to screen for a chromosomal knockout of the cop (lanes 3 - 8). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb DNA Ladder, samples were run on a 1% agarose gel. Lanes 3 - 8 show mutant bands, indicating retention of a Δcop genotype. A faint part of a band can be seen in lane 3, however this is likely overflow from the control well.

Starter cultures were grown from single colonies of cells in Pacaud medium (2.1.1.2) with 2 g/L yeast extract and 15 mM glucose. Three replicate cultures were grown, referred to here as Δ*cop* 1, Δ*cop* 2 and Δ*cop* 3. Glucose was used as the carbon source for the starter cultures instead of methanol since Δ*cop* mutants were expected to be unable to grow in the presence of methanol. Due to the risk of glucose carryover, cells were centrifuged to remove the glucose containing medium and were washed using an inert medium without a carbon source. Serum bottles were set up as standard (2.1.2.1) using Pacaud medium (2.1.1.2) with yeast extract present at 2 g/L and methanol at 200 mM.

This initial growth curve (Figure 31) displayed an unexpected phenotype. It was expected that without *cop*, cells could not metabolise methanol. However, after 96 hours, the Δ*cop* 2 mutant culture showed evidence of cell growth through an increase in OD_{620}. This indicated that cells within the population were capable of utilising methanol. Over the measured time, the mutant culture achieved an OD_{620} roughly half that of the wild type after 168 hours. The Δ*cop* 1 mutant also showed some evidence of growth at around this time. This suggested that the gene was still present. Accordingly, the cells from 168 hours of growth were screened internally and externally using colony PCR with the primers cop External Forward and cop External Reverse and the primers cop internal Forward and cop internal Reverse (Figure 32).

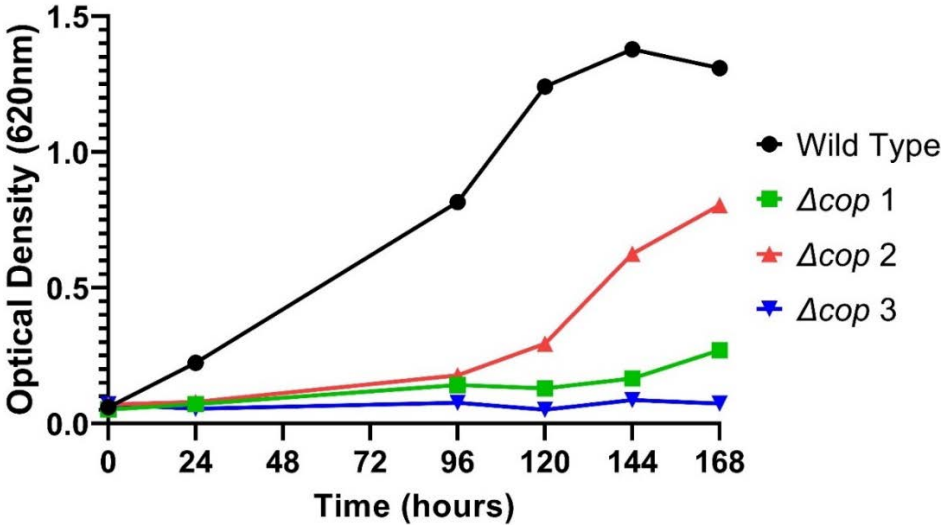

Figure 31 - Growth of Δcop and wild-type E. limosum NG-6849 strain with methanol as carbon source.

Cells were grown in serum bottles were set up using Pacaud medium with yeast extract present at 2 g/L and methanol at 200 mM. Optical density was measured at 620 nm, every 24 hours except at 48 and 72 hours. The wild type was grown as a single culture, Δcop 1, Δcop 2 and Δcop 3 are replicates from a single generated mutant culture.

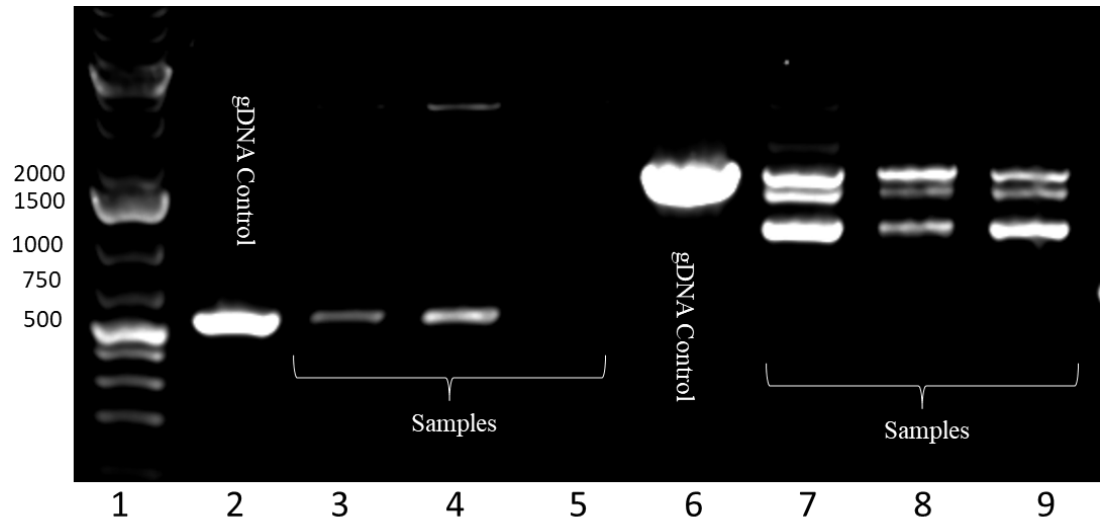

Figure 32 - Screen of putative E. limosum Δcop mutants using primers binding within and outside of cop.

A standard DreamTaq colony PCR was performed on 1µL of the 168-hour end points of the cultures shown in Figure 31, E. limosum NG-6849 Δcop and its corresponding wild-type grown on methanol for 168 hours. Primers (cop internal Forward and cop internal Reverse) binding within the coding region were used to screen for a chromosomal knockout of the cop (lanes 3 - 5). Primers (cop External Forward and cop External Reverse) which bound outside the coding region were used to screen for a chromosomal knockout of the cop (lanes 7 - 9). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2 and 6. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 1% agarose gel. All internal PCRs aside from one showed the gene present, with lane 5 being anomalous (as lane 9 is from the same colony), external primers showed a mixed genotype, suggesting the presence of native genotype cells.

Screening (Figure 32) showed a return to a mixed genotype, indicating that a portion of the population had returned to a native genotype. This could be explained in several ways. The simplest explanation was the presence of a wild-type contaminant, however, colony PCR (Figure 30) performed on the colonies used for this growth assay showed no wild-type banding. As seen previously (Figure 27), the wild-type product is produced even in cases of a mixed genotype background, so it is likely that any wild-type contamination would show some banding. The overexposure used when imaging the agarose gel was intentional to try to find any faint wild-type bands that were present and still none were present.

It is also possible that a fully functional *cop* gene is present on a retained pSBRCas9-Cpf-cop knockout plasmid and is being expressed from the plasmid. If such a plasmid was present in the majority of cells, a long lag in methanol growth would not be expected, as the copy number of *cop* would likely be higher than chromosomal. If, however, said plasmid was only in a minority of cells, then perhaps the long lag phase could be accounted for. This may also explain the sharp increase in growth, as more of the population contains the plasmid. As only plasmid-maintaining cells replicate, a point would be reached where the doubling time, for the population, relative to itself, would reach that of wild-type *E. limosum*.

This could explain the apparent lack of antibiotic growth. If only a tiny minority of cells maintain the plasmid, then the number of viable colony forming units may have been too low to grow visible colonies in the timescales investigated. If almost all cells die, then insufficient mass may remain to allow for the creation of easily observed colonies. Indeed, living cells may have been present but too small to be readily observed without microscopy. However, this does not explain the colony PCR results observed (Figure 32), with primers that bind outside of the native *cop* region producing wild-type bands, indicating the presence of a gene of the size of *cop* in that region.

Furthermore, based upon the plasmid structure of pSBRCas9-Cpf-*cop*, there should not be ability for an exchanged wild-type gene to be transcribed from the plasmid. The presence of stop codons and terminator sequences in the plasmid should prevent

any effect of the *C. perfringens* P_{fdx} promoter on any integrated genes, and no other promoter exists on the plasmid that could allow for transcription. This is compounded by the fact that if the *C. perfringens* P_{fdx} were able to the drive synthesis of the exchanged wild-type *cop* gene, then a guide must also be produced. This guide would target the *cop* located on the plasmid and result in Cas9 creating double-stranded breaks, destroying the plasmid. This means that gene expression from the plasmid is extremely unlikely and would require significant mutation in the guide sequence or Cas9 to be possible.

It is possible, based on the colony PCR (Figure 32), that genomic integration of *cop* is occurring somehow, or a low level of contamination with wild type cells occurred. This would align with both the colony PCR data (Figure 32) and the growth curve data (Figure 31). This aside it is also possible that another gene is able to rescue the phenotype forming methyl-THF and allowing survival, perhaps this gene is either less efficient or is normally expressed at a low level, both of which would explain the lag phase observed. It is also possible that the selective pressure resulted in selection for cells which expressed this gene more highly or those which carried a mutant form of the gene perhaps with a degree of site promiscuity or some other mutation which would result in methyl-THF formation as observed in chapter 3 *E. limosum* is able to change its phenotype under selective pressure.

If genomic integration was occurring in this fashion, then it is likely to occur only in a minority of the population, as it seems from plate analysis of antibiotic resistance that few (if any) cells possess resistance. This would result in a long lag phase before exponential-like growth due to the small number of initial cells.

4.3.2.4 Generation of *E. limosum* NG-6849 Δ*mti* mutants

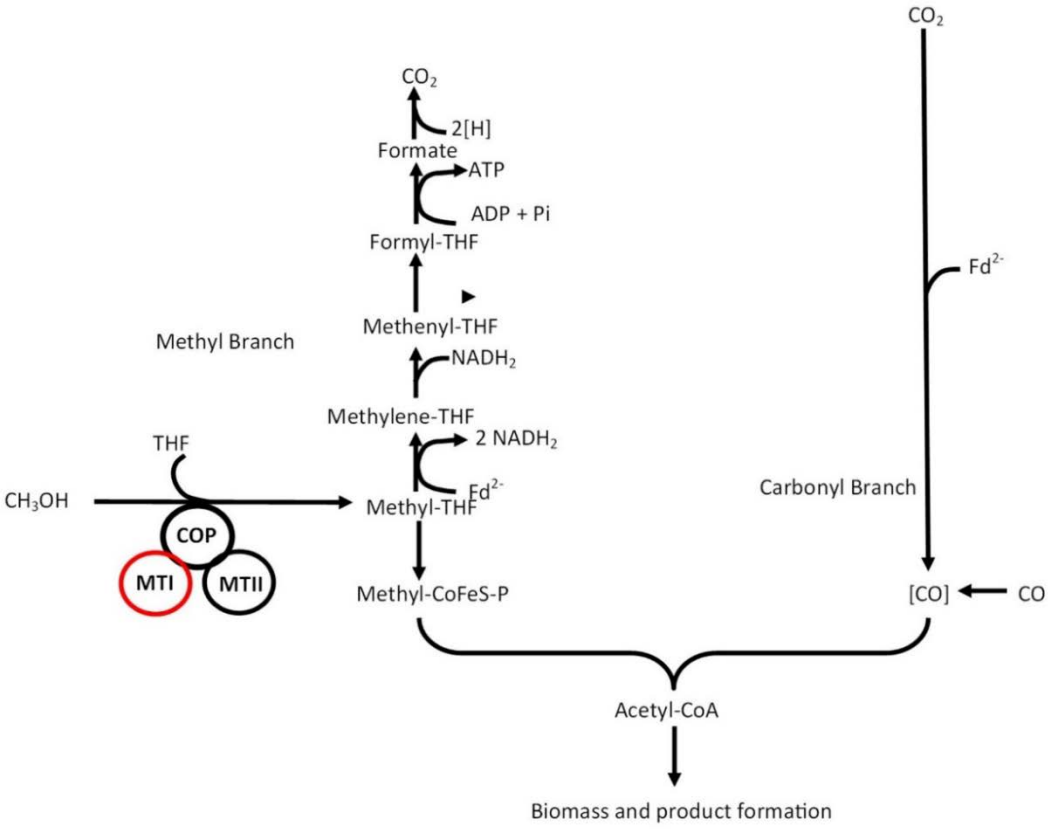

Figure 33 - Hypothetical pathway for methanol utilisation in E. limosum

Enzymes hypothesised to initiate methanol breakdown are displayed in the black and red circles and include the proteins encoded by the cop, mti and mtiI, red indicating the gene to be knocked out in the following section. Based on the hypothetical methanol utilisation pathway proposed in (325).

The *E. limosum mti* gene encodes an Mta18 (WP_038352386.1) like protein, mti, a methanol-coronoid methyltransferase protein and a homologue of the *A. woodii* protein MtaB (106,108). This protein forms a complex with an MtaA like protein, mtiI and a MtaC19 like protein, cop, which together act to metabolise methanol (105) by forming methyl-THF that is metabolised further (Figure 33). Deletion of these genes was of interest to Professor Peter Dürre's group (University of Ulm), collaborators on the Era CoBiotech BioMetChem project, under which the present study was conducted. Deletion of this gene was predicted to prevent growth on methanol.

The CRISPR-Cas9 knockout vector, pSBRCas9-Cpf-*mti*, was created (2.6.1), making use of a synthetically derived guide and homology arm set generated by IDT (Coralville, IA, USA) and using pSBRCas9-Cpf as the backbone, a constitutive CRISPR-Cas9 vector with *C. perfringens* P_{fdx} driving guide RNA production. A knockout of *mti* was attempted as detailed previously (2.6.5.1). Transformed NG-6849 cells were allowed to recover on antibiotic selection plates (2.1.1.2) for a week before screening (Figure 34) using colony PCR (2.5.2.3) with the primers mti External Forward and mti External Reverse.

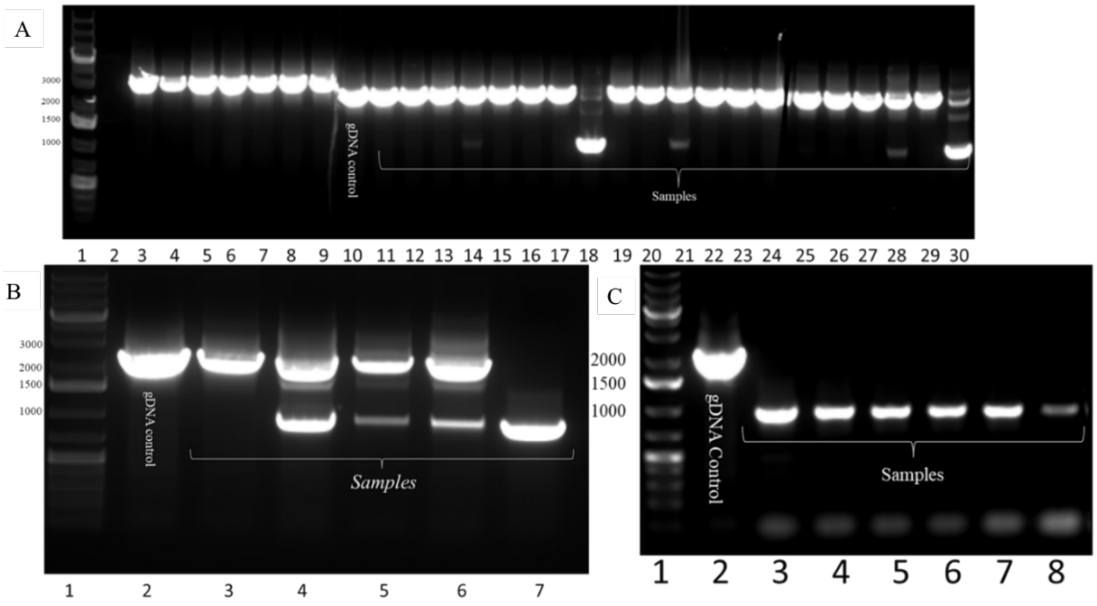

Figure 34 - Screening for E. limosum mti deletion mutants by colony PCR

A standard DreamTaq colony PCR was performed on E. limosum NG-6849 transformed with pSBRCas9-Cpf-mti. Primers (mti External Forward and mti External Reverse) were used to screen for a chromosomal knockout of the mti gene. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2 in images 2 and 3 and in lane 10 in image 1. The ladder used (in lane 1) was GeneRuler 1 kb Plus DNA Ladder, samples were run on a 1% agarose gel.

- ***A.*** *Lanes 2 - 9 are for an unrelated experiment. 19 colonies were screened for a knockout of mti. Deletion bands appeared in lanes 14,18, 21, 28 and 30, wild-type bands were present suggesting a mixed colony genotype. Mixed colonies were streaked to single colonies on selective medium and allowed to recover for a week before rescreening using colony PCR; a single colony from each plate was taken for assessment (Figure 34B).*
- ***B.*** *The 5 mixed colonies from Figure 34A were rescreened for a knockout of mti. Deletion bands appeared in lanes 4, 5, 6, and 7, with wild-type bands present in all but lane 7, suggesting a pure deletion genotype for that colony. Due to the presense of some ghosting, colony 7 was streaked out to single colonies and rescreened (Figure 34C).*
- ***C.*** *6 colonies, purported pure knockouts Figure 34B, were rescreened for a knockout of mti. Deletion bands appeared in lanes 3 - 8.*

Initial colony PCR screening failed to identify any mutants (Figure 34A), and so colonies were plated onto selective plates (2.1.1.2) to ensure the maintenance of the vector, and left for another week before rescreening (Figure 34B). Rescreening identified a single pure *mti* mutant. Due to ghosting on the gel, it was thought prudent to streak to single colonies and screen again. Rescreening (Figure 34C) identified six knockouts of the chromosomal *mti* gene. Successful colonies were stocked (2.1.2.5), and a plasmid loss procedure performed as described in 2.6.7. After several weeks of growth and patch plating no loss of ability to grow on selective medium was observed, indicating no loss of plasmid; this suggests that retention of the plasmid is in some way essential. Cells were stocked (2.1.2.5) to await future work.

4.3.2.4.1 Δ*mti* gene essentiality

A potential explanation for the inability of transformed cells to lose their resistance to antibiotic selection is the continued maintenance of the plasmid. This may have been due to genetic essentiality of *mti*. If a double crossover had occurred, and the gene was maintained and expressed from the pSBRCas9-Cpf-*mti* plasmid, then plasmid loss would not occur. However, as noted in the Δ*cop* mutant characterisation (4.3.2.3.1), it was unlikely, unless a plasmid has significant mutations, that an exchanged wild-type allele located on the plasmid could be expressed from the plasmid.

As it appeared that the *mti* gene may have been retained, internal screening colony PCRs were performed (Figure 35) using the primers mti Internal Forward and mti Internal Reverse. It was possible that after homologous recombination, the gene was maintained on the plasmid or had inserted elsewhere in the genome. A colony PCR with internal screening primers should bind within the gene and amplify it irrespective of its location in the cell; this would confirm whether *mti* was still present, either on a plasmid or reintegrated into the genome.

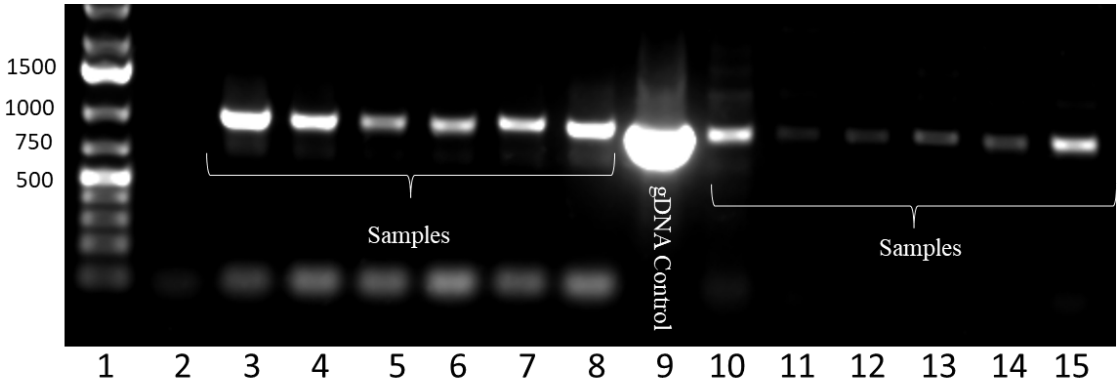

Figure 35 - Screening for E. limosum mti by colony PCR internal screening for deletion

A standard DreamTaq colony PCR was performed on 12 colonies of E. limosum NG-6849 transformed with pSBRCas9-Cpf-mti. Primers (mti Internal Forward and mti Internal Reverse) were used to screen for a chromosomal knockout of the mti (lanes 3 - 15). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 9. The ladder used (in lane 1) was GeneRuler 1 kb DNA Ladder, samples were run on a 1% agarose gel. Wild-type-sized internal fragments were seen for all colonies, suggesting the presence of the gene, likely on a plasmid.

Screening with internal primers (Figure 35) for *mti* revealed that *mti* was present in all screened colonies. As *mti* was still present, a plasmid miniprep (2.2.2) from the Δ*mti* mutant was performed and introduced into *E. coli* (2.2.1.1). Once colonies arose, plasmid minipreps were performed and the obtained DNA sequenced with primers Cas_pRECAS_seq_F and pCB102_pRECAS_seq_r. Only short, low-quality reads could be obtained, despite many additional attempts and different primers. The presence of internal *mti* fragments while screening with external primers suggested a chromosomal knockout, providing evidence for gene retention. While sequencing of retained plasmids was not successful, the fact that extracted genetic material was successfully introduced into *E. coli* and viable colonies arose on antibiotic selection indicates that a derivative of the originally introduced plasmid was retained in *E. limosum*, potentially expressing the *mti* gene. It is also possible that the antibiotic resistance gene had permanently integrated into the genome. This would explain the screening results observed and explain the low-quality reads of extracted DNA, although it would not explain how transformed *E. coli* was able to grow on selection.

4.3.2.5 Generation of *E. limosum* NG-6849 Δ*mtiI* mutants

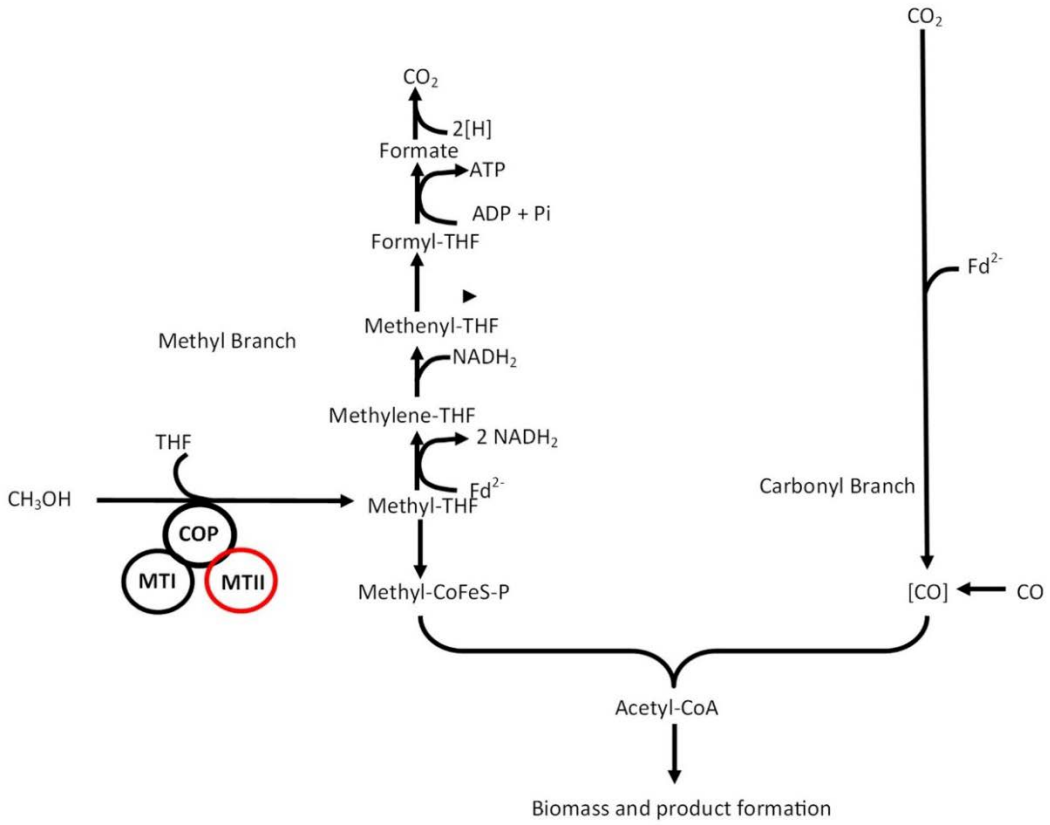

Figure 36 - Hypothetical pathway for methanol utilisation in E. limosum

Enzymes hypothesised to initiate methanol breakdown are displayed in the black and red circles and include the proteins encoded by the cop, mti and mtiI, red indicating the gene to be knocked out in the following section. Based on the hypothetical methanol utilisation pathway proposed in (325).

The *E. limosum mtiI* gene encodes an MtaA (WP_038352459.1) like protein, mtiI, , a methyltetrahydrofolate: cobalamin methyltransferase protein (108) and is homologous to the *A. woodii* protein with the same name (106). MtaA, together with an Mta18 like protein, mti, and MtaC19 like protein, cop, forms a complex that metabolise methanol (105) to methyl-THF, which is further metabolised (Figure 36). Mutants were attempted as described previously (2.6.5) using the plasmid pSBRCas9-Cpf-*mtiI*.

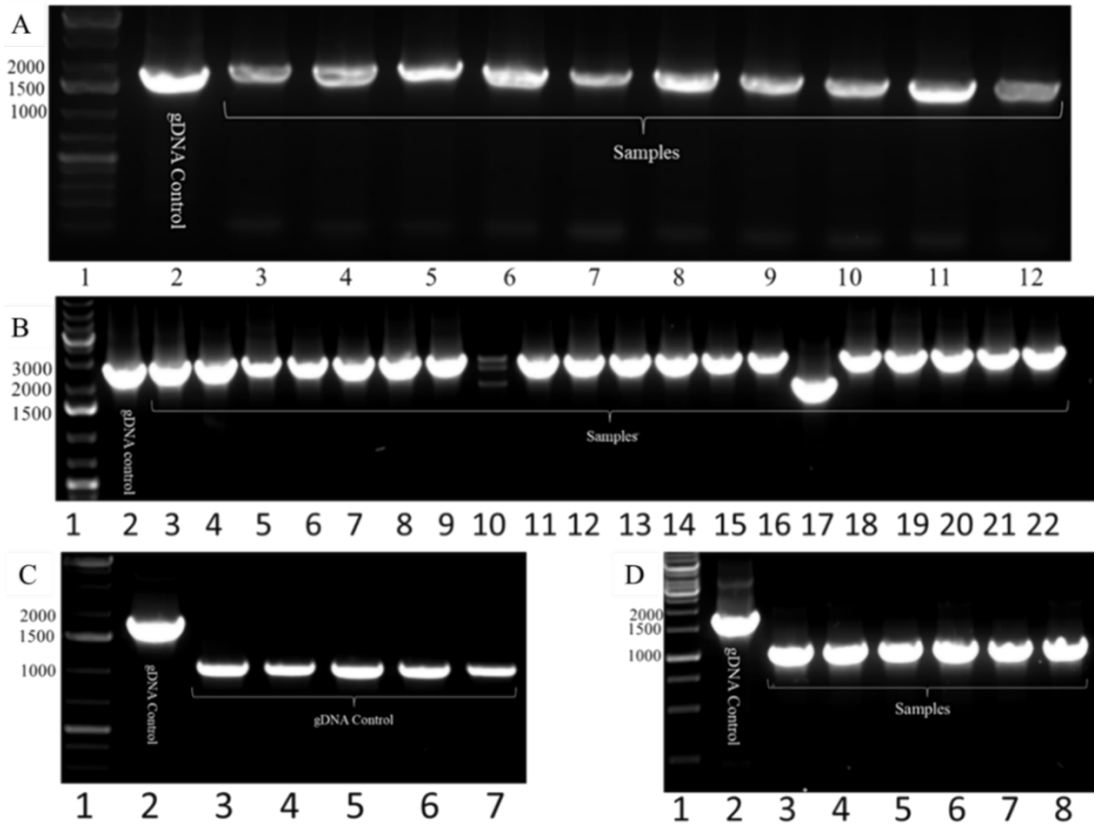

Figure 37 - Screening for E. limosum mtiI deletion mutants by colony PCR

A standard DreamTaq colony PCR was performed on E. limosum NG-6849 transformed with pSBRCas9-Cpf-mtiI. Primers (mtiI External Forward and mtiI External Reverse) were used to screen for a chromosomal knockout of the mtiI. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb DNA Ladder, samples were run on a 1% agarose gel. Repeats were performed due to previous concerns of wild-type contamination of mutant lines.

- ***A.*** *10 colonies (lanes 3 - 12) were screened for a knockout of mtiI. Wild-type bands of ~1.7 kb appeared in screened colonies indicating no mixed genotypes and no pure mutants. As the initial round of screening using colony PCR displayed no mutants, mixed or pure, cells were allowed to grow on plated for an additional week (2.5.2.3) and rescreened as shown in B.*
- ***B.*** *20 colonies (lanes 3 - 22) were rescreened for a knockout of mtiI. Deletion bands (1.5 kb) appeared in lane 17 indicating a pure mutant. The pure mutant created (lane 17) using CRISPR-Cas9 was streaked out to single colonies on standard selective plates (2.1.1.2) and screened once again with colony PCR as shown in C.*
- ***C.*** *5 colonies (lanes 3 - 7) were rescreened for a knockout of mtiI. Lanes 3 - 7 show a mutant band indicating pure ΔmtiI mutants. ΔmtiI mutants were streaked to fresh antibiotic selection plates and allowed to grow for a further week before being screened again as shown in D.*
- ***D.*** *6 colonies (lanes 3 - 8) were rescreened for a knockout of mtiI. Lanes 3 - 8 show a mutant band indicating pure ΔmtiI mutants.*

Screening (Figure 37) identified six colonies that appeared to be *mtiI* chromosomal deletion mutants. Successful pure deletion colonies were stocked (2.1.2.5), and plasmid loss was performed (2.6.7). Once the cells no longer grew on antibiotic selection plates, they cells were screened to ensure the maintenance of the Δ*mtiI* genotype. Δ*mtiI* mutant cells were stocked (2.1.2.5) and carried forward for

phenotype assessment (4.3.2.5.1). Screening (Figure 38) showed that cells which had lost the ability to grow on antibiotic selection had retained chromosomal deletion of *mtiI*.

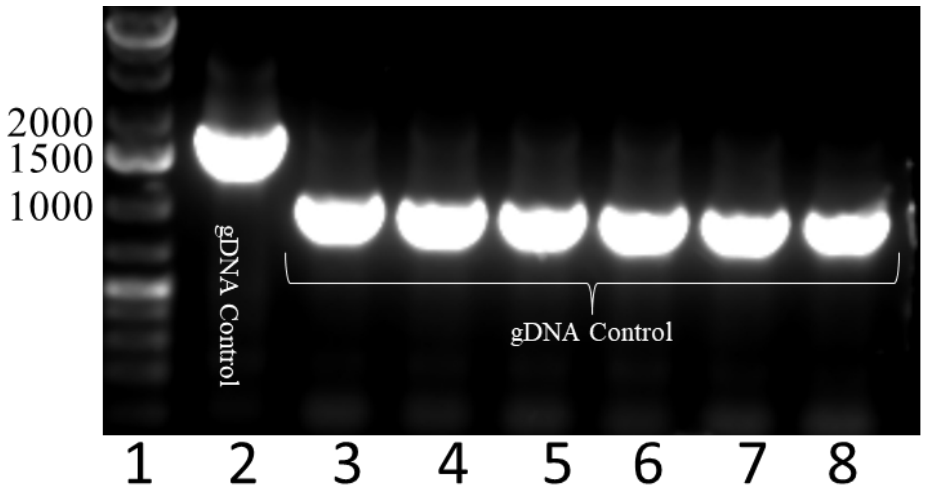

Figure 38 - Screening for E. limosum mtiI deletion mutants by colony PCR

A standard DreamTaq colony PCR was performed on 6 colonies of E. limosum NG-6849 transformed with pSBRCas9-Cpf-mtiI, these cells were previously identified as pure ΔmtiI mutants. Primers (mtiI External Forward and mtiI External Reverse) were used to screen for a chromosomal knockout of the mtiI (lanes 3 - 8). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2, samples were run on a 1% agarose gel. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder. Lanes 3 - 8 show a mutant band indicating pure ΔmtiI mutants.

4.3.2.5.1 Characterisation of *E. limosum* NG-6849 Δ*mtiI* mutants

A Δ*mtiI* mutant would theoretically be unable to grow on methanol because there would be no way to pass a methyl group from methanol onto THF (in the Wood-Ljungdahl pathway). However, *E. limosum* has different methyltransferases (of the MttB family) that work on different substrates. Each of these has an associated cobalamin protein and a second methyltransferase (as Mta18 has MtaC19 and MtaA). It is possible that there could be interaction between Mta18/MtaC19 and one of these MtaA-like enzymes. The creation of this mutant would determine whether this occurs to a significant enough degree to allow growth on methanol without the MtaA protein being produced.

A phenotype assessment for growth on methanol was performed (Figure 39) from single colonies of both the Δ*mtiI* mutant and wild-type NG-6849 strains. Starter cultures were grown from single colonies of cells in a Pacaud medium (2.1.1.2) with 2

g/L yeast extract and 15 mM glucose. Three replicate cultures were grown, referred to here as Δ*mtii* 1, Δ*mtii* 2 and Δ*mtii* 3Glucose was used as the carbon source opposed to methanol as Δ*mtiI* mutants were expected to be unable to grow in the presence of methanol. Due to the risk of glucose carryover, cells were centrifuged to remove medium and were resuspended in inert medium without a carbon source. Once this was done, serum bottles were prepared as standard (2.1.2.1) using Pacaud medium (2.1.1.2) with yeast extract present at 2 g/L and methanol at 200 mM.

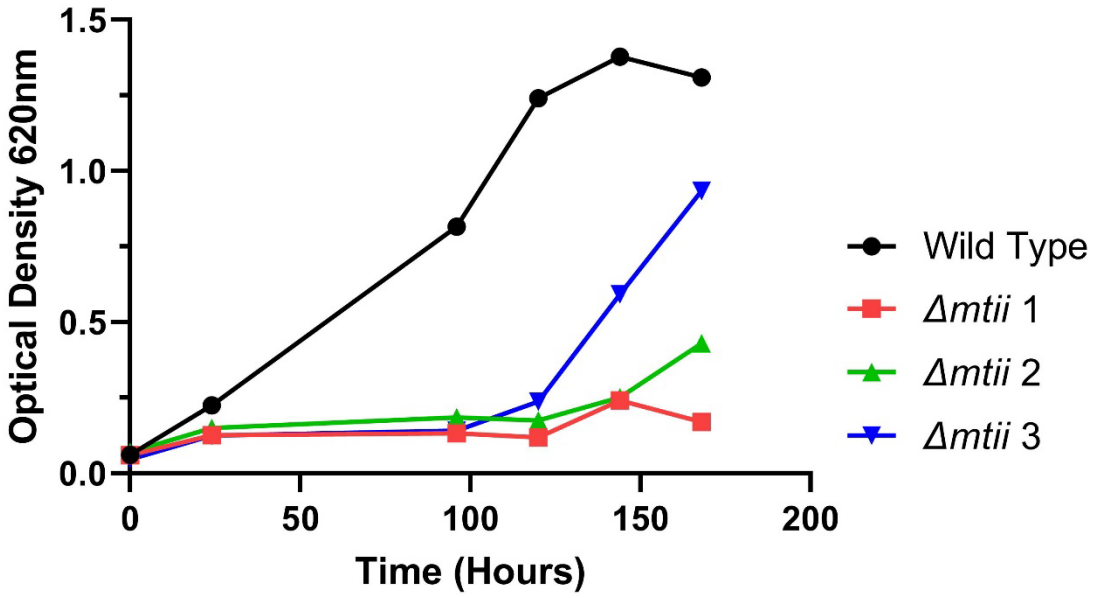

Figure 39 - Graph displaying growth of ΔmtiI and wild-type E. limosum NG-6849 strain utilising methanol as a carbon source.

Cells were grown in serum bottles were set up using Pacaud medium with yeast extract present at 2 g/L and methanol at 200 mM. Optical density was measured at 620 nm, every 24 hours except at 48 and 72 hours. The wild type was grown as a single culture, Δmtii 1, Δmtii 2 and Δmtii 3 are replicates from a single generated mutant culture. Wild-type E. limosum NG-6849 was grown as singlet.

Initially, both Δ*mtiI* mutant and wild-type cell lines behaved as expected, with no growth observed in the mutant. However, after around 120 hours, one of the replicates of Δ*mtiI* mutants started to increase biomass, indicating an ability to grow on methanol. Maximal biomass was approached by 168 hours and was like that of the wild-type strain. The second mutant replicate, Δ*mtiI* 2, also started to grow around this time. This effect is comparable to what was observed for Δ*cop* mutants. As in the case of Δ*cop* (4.3.2.3.1), colony PCR was performed on biomass from the end point of growth at 168 hours, to assess the presence and location *mtiI* using primers binding within and outside the gene (Figure 40).

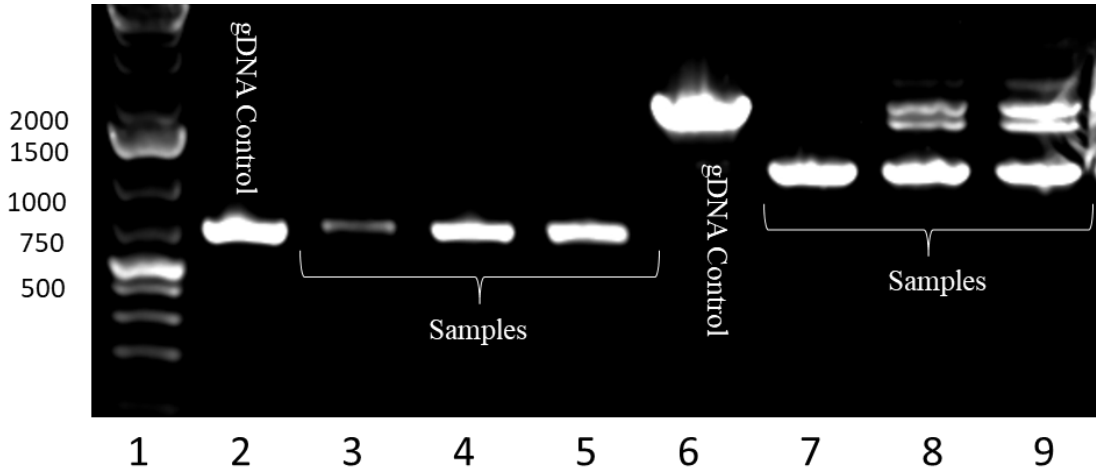

Figure 40 - Screening for E. limosum re-emergence of wild-type band in mtiI mutants after growth experiment by colony PCR

A standard DreamTaq colony PCR was performed on 1µL of the 168-hour end points of Figure 39, ΔmtiI E. limosum NG-6849 grown on methanol for 168 hours. Primers (mtiI internal Forward and mtiI internal Reverse) were used to screen for the absence of the mtiI gene (lanes 3 - 6). Primers (mtiI External Forward and mtiI External Reverse) were used to screen for a chromosomal knockout of the mtiI gene (lanes 8 - 10). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2 and 6. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 1% agarose gel. All internal PCRs showed the gene present, external primers showed a mixed genotype in all but lane 7, suggesting a return to a wild-type genotype for some of the cell mass. Lanes 2 and 7, representing mtiI 1, which did not grow on methanol even after 168 hours, maintained the chromosomal genotype of the ΔmtiI mutant.

Screening (Figure 40) showed the presence of the *mtiI* gene in all samples when screened with internal primers. However, the first replicate of the Δ*mtiI* cell line (Lane 7, Figure 40) did not show wild-type banding with external primers and appears to be a pure deletion. This suggests that there is retention of the plasmid containing the *mtiI* gene or the genomic integration of the gene elsewhere in the genome. It is likely that this is occurring in a minority of the population due to the faint banding seen in lane 3 (Figure 40). Furthermore, the lack of wild-type banding suggests that no genetic restoration to the wild type has occurred, which is consistent with the growth data (Figure 39). As with Δ*cop* (4.3.2.3.1), plasmid sequencing was attempted but again proved fruitless.

4.3.2.6 Generation of *E. limosum* NG-6849 Δ*metF1* mutants

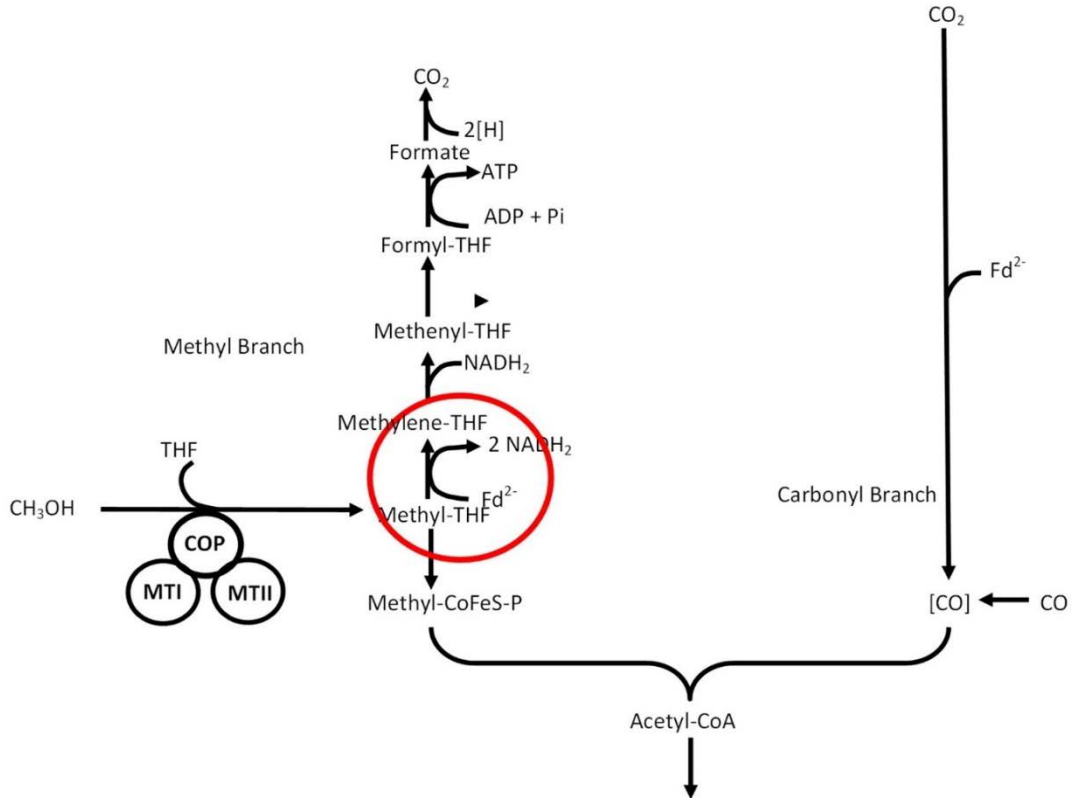

Figure 41 - Proposed role of metF1 gene product in the E. limosum methanol utilisation pathway

The reaction catalysed by MetF1 is highlighted by the red circle. The directionality of the methyl branch, as in towards CO_2 production, in indicative of methanol growth; direction would be reversed on CO_2/H_2 growth.

E. limosum contains two copies of *metF*, one of which (WP_038351520.1) is not associated with an adjacent *metV* (WP_038351866.1) and is referred to as *metF1* in this text. This is unlike the other copy of *metF* (WP_081571099.1) which is part of a protein complex (Figure 41) believed to function together with *metV* as part of the Wood-Ljungdahl pathway (Dr. Helge M. Dietrich, Dr. Christian Öppinger, and Professor Volker Müller, Goethe University Frankfurt, personal communication).

metF1 codes for a 5,10-methylenetetrahydrofolate reductase, also known as methylene THF reductase (108). Methylene THF reductase is believed to be involved in the reverse methyl branch of the Wood-Ljungdahl pathway, the reverse direction on the methyl branch, when feeding on methanol. Its function is to convert methylene THF into methyl-THF and vice versa. For the *metF*, the mechanism was hypothesised to involve an NAD(P)H/ NAD(P)+ dependency (Dr. Helge M. Dietrich, Dr. Christian

Öppinger, and Professor Volker Müller, Goethe University Frankfurt, personal communication). Mutants were attempted as described previously (2.6.5.1) using pSBRCas9-Cpf-*metF1*. Colonies which arose on plates were screened using colony PCR (Figure 42) with the primers metF1 External Forward and metF1 External Reverse.

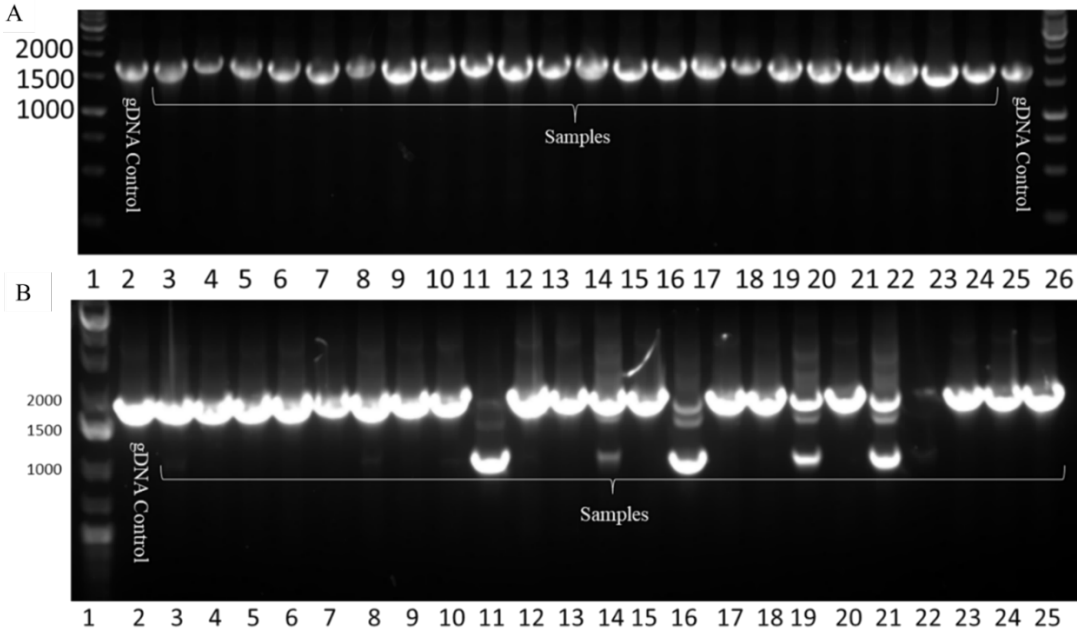

Figure 42 - Screening for E. limosum metF1 deletion mutants by colony PCR

A standard DreamTaq colony PCR was performed on E. limosum NG-6849 transformed with pSBRCas9-Cpf-metF1. Primers (metF1 External Forward and metF1 External Reverse) were used to screen for a chromosomal knockout of the metF1. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2 (and lane 25 in 1). The ladder used (in lane 1) was GeneRuler 1 kb DNA Ladder, samples were run on a 1% agarose gel.

- ***A.*** *22 colonies (3 - 24) were screened for knockout of metF1. wild-type bands (of ~1.7 kb) appeared in screened colonies indicating no mixed genotypes and no pure mutants. Initial colony PCR screening failed to identify any mutants and so colonies were plated onto selective plates (2.1.1.2) and left for another week before rescreening as shown in B.*
- ***B.*** *23 colonies (3 - 25) were rescreened for knockout of metF1. Mutant bands (of ~1 kb) appeared in screened colonies 11, 14, 16, 19 and 21. indicating mixed genotypes in all cases due to the presence of wild-type sized fragments.*

Initial screening by colony PCR (Figure 42A) revealed the absence of mutants. Candidate colonies were therefore streaked to single colonies on selective plates, incubated for another week, and then rescreened (Figure 42B). The second round of screening suggested the emergence of Δ*metF1* mutants, but the colonies were mixed with the wild type. Mixed colonies were streaked to single colonies on selective medium (2.1.1.2) and allowed to recover for a week before rescreening using colony PCR (Figure 43). A single colony from each plate was taken for assessment.

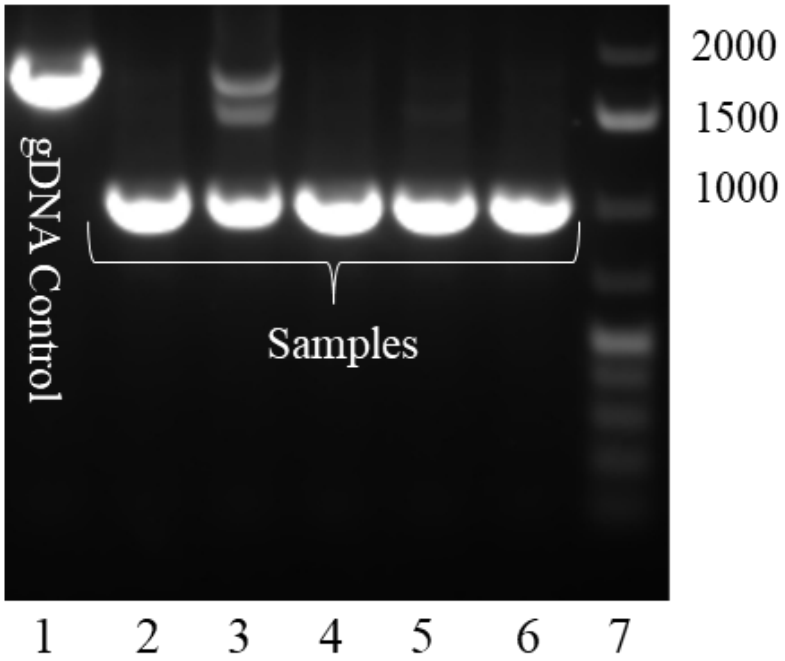

Figure 43 - Screening for E. limosum metF1 deletion mutants by colony PCR

A standard DreamTaq colony PCR was performed on 5 colonies of putative E. limosum NG-6849 ΔmetF1 (2 - 6). Colonies were those which in previous screenings (Figure 42B) had appeared to be of mixed genotype. Primers (metF1 External Forward and metF1 External Reverse) were used to screen for a chromosomal knockout of the metF1 (lanes 2 - 6). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 1. The ladder used (in lane 7) was GeneRuler 1 kb Plus DNA Ladder, samples were run on a 1% agarose gel. Lanes 2, 4, 5 and 6 display a single band ~600 - 800 bases less than the wild-type, indicating a pure metF1 knockout. Lane 3 has a wild-type band in addition to the deletion band indicating a mixed mutant colony.

Screening (Figure 43) identified four of the colonies which appeared to be successful pure chromosomal deletions of *metF1*. Successful pure deletion colonies were stocked (2.1.2.5) and plasmid loss performed (2.6.7). Cells which did not grow on selection but grew without selection were screened again (Figure 44) to ensure they had retained their Δ*metF1* mutant genotype.

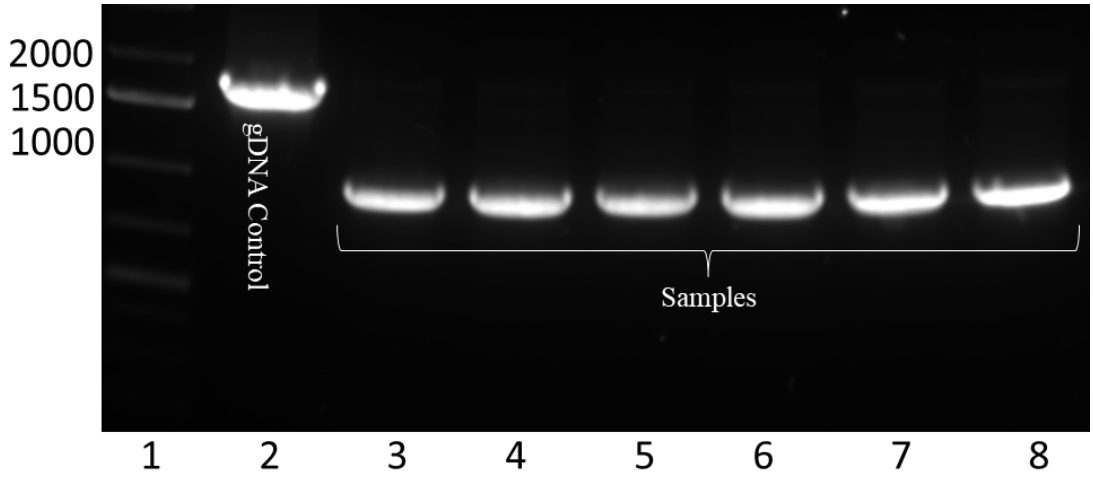

Figure 44 - Screening for E. limosum metF1 deletion mutants by colony PCR

A standard DreamTaq colony PCR was performed on 6 colonies of E. limosum NG-6849 ⊿metF, following plasmid loss. Primers (metF1 External Forward and metF1 External Reverse)) which found external to metF1 were used to rescreen for a chromosomal knockout of the metF1 (lanes 3 - 8). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 1% agarose gel.

Screening (Figure 44) performed on cells which has lost the ability to grow on antibiotic selection, showed a retention of the chromosomal deletion of *metF1*. Due to the issues seen in *cop* (4.3.2.3), *mti* (4.3.2.4) and *mtiI* (4.3.2.5) it seemed prudent to screen again with internal primers, metF1 internal Forward and metF1 internal Reverse, (Figure 45) to ensure a copy of the gene was not retained on a plasmid or at an alternate site in the genome.

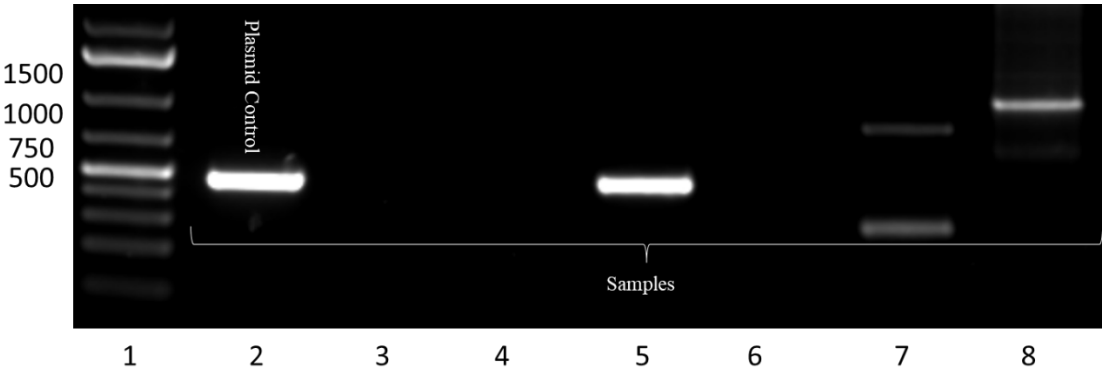

Figure 45 - Screening for E. limosum metF1 by colony PCR screening

A standard DreamTaq colony PCR was performed on 6 colonies of E. limosum NG-6849 which external colony PCR purported to be ⊿metF1 mutants (Figure 44). Primers (metF1 internal Forward and metF1 internal Reverse) which bound within metF1 were used to screen for retention of the metF1 gene (lanes 3 - 8). A control of pSBRCas9-Cpf-metF1 plasmid was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 1% agarose gel. Colony 5 displayed retention, and colonies 7 and 8 displayed some non-specific banding which may indicate a retained mutated plasmid.

Screening (Figure 45) identified a single colony previously having been confirmed as being a Δ*metF1* mutant retaining the *metF1* gene. The screening was done after an additional streaking to single colonies and before applying a selective pressure in favour of the presence of *metF1* for genomic reintegration. This corroborates the working hypothesis that a mutated plasmid is maintained in a minority of cells. This finding highlights a fundamental issue with the CRISPR-Cas9 system used in this study and emphasises the need to solve this issue or use an alternative genetic modification system. Due to the expectation of a repeat of the results observed with *cop*, *mti*, and *mtiI* the characterisation of the Δ*metF1* generated by CRISPR-Cas9 was not performed.

4.3.2.7 Generation of *E. limosum* NG-6849 Δ*metVmetF* mutants

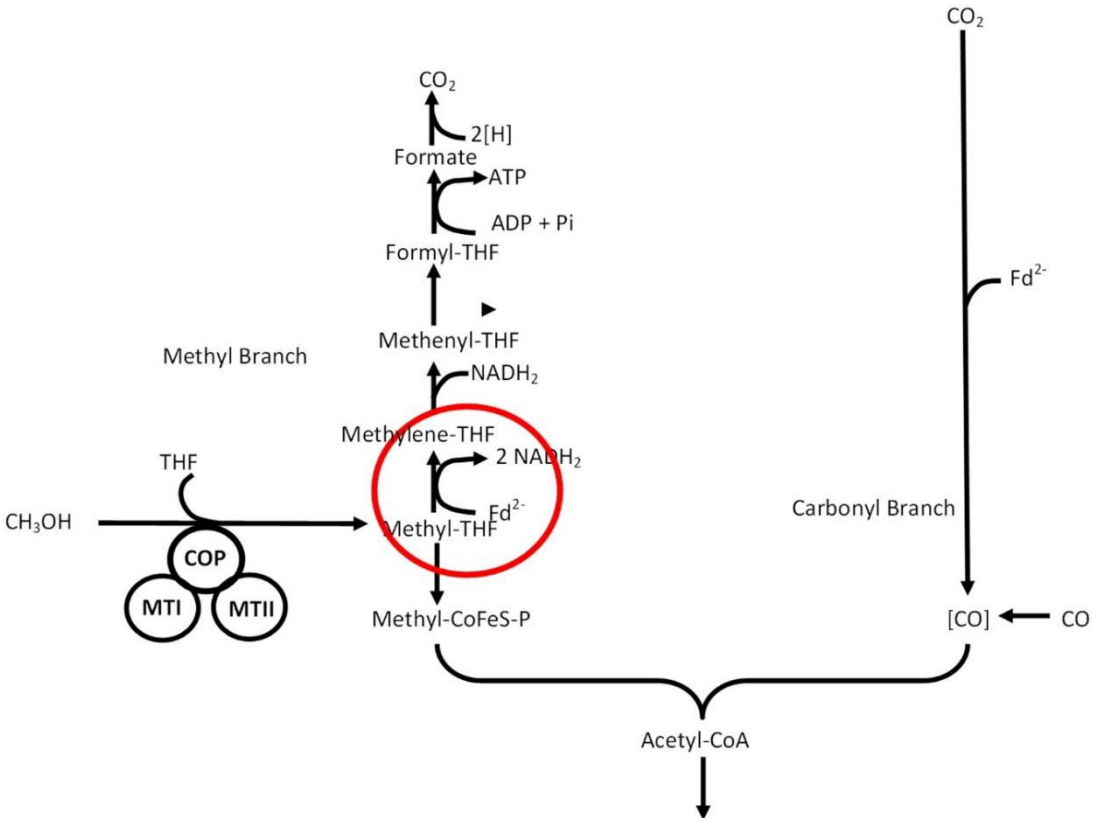

Figure 46 - Proposed role of metVmetF gene product in the E. limosum methanol utilisation pathway

The reaction catalysed by MetVF is highlighted by the red circle. The directionality of the methyl branch, as in towards CO_2 production, in indicative of methanol growth; direction would be reversed on CO_2/H_2 growth.

metVmetF is a pair of genes located sequentially with some overlap (108). These genes are homologous to the genes of the same name in *A. woodii*. They encode the

two methylenetetrahydrofolate reductase (108) subunits, WP_038351866.1 and WP_081571099.1, respectively. It was hypothesised that these genes involved in the reverse methyl branch of the Wood-Ljungdahl pathway (Figure 46), converting some of the methylene-THF to methyl-THF (106). Deletion of these genes would help determine whether their function was conserved in *E. limosum* and if their function could be compensated for by another gene. Mutants were attempted as described previously (2.6.5.1) using pSBRCas9-Cpf-*metVmetF*.

Screening of mutants via colony PCR (2.5.2.3) using the primers metVmetF External Forward and metVmetF External Reverse, (Figure 47) took several rounds to isolate pure knockout colonies. A mixed mutant (Figure 47C) was identified after the second round of re-streaking and streaked to single colonies over two plates. A further round of screening (Figure 47D) identified 8 pure Δ*metVmetF* mutant colonies from the descendants of the mixed colony. The final round of selection involved screening these two plates, hence the colocation of mutants in the final gel (Figure 47D). One plate had all mutant colonies, while the other had none. Successful pure deletion colonies were stocked (2.1.2.5), and plasmid loss was performed (2.6.7). After several weeks of growth and patch plating, no loss of ability to grow on selective medium was observed. Cells were stocked (2.1.2.5) to await future work. Of concern TraDIS data (Dr. Elizabeth Redfern, University of Nottingham, personal communication) later showed that the *metF* portion of *metVmetF* was essential for growth on glucose. This is discussed further later in this work (4.3.4.8), but in short, it lends yet more weight to the hypothesis of low-level plasmid retention and genomic integration.

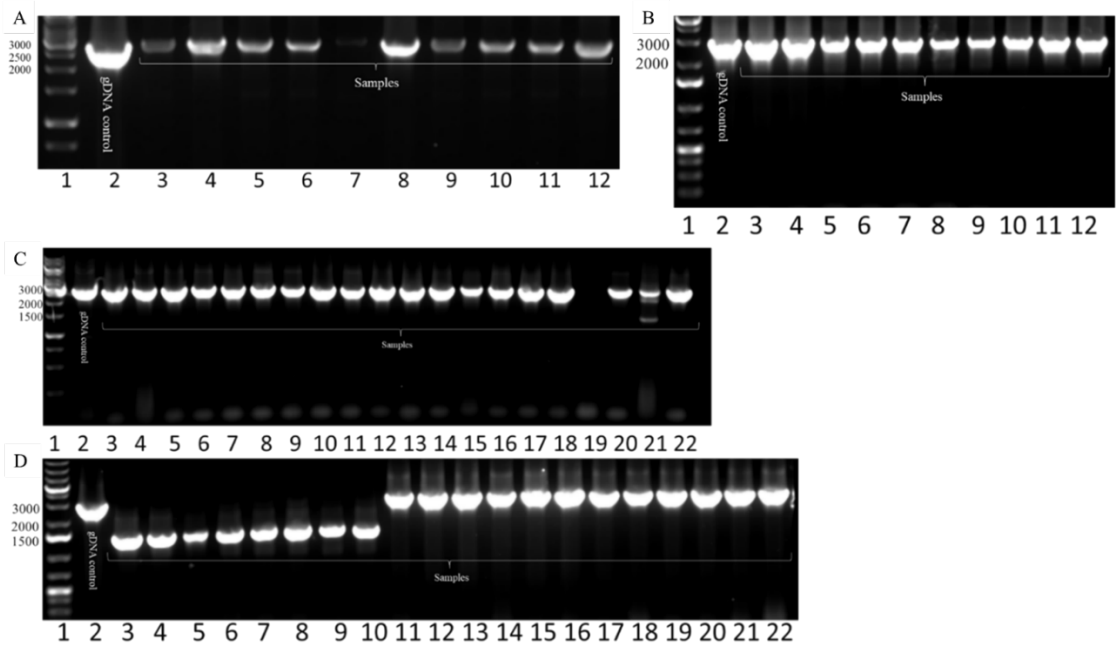

Figure 47 – Screening for E. limosum metVmetF deletion mutants by colony PCR

A standard DreamTaq colony PCR was performed on E. limosum NG-6849 transformed with pSBRCas9-Cpf- metVmetF. Primers (metVmetF External Forward and metVmetF External Reverse) were used to screen for a chromosomal knockout of the metVmetF. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb DNA Ladder for A and GeneRuler 1 kb DNA Ladder for B – D, samples were run on a 1% agarose gel.

- ***A.*** *10 colonies (lanes 3 – 12) were screened for knockout of metVmetF. No mutants appeared so cells were allowed to recover for 1 week before screening again. Initial colony PCR screening failed to identify any mutants and so colonies were plated onto selective plates (2.1.1.2) and left for another week before rescreening (Figure 47B).*
- ***B.*** *10 colonies (lanes 3 – 12) were rescreened for knockout of metVmetF No mutants appeared so cells were allowed to recover for 1 week before screening again. Further colony PCR screening failed to identify any mutants and additional colonies were rescreened (Figure 47C).*
- ***C.*** *20 colonies (lanes 3 – 22) were rescreened for knockout of metVmetF. A mix mutant appeared (lane 21), with a band of ~1.5 kb, however it is worth noting a band at ~2.5kb (slightly below the wild-type band of ~3 kb) is present. Mixed colonies were streaked to single colonies on selective medium and allowed to recover for a week before rescreening using colony PCR; due to the single mixed mutant several colonies were taken for assessment (Figure 47D).*
- ***D.*** *20 colonies (lands 3 – 22) were rescreened for a chromosomal knockout of the metVmetF. Pure mutants were present in lanes 3 – 10, these mutants were carried forward for plasmid loss.*

4.3.2.8 Generation of *E. limosum* NG-6849 Δ*act*Δ*metVmetF* mutants

The relevance of *metVmetF* is discussed in 4.3.2.7. With the creation of a successful Δ*act* mutant an opportunity to create a double mutant background arose. Using the existing Δ*act* mutant and the recently successful pSBRCas9-Cpf-*metVmetF* CRISPR-Cas9 knockout vector, the generation of a Δ*act*Δ*metVmetF* mutant was attempted as detailed previously (2.6.5.1). Screening (Figure 48) failed to identify any mutants, so cells were streaked onto selective plates and rescreened for seven weeks, with no mutants arising. A single gel is displayed due to the redundancy of six additional near identical gels all displaying failure. This was perplexing because the vector had previously worked in a wild-type background (4.3.2.7).

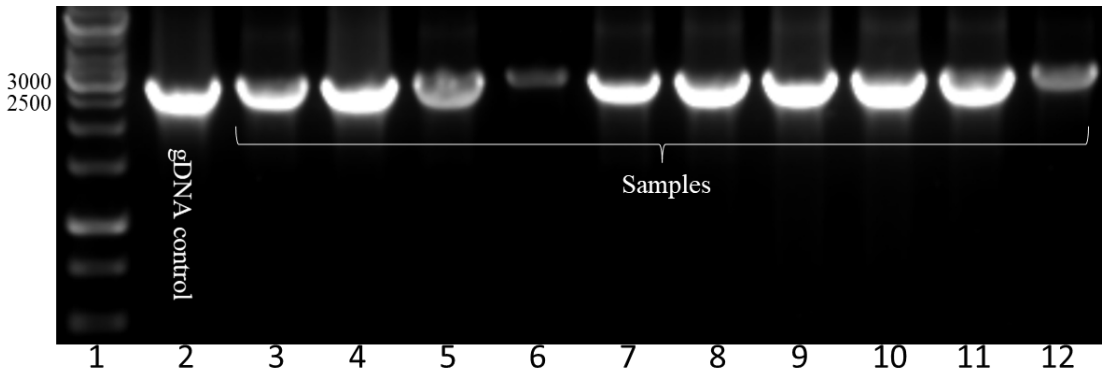

Figure 48 – Screening for E. limosum metVmetF deletion mutants by colony PCR in a E. limosum NG-6849 Δact background

A standard DreamTaq colony PCR was performed on 10 colonies of E. limosum NG-6849 Δact transformed with pSBRCas9-Cpf-metVmetF. Primers (metVmetF External Forward and metVmetF External Reverse) were used to screen for a chromosomal knockout of the metVmetF (lanes 3 – 12). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 1. The ladder used (in lane 1) was GeneRuler 1 kb DNA Ladder, samples were run on a 1% agarose gel. No mutants were isolated.

The failure of a knockout plasmid to generate mutants, which appeared to have been successful previously (4.3.2.7), was unexplained until the phenotype assay work for Δ*cop* (4.3.2.3.1), Δ*mti* (4.3.2.4.1) and Δ*mtiI* (4.3.2.5.1) had been performed, suggesting plasmid retention. As Δ*act* was similarly generated, using the same CRISPR-Cas9 system, it is probable that it was similarly affected. This could result in the observed failure, as it may not be necessary for cells to maintain the introduced pSBRCas9-Cpf-*metVmetF* plasmid. If the prior knockout plasmid is being retained, then antibiotic selection could be conferred by it. This, in turn, would result in its loss within the population due to a lack of maintenance pressure. However, later work

discovered that this gene pair was, in fact, essential (4.3.4.8), though the working theory at the time was as detailed prior.

4.3.2.9 Generation of *E. limosum* NG-6849 Δ*pbt1* mutants

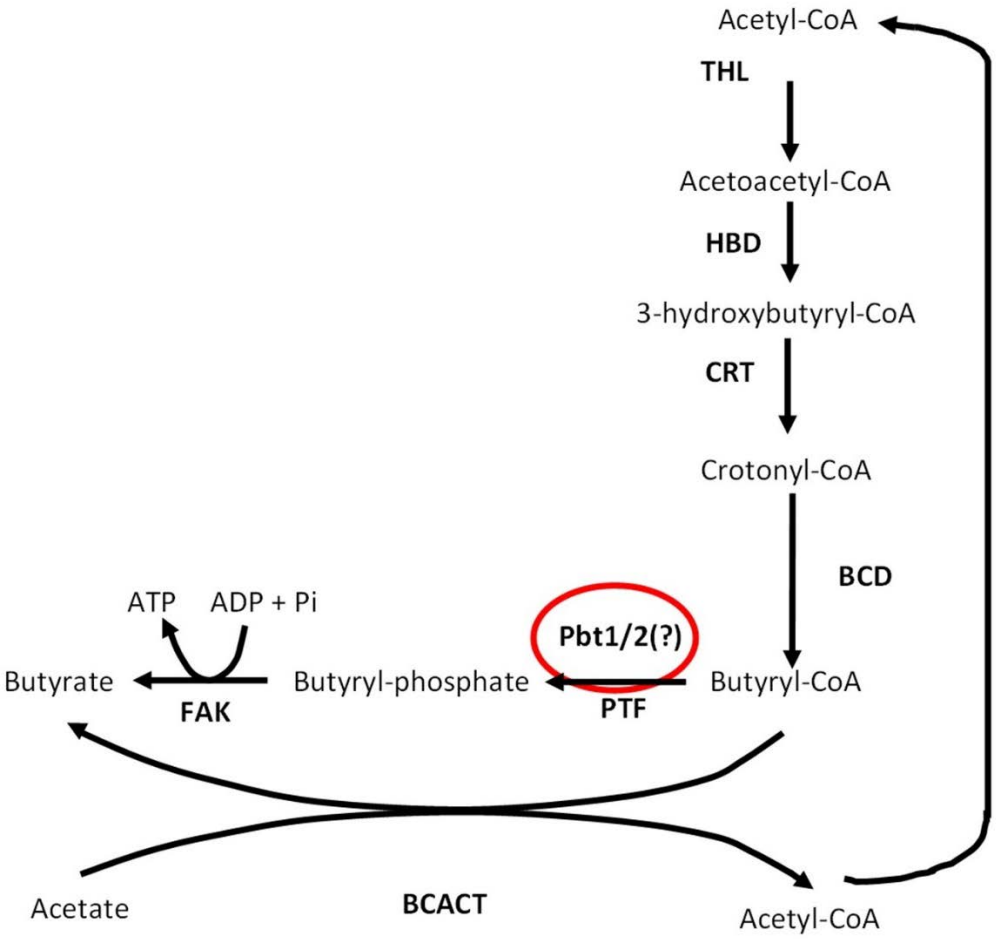

Figure 49 – Putative reaction catalysed by phosphate butyryltransferase (Ptb).

Pathway leading from acetyl-CoA to butyrate, with the Ptb reaction highlighted in red. BCACT butyryl CoA:acetate CoA transferase; CRT, crotonase; FAK, fatty acid kinase; HBD, 3-hydroxybutyryl-CoA dehydrogenase; BCD butyryl-CoA dehydrogenase; PTF, phosphotransferase; THL, thiolase

E. limosum contains two putative genes for a phosphate butyryltransferase, called *pbt1* and *pbt2*, respectively. These proteins are purported to catalyse the conversion of butyryl-CoA to butyryl phosphate (Figure 49). It was hypothesised (Professor Philippe Soucaille, INSA, University of Toulouse, personal communication) that the phosphate butyryltransferase may have a degree of site promiscuity, which may result in these two genes being the phosphotransferase(s) that catalyses this step.

Of the two phosphate butyryltransferase genes *pbt1*, WP_038352195.1, is the more highly expressed of the two (Dr. Yoseb Song, Massachusetts Institute of Technology, personal communication). Deletion of this gene might, therefore, be expected to have a greater phenotypical effect than *pbt2*. The CRISPR-Cas9 knockout vector, pSBRCas9-Cpf-*pbt1*, was created (2.6.1), and used to attempt a knockout of *pbt1* as detailed previously (2.6.5.1). Recovered colonies were screened for genomic deletion of *pbt1* (Figure 50) using colony PCR (2.5.2.3) with the primers pbt1 External Forward and pbt1 External Reverse.

Initial colony PCR (Figure 50A) did not identify any mutants. Colonies were therefore streaked to single colonies on thiamphenicol selection plates and left a week before screening again. The second round of screening (Figure 50B) again did not identify mutants so again colonies were streaked. The third round of screening (Figure 50C) however identified several apparent pure Δ*pbt1* mutants. Plasmid loss (2.6.7) was then performed. Once cells were no longer able to grow under antibiotic selection, colonies were again screened to confirm the deletion prior to stocking (2.1.2.5) for future work. Screening (Figure 51) identified the retention of the Δ*pbt1* mutant genotype in colonies which were unable to grow on antibiotic selection, indicative of plasmid loss.

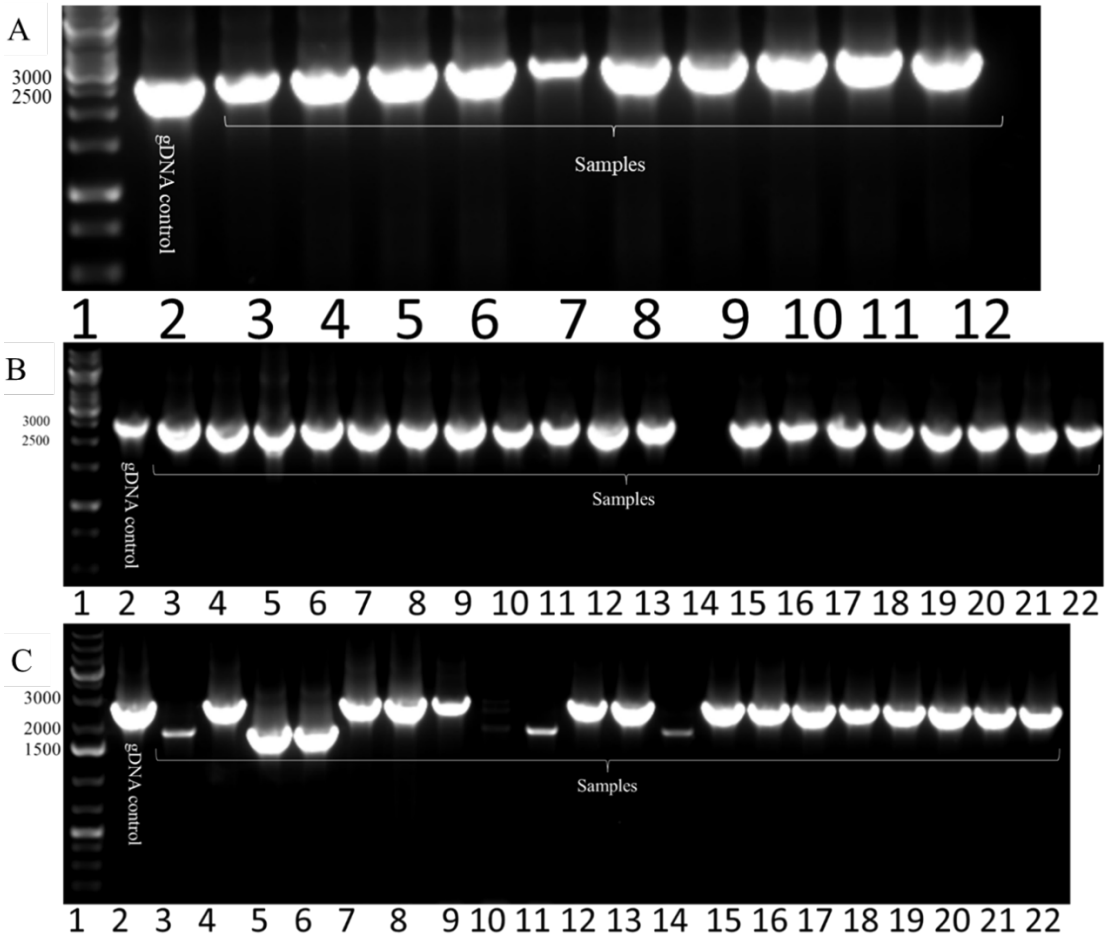

Figure 50 – Screening for E. limosum pbt1 deletion mutants by colony PCR in a wild-type E. limosum NG-6849 background

A standard DreamTaq colony PCR was performed on E. limosum NG-6849 transformed with pSBRCas9-Cpf- pbt1. Primers (pbt1 External Forward and pbt1 External Reverse) were used to screen for a chromosomal knockout of the pbt1. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb DNA Ladder, samples were run on a 1% agarose gel.

- ***A.*** *10 colonies (lanes 3 – 12) were screened for knockout of pbt1. No mutants appeared so cells were allowed to recover for 1 week before screening again as shown in B.*
- ***B.*** *20 colonies (lanes 3 – 22) were rescreened for knockout of pbt1. No mutants appeared so cells were allowed to recover for 1 week before screening again. Further screening failed to identify any mutants and so additional colonies screened using colony PCR as shown in C.*
- ***C.*** *20 colonies (lanes 3 – 22) were rescreened for knockout of pbt1. Pure mutants appeared in lanes 3, 5, 6, 11 and 14, these were carried through for further work.*

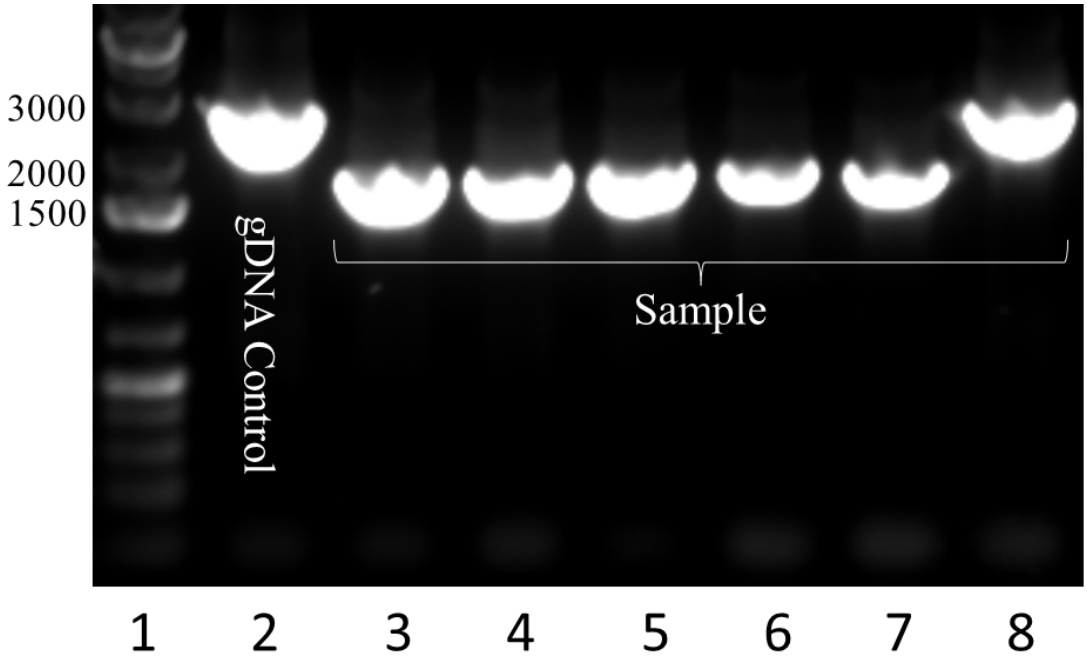

Figure 51 – Screening for E. limosum pbt1 deletion mutants by colony PCR in a wild-type E. limosum NG-6849 background

A standard DreamTaq colony PCR was performed on 6 colonies of E. limosum NG-6849 transformed with pSBRCas9-Cpf-pbt1. Primers (pbt1 External Forward and pbt1 External Reverse) were used to rescreen for a chromosomal knockout of the pbt1 (lanes 3 – 7) in cells which had purportedly undergone plasmid loss. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 1. The ladder used (in lane 1) was GeneRuler 1 kb DNA Ladder. Pure mutants appeared in lanes 3 – 7, these were stocked to await future work.

4.3.2.9.1 Δ*pbt1* ATCC 8486 growth on glucose

A knockout of *pbt1* has previously been conducted at the University of Nottingham in the original ATCC 8486 strain (117). It had been suggested (Professor Philippe Soucaille, INSA, University of Toulouse, personal communication) that as *pbt1* is a phosphate butyryltransferase, it may be involved in butyrate production; this possibility was explored due to the failure of the *act* knockout (4.3.2.1.1) to perturb butyrate production. The phenotype of the ATCC 8486 Δ*pbt1* mutant was therefore investigated, and butyrate production levels were assessed using HPLC (2.1.2.4.2). This work was performed concurrently with the generation of the Δ*pbt1* mutant in the NG-6849 strain.

Starter cultures were grown from single colonies of cells in Pacaud medium (2.1.1.2) with 2 g/L yeast extract and 15 mM glucose. Serum bottle growth was performed in triplicate using a control of wild-type *E. limosum* ATCC 8486. Serum bottles were set

up as standard (2.1.2.1) using Pacaud medium (2.1.1.2) with yeast extract present at 2 g/L and 15 mM glucose. Growth on methanol was not initially performed due to the preliminary data from Δ*act* characterisation (4.2.1.) which showed butyrate production for both carbon sources.

Optical density measurements (2.1.2.4.1) were used as a means of tracking biomass formation and were taken at 24-hour intervals throughout growth, including at 0 hours. HPLC samples were taken concurrently with optical density samples and samples were processed and analysed (2.1.2.4.2). As with Δ*act* characterisation (4.3.2.1.1), butyrate production levels and acetate production levels were of interest for this analysis.

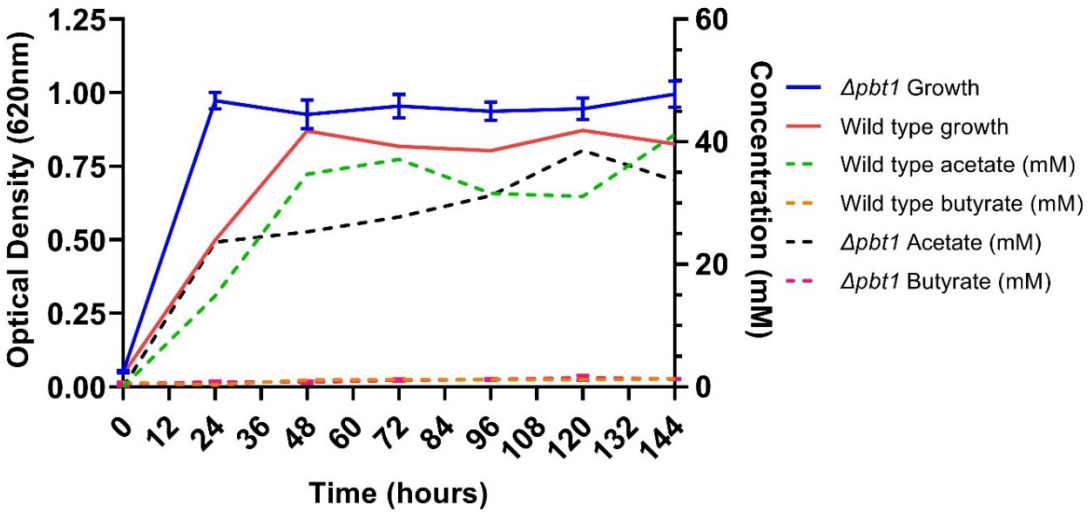

Figure 52 – Acetate and butyrate levels throughout E. limosum ATCC 8486 wild-type and E. limosum ATCC 8486 Δpbt1 E. limosum growth on glucose

Wild-type and Δpbt1 E. limosum NG-6849 were grown on glucose in batch culture in serum bottles and an anaerobic gas headspace (~8% hydrogen, ~77% N_2, ~15% CO_2) of 200 mL. Spectrophotometer readings for absorbance at 620 nm and samples for HLPC were taken every 24 hours. Δpbt1 mutants were grown in triplicate with the average values for OD and concentration of acetate and butyrate given. The wild type control was grown as a single flask.

Both the wild-type and Δ*pbt1* ATCC 8486 strains grew comparably, reaching maximum OD_{620} of 0.87 and 0.99, respectively (Figure 52). Acetate and butyrate production levels were similarly comparable, with respective maximal acetate titres of 41 mM and 57 mM, and butyrate titres of 1.31 mM and 1.38 mM, respectively. Butyrate levels were consistent among wild-type and Δ*pbt1* ATCC 8486. No

significant decrease occurred in the Δ*pbt1* background, and therefore knocking out *pbt1* alone did not disrupt butyrate production. Due to the lack of the expected phenotype, it did not seem necessary to characterise the Δ*pbt1* mutant in the NG-6849 strain once it arose using CRISPR-Cas9.

4.3.2.10 Generation of *E. limosum* NG-6849 Δ*pbt2* mutants

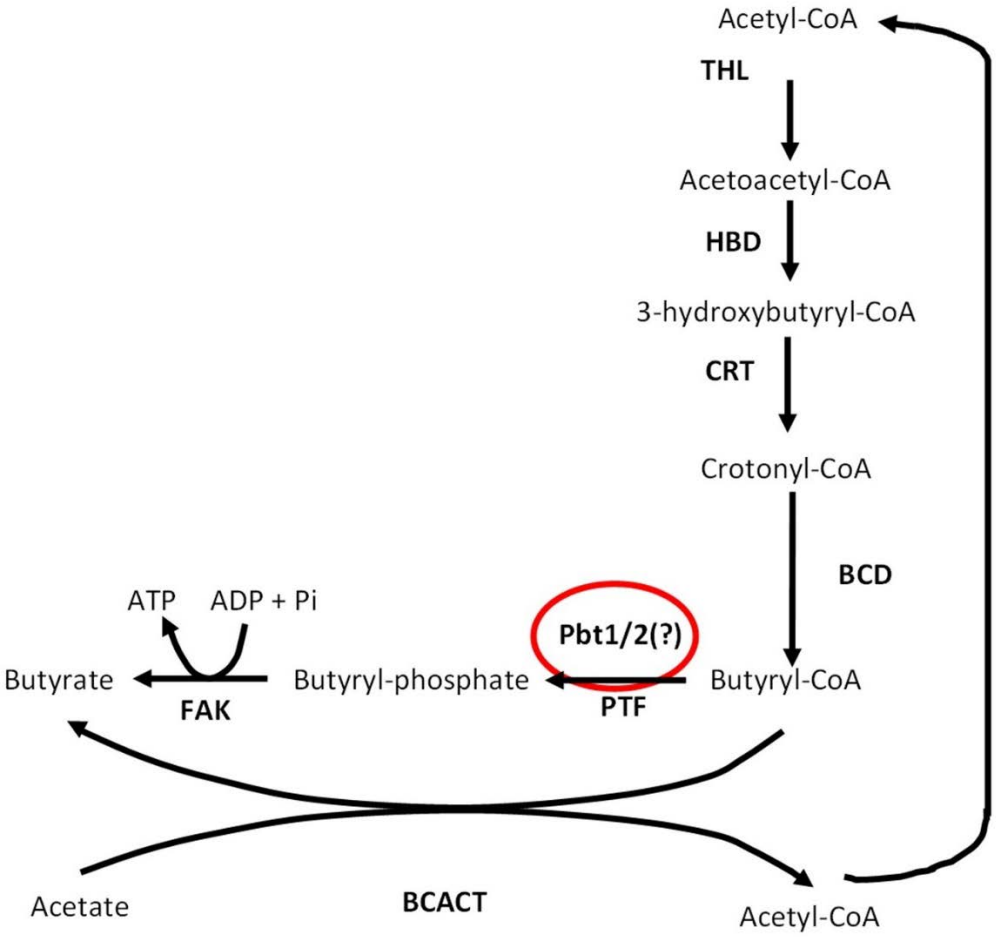

Figure 53 – Putative reaction catalysed by phosphate butyryltransferase.

BCACT butyryl CoA:acetate CoA transferase; CRT, crotonase; FAK, fatty acid kinase; HBD, 3-hydroxybutyryl-CoA dehydrogenase; BCD, butyryl-CoA dehydrogenase; PTF, phosphotransferase; THL, thiolase

As a follow-up to the work done with *pbt1*, *pbt2* (WP_052237430.1) became a target for deletion (Figure 53), both in a wild-type NG-6849 background and in a NG-6849 Δ*pbt1* mutant. By deleting both genes, the possibility of *pbt2* being upregulated to take over the function of *pbt1* is mitigated. The CRISPR-Cas9 knockout vector, pSBRCas9-Cpf-*pbt2*, was created (2.6.1), making use of a synthesised guide and homology arm set and knockout of *pbt2* attempted as detailed previously (2.6.5.1).

However Δ*pbt2* mutant could not be obtained using this CRISPR-Cas9 system despite several different guides being employed.

4.3.3 CRISPR-Cas9 plasmid loss issues

As noted throughout the CRISPR-Cas9 knockout section, an unusual effect was noticed. It appeared that in the case of some mutants (4.3.2.3, 4.3.2.4, 4.3.2.5), a reversion to a native genotype occurred when cells were placed under selective pressure that selected for the native genotype. While several hypotheses were in play when this was first identified (4.3.2.3.1) it seems unlikely that spontaneous mutations could occur in the case of multiple independent knockouts or alternative genes replace function in all cases. Further neither of those explanations explained the colony PCRs which showed a return of cells which were previously identified as pure mutants to a wild type genotype; the absence of wild type bands despite 35 cycles in the prior colony PCRs suggests wild type contamination was also unlikely. Therefore, the working hypothesis became that a minority of a mutant population retained the post-allelic exchange vector, allowing for reintegration and restoration to a native genotype. Once this occurred, the now wild-type cell line would dominate in the selective conditions. The presence of the plasmid was confirmed using colony PCRs, using the primers Cas_pRECAS_seq_F and pCB102_pRECAS_seq_r, on cells which displayed a pure knockout genotype (Figure 54).

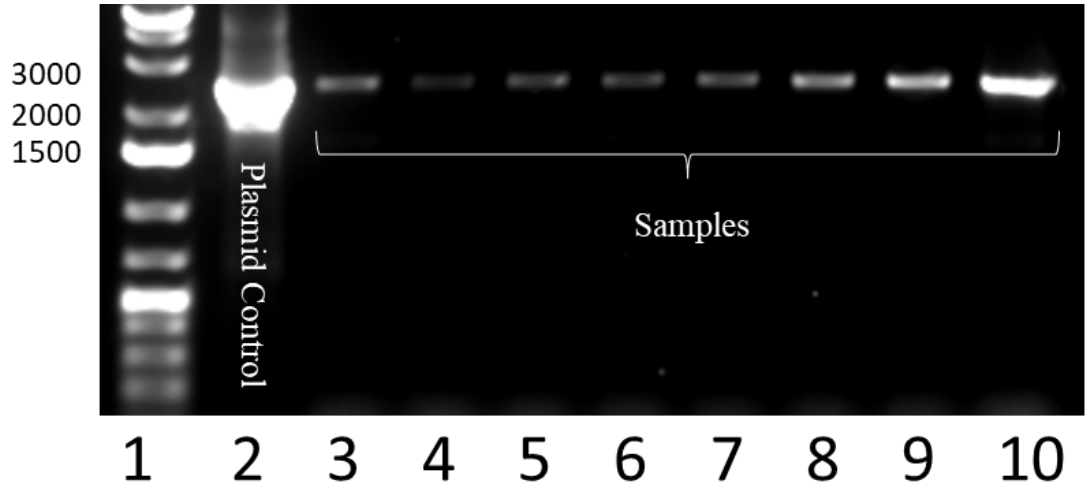

Figure 54 – Screening for plasmid maintenance in E. limosum NG-6849 Δcop by colony PCR using primers directed against the plasmid backbone.

A standard DreamTaq colony PCR was performed on 8 colonies of E. limosum Δcop. Primers binding to the pSBRCas9-Cpf-cop backbone (Cas_pRECAS_seq_F and pCB102_pRECAS_seq_r) were used to internally screen for maintenance of the plasmid. A control of pSBRCas9-Cpf-act DNA was used (lane 2). The ladder used (in lane 1) was GeneRuler 1 kb DNA Ladder, samples were run on a 1% agarose gel. All samples (lanes 3 – 10) showed products of the same size as the pMTL plasmid control.

Screening (Figure 54) identified the maintenance of a plasmid compatible with the primers used, likely pSBRCas9-Cpf-*cop*, in *E. limosum NG-6849* Δ*cop* cells generated using CRISPR-Cas9. As it seemed that plasmids were being retained, questions arose as to why they were not lost. After allelic exchange, the presence of the target on the plasmid should lead to plasmid destruction, as sgRNA would target Cas9 to the plasmid site of the target and thus cause a double-stranded break and destroy the plasmid. It seemed likely, therefore, that a mutation was occurring in the plasmids, rendering this process non-functional. Plasmids were extracted (2.2.2) from Δ*cop* (4.3.2.3), Δ*mti* (4.3.2.4) *and* Δ*mtiI* (4.3.2.5) mutants of *E. limosum*. Plasmids were sequenced (2.2.7), but due to poor quality reads, the resultant sequencing data was of no use, with even the best being unable to align well against any in silico plasmid sequences. While this could mean that the mutations are plentiful in the plasmid and affecting many sites at which sequencing primers bind, it does not reconcile with the issues surrounding reads of regions. If a primer bound to a region in PCR and produced a clean product, it would stand to reason that the product would have a region compatible with the primer, yet sequencing did not show this.

The discovery of plasmid maintenance in Δ*cop* and Δ*mtiI* mutants (4.3.2.3 and 4.3.2.5 respectively), combined with an inability for Δ*mti* (4.3.2.4) Δ*metVmetF* (4.3.2.7) and to lose its plasmid (4.3.2.4), raised questions regarding all the CRISPR-Cas9 mutants created. Hence, all were assessed to determine if they were maintaining plasmids. DNA obtained from cells harvested from the Δ*act* phenotypic assay was amplified using colony PCR (Figure 55) to screen for the presence of *act* using primers, act internal_forward and act internal_reverse, that bound within the gene. These cells had previously displayed a knockout of *act* at its native location.

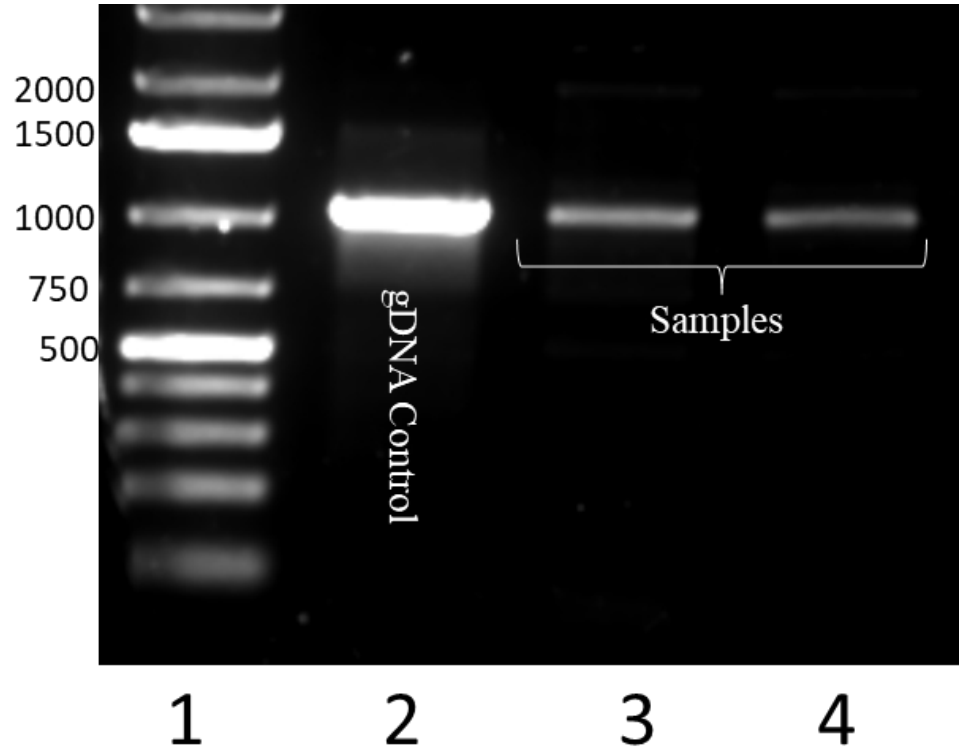

Figure 55 – Screening for the presence of act by colony PCR with internal primers in suspected E. limosum NG-6849 Δact mutants

A standard DreamTaq colony PCR was performed on 2 colonies of E. limosum ⊿act. Primers (act internal_forward and act internal_reverse) were used to internally screen for a knockout of act (lanes 3 – 5). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder. All internal PCRs showed the gene present, suggesting a maintenance of genes on the plasmid.

With screened cells displaying retention of the *act* gene somewhere within the genome in at least some cells of the colonies, they were plated onto standard selection plates (2.1.1.2) and were left to grow for a week. No colonies arose for any cells grown, suggesting that retained plasmid, if present, was unable to confer antibiotic resistance. It is possible that these plasmids were being maintained in cells at very low

levels or existing as a mixture of integrated and free-living plasmid. This is evidenced by the growth seen with Δ*cop* and Δ*mtiI* mutants (4.3.2.3 and 4.3.2.5, respectively), with Δ*cop* and Δ*mtiI* cells requiring reintegration of the genes to grow on methanol. However, it is worth noting that Δ*hisI* mutant (4.3.2.2) did not behave similarly, with no growth on histidine deficient medium occurring.

In the case of Δ*cop* and Δ*mtiI* mutants despite the plasmid loss issue, it is likely that the long lag observed in the pure mutant genotype condition partially confirms the expected phenotype of being unable to grow without those genes. However, to ensure reliability of data further experiments would need to be carried out using a more robust genetic modification system.

4.3.3.1 Removal of retained plasmids

The retention of Cas9 knockout plasmids following gene deletion complicates the creation of additional knockouts in a mutant strain. To address this a modified CRISPR-Cas9 knockout plasmid, with the homology arms and Gram-positive replicon removed, was created. These modifications to the plasmid would result in a suicide plasmid in a Gram-positive organism. This vector has two effects: it kills any cells maintaining a native genotype and destroy any retained plasmids that contain the target genes. The plasmid was created using NEBuilder® HiFi DNA Assembly (2.2.9), removing the homology arms and Gram-positive replicon (Figure 56).

The Δ*act* mutant background was chosen for this piece of work due to the importance of getting a clean knockout to allow for the creation of multiple mutations in the same background. NEBuilder® HiFi DNA Assembly (2.2.9) was used to create the suicide variant of the *act* CRISPR-Cas9 knockout vector pSBRCas9-Cpf-*act* using the primers *act* suicide_fwd and *act* suicide_rev. Sequencing of the resultant plasmid was performed (2.2.7) using the primer Cas_pRECAS_seq_F/JM115 and suicide vector seq reverse primers. The resulting plasmid was introduced into *E. coli* (2.2.1.1) and the standard procedure of screening colonies performed (2.2.2).

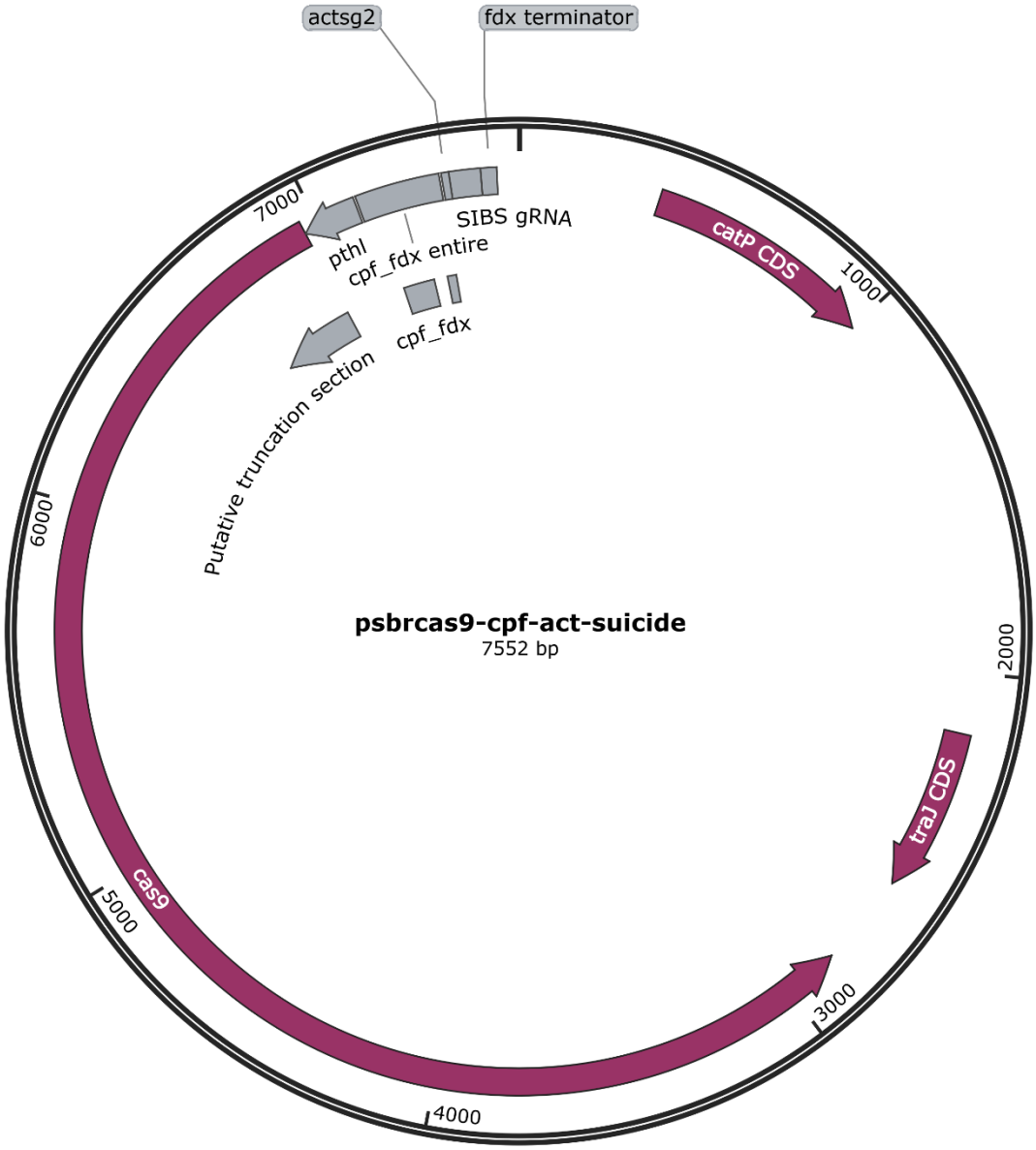

Figure 56 - Plasmid map of pSBRCas9-Cpf-act-suicide

Modified pSBRCas9-Cpf-act lacking homology arms and gram-positive replicon. C. perfringens P_{fdx} drives guide RNA production. Guide RNA will target atc and destroy any plasmid containing it and killing any cell maintaining it. Gram-negative replicon is maintained to allow for maintenance in E. coli.

The CRISPR-Cas9 derived Δ*act E. limosum* NG-6849 strain (4.3.2.1) was transformed (2.3) with the confirmed suicide plasmid, known as pSBRCas9-cpf-*act*-suicide. PCR was used to screen cells (Figure 57) with purpose of investigation of plasmid retention. The primers Cas_pRECAS_seq_F/JM115 and pCB102_pRECAS_seq_r/ JM116 were used as they bind within the plasmid and

should produce a product even if the region between them were to vary considerably due to mutations. However, such a PCR would not be able to ensure that the screened cells were mutants. Therefore, external screening primers were also used in addition to assess the genotype using the same Δ*act* cells used for the plasmid screening.

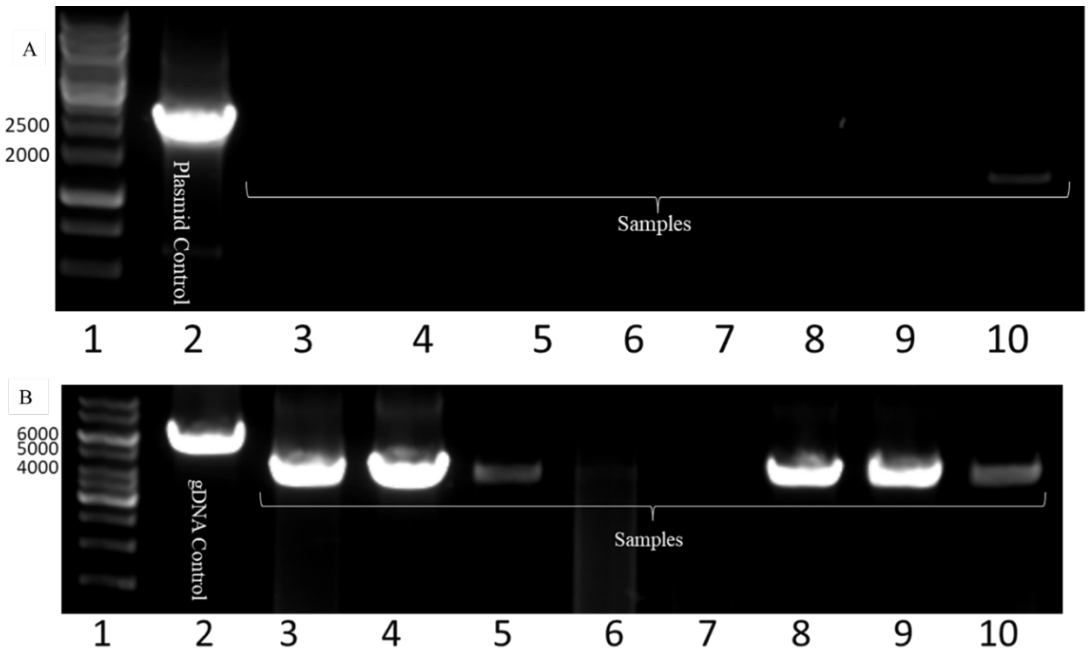

Figure 57 - PCR screen of E. limosum NG-6849 Δact cells for (A) loss of knockout plasmid and (B) retention of act deletion following transformation with pSBRCas9-Cpf-act-suicide.

- **A.** *The ladder used (in lane 1) was GeneRuler 1 kb DNA Ladder. Lane 2 contained a pSBRCas9-Cpf-act control. Lanes 3 - 10 contained the screened cells. Reaction was amplified by a standard using the primers Cas_pRECAS_seq_F/JM115 and pCB102_pRECAS_seq_r/JM116, producing product ~2.5kb in length in case of the maintained plasmid. As observed in lanes 3 - 10 no cells retained plasmid, although lane 10 did show some banding, suggesting regions may have been retained (perhaps partially integrated into the genome).*
- **B.** *The ladder used (in lane 1) was GeneRuler 1 kb Plus DNA Ladder. Lane 2 contained a wild-type control. Lanes 3 - 10 contained the screened clones. Reaction was amplified by a standard using the primers act TAT SCR F and act TAT SCR R, producing a wild-type product ~6kb in length and a deletion product of ~5kb. Lanes 3, 4, 5, 8, 9 and 10 displayed clear deletion bands.*

Colony PCRs of pSBRCas9-Cpf-*act*-suicide transformed Δ*act* mutants (Figure 57) did not show banding consistent with plasmid retention and showed a lack of *act* at the native location (Figure 57). This suggests that a suicide-based system, utilising a modified version of a functional CRISPR-Cas9 knockout vector, could work to remove retained plasmids. Furthermore, it suggests that issues arising are due to a lack of Cas9 activity, strongly indicating that mutations arise in the plasmid borne Cas9 in *E. limosum* in a minority of the population. This is perhaps unexpected given

that the constitutive Cas9 used as part of this work was mutated by *E. coli*, likely meaning that an induced system would be a better choice for CRISPR-Cas9 to work in *E. limosum*.

While this suicide system deals with the issue of retained plasmids in existing mutants, it is not a solution to the problem but rather a remediation of failure. Initial CRISPR-Cas9 knockouts must be performed prior to a second transformation in a pure mutant background, which itself needs further screening. This process would take more than a month for each mutant, introducing the risks of SNPs arising, which would necessitate genomic DNA sequencing to ensure any phenotypical changes observed were due to the knockout. This SNP risk is in addition to the many streakings performed to get to the point of knockout with the CRISPR-Cas9 system as it is. The same would likely be the case for a knock-in. Therefore, an alternative system was needed.

4.3.3.2 Alternative Cas systems

If the issue with the CRISPR-Cas9 system was due to mutations of the plasmid Cas9, then using an alternative Cas system may have been a viable solution. The RiboCas CRISPR-Cas9 system (218) is a CRISPR-Cas9 system that had much success in many other organisms. This system has Cas9 under inducible control, with tight repression when the correct inducer, lactose, is not present. This minimises stress placed upon cells, including *E. coli*, which had previously mutated the constitutively expressed plasmid based CRISPR-Cas9. Such mutations are unlikely in a system with good repression. Despite this system's many potential advantages, function was never observed in *E. limosum*.

The native CRISPR-Cas3 system (1.5.3.3.2) was also investigated after the failures of the RiboCas system. It had shown some promise with *orfD* being deleted (Dr. James Millard, University of Nottingham, personal communication). However, in addition to an inability to replicate this knockout, no other mutants were successfully created with this system (data not shown). As such, a totally different means of counter-

selection was sought, culminating in the use of a suicide plasmid-based, toxin-antitoxin counter-selection system (4.4).

4.3.4 Toxin-antitoxin system

Throughout this work, it became clear that CRISPR-Cas9 had issues in *E. limosum*; an alternative system was, therefore, desirable. Such a system would need to improve upon the existing CRISPR-Cas9 system, allowing for clean deletions, use easy-to-assemble vectors, allow for rapid and reliable genetic modifications, and negate the issues of plasmid retention and eventual genomic reintegration previously deleted genes.

The type I toxin-antitoxin system (referred to as TAKOS for toxin-antitoxin knockout system) developed by Dr. James Millard at the University of Nottingham (227) drew on the previous use of such systems in other bacteria, notably in the anaerobe *C. difficile* (228). It was based on a toxin-antitoxin system derived from *Eubacterium callanderi* KIST612 (formerly *E. limosum* KIST612). One of its principal attractions was that it made use of suicide vectors and should, therefore, avoid the plasmid retention issues that were seen in CRISPR-Cas9 (4.3.3).

The heart of the system is a tetracycline-controlled inducible promoter that drives the expression of the encoded toxin gene *relE*. The *relB* antitoxin gene is driven by the same constitutive promoter that drives the system in *E. callanderi*. Next to it are the homology arms of the target, either for a knockout or a knock-in. This is then followed by a Gram-negative replicon, which is essential for maintenance in *E. coli*, and an antibiotic resistance gene that allows for selection. The structure of the plasmids was as follows (Figure 58). Due to this system being tetracycline-controlled, anhydrotetracycline was supplemented into the medium after the initial transformation (2.1.1.2.1). With the issues that had arisen with the CRISPR-Cas9 system (4.3.3), all CRISPR-Cas9 mutant generation was replicated using the toxin-antitoxin system, and phenotype characterisations were repeated.

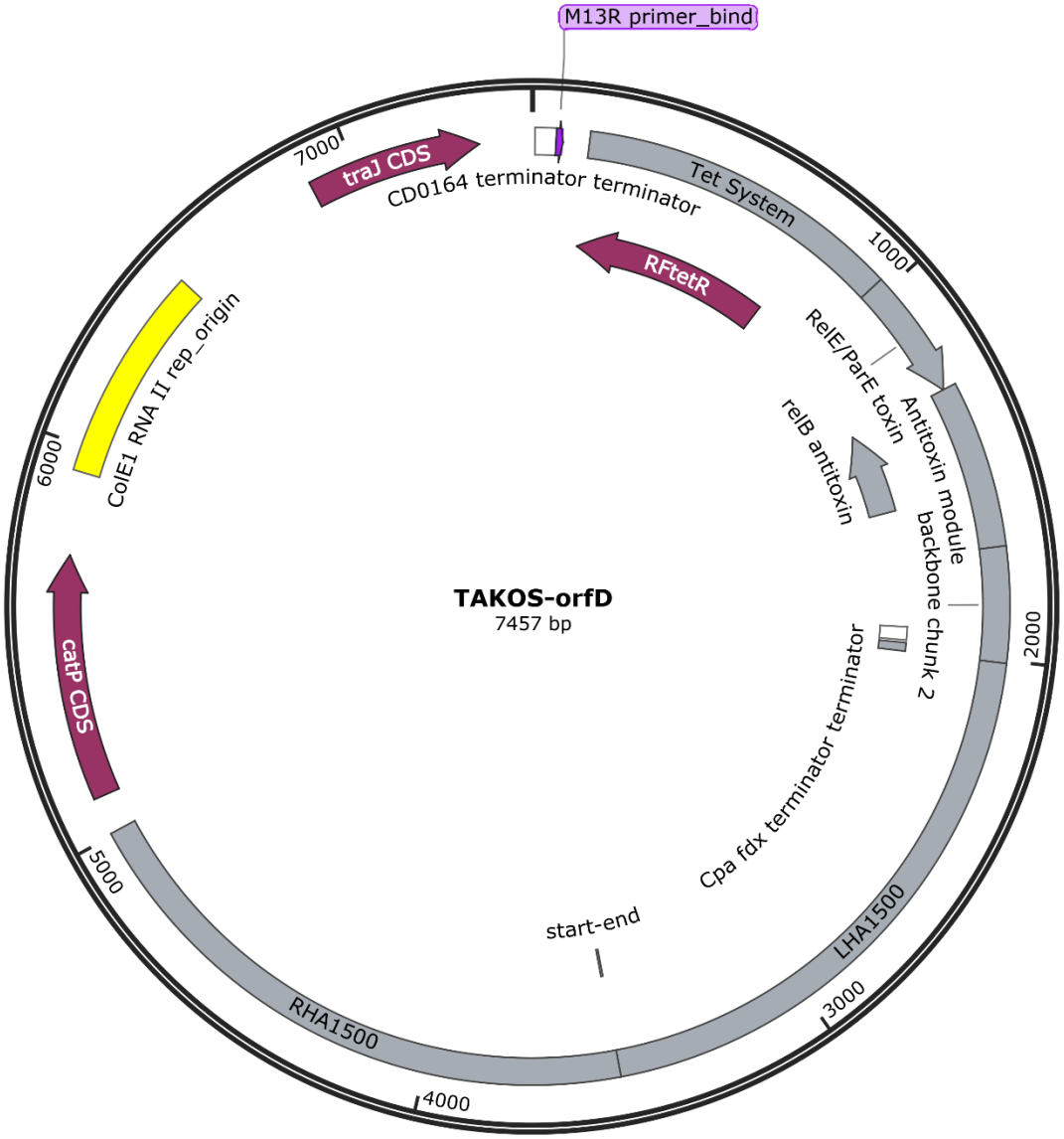

Figure 58 - Plasmid map of toxin-antitoxin plasmid TAKOS-orfD for the deletion of orfD

4.3.4.1 Generation of *E. limosum* NG-6849 Δ*orfD* mutants using toxin-antitoxin system

To ensure the toxin-antitoxin system functioned correctly, the protocol was reliable, and the results obtained by a prior researcher (Dr. James Millard, University of Nottingham) were repeatable, a knockout using a previously constructed vector, TAKOS-*orfD*, was performed. The procedure for induction was followed as previously stated (2.1.1.2.1), and colony PCRs (2.5.2.3) were performed (Figure 59) with the primers orfD TAT SCR F and orfD TAT SCR R.

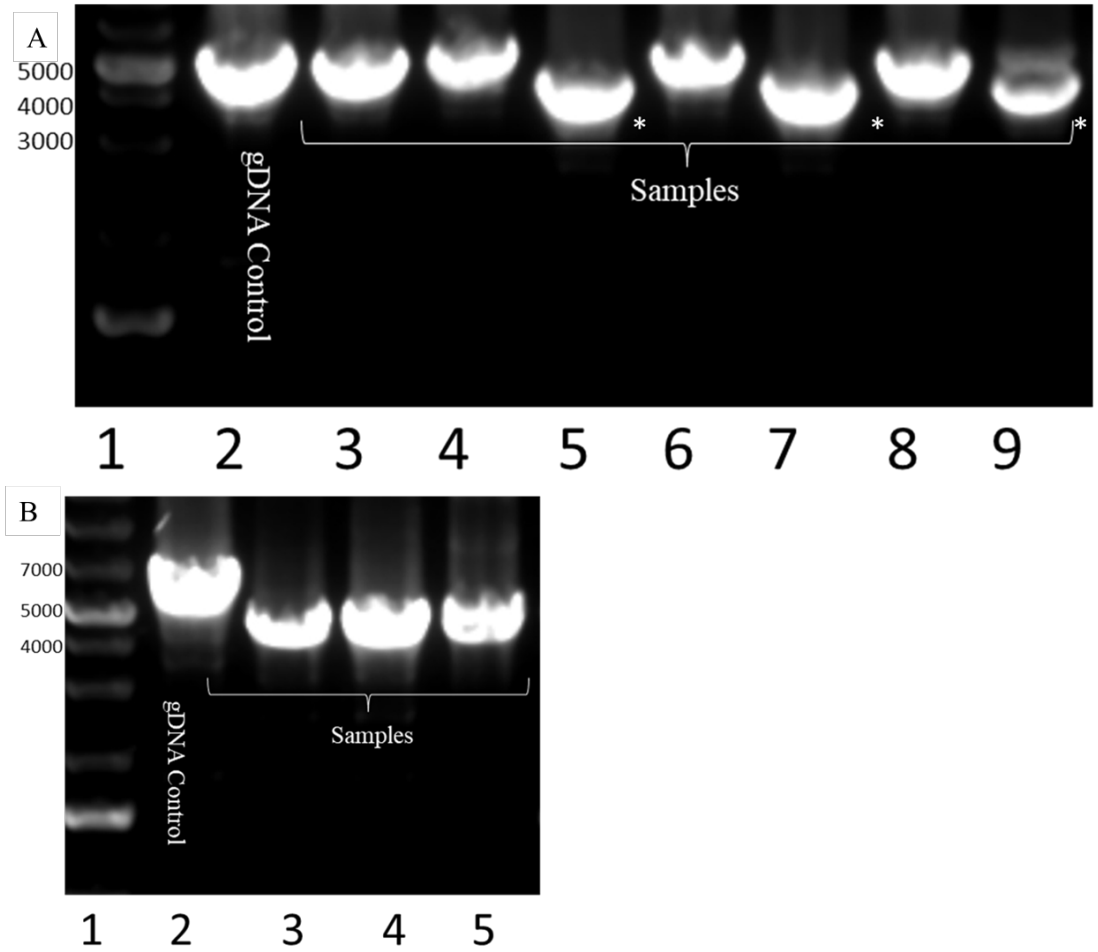

Figure 59 - Screening E. limosum NG-6849 for the deletion of orfD by colony PCR

A standard DreamTaq colony PCR was performed on E. limosum NG-6849 transformed with TAKOS-orfD. Primers (orfD TAT SCR F and orfD TAT SCR R) were used to screen for a chromosomal knockout of the orfD gene. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder.

- *A. 1% agarose gel displaying 7 colonies (lanes 3 - 9) were screened for knockout of orfD. wild-type bands (of ~5 kb) appeared in all screened colonies except for lanes 5, 7 and 9 where a smaller band of the expected deletion size (~4.2 kb) was observed; putative mutants are highlighted with "*" next to the band.*
- *B. 0.8% agarose gel displaying the 3 previously screened, "*", colonies putatively identified as ΔorfD (Figure 59A) were rescreened. Mutant bands (~4.2 kb) were observed in all screened colonies with wild-type bands (of ~5 kb) not appearing in any, indicating pure deletions.*

Screening of cells transformed with the TAKOS-*orfD* vector identified three pure mutants (Figure 59A). However, due to the product size and agarose gel percentage (1%), the mutant products were very close to that of the wild type. The three apparent pure Δ*orfD* mutant colonies were screened again using colony PCR (Figure 59B) and run on a 0.8% agarose gel to try to make the difference more distinct for easier identification between the wild-type and mutant products. The second round of colony PCR screening (Figure 59B) using a 0.8% agarose gel allowed for easier

identification of mutant bands. 0.8% agarose would be used for future toxin-antitoxin system-generated mutants and primers were designed to minimize this issue. Internal screening for *orfD* was performed (Figure 60), with the primers orfD_Int2_F and orfD_Int2_R, to ensure no reintegration elsewhere in the genome. It would also screen for plasmids that had undergone homologous recombination and contained the wild-type gene region. This should be minimal due to the suicide nature of the plasmids used.

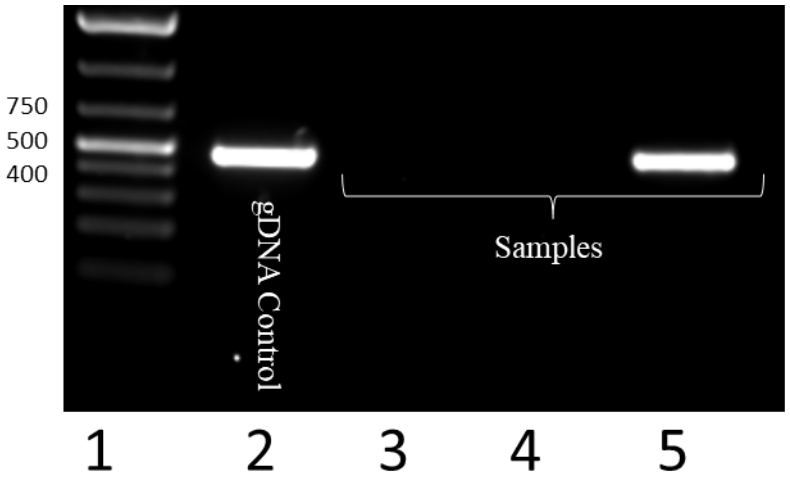

Figure 60 - Screening E. limosum NG-6849 for the deletion of orfD by colony PCR

A standard OneTaq colony PCR was performed on 3 colonies of E. limosum NG-6849 transformed with TAKOS-orfD. Primers (orfD_Int2_F and orfD_Int2_R) were used to screen for a chromosomal presence of orfD (lanes 3 - 5). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 0.8% agarose gel. wild-type bands (of ~500 b) appeared in lanes 2 and 5, but not in 3 and 4, suggesting that mutants 3 and 4 were indeed pure knockouts.

Internal screening (Figure 60) of colonies identified as Δ*orfD* mutants using external primers showed that 2 out the 3 colonies were pure mutants with no retention of *orfD* either on a retained plasmid or reintegrated elsewhere in the genome. The colonies that displayed an apparent mutant band using external primers, and no product using internal primers were stocked (2.1.2.5).

PCRs were repeated (Figure 61), on different colonies, with an improved set of primers (orfD TAT SCR2 F and orfD TAT SCR2 R), which bound outside one of the homology arms (preventing products from any retained plasmid) and within one of the homology arms close to the gene of interest. This resulted in smaller products which allowed for reliable sanger sequencing of PCR. The improved primers also

eliminated the issue observed in Figure 59, where mutant and wild-type bands were too close on the gel, resulting in colony 7 being erroneously suspected of being a knockout (which Figure 60 showed was not the case).

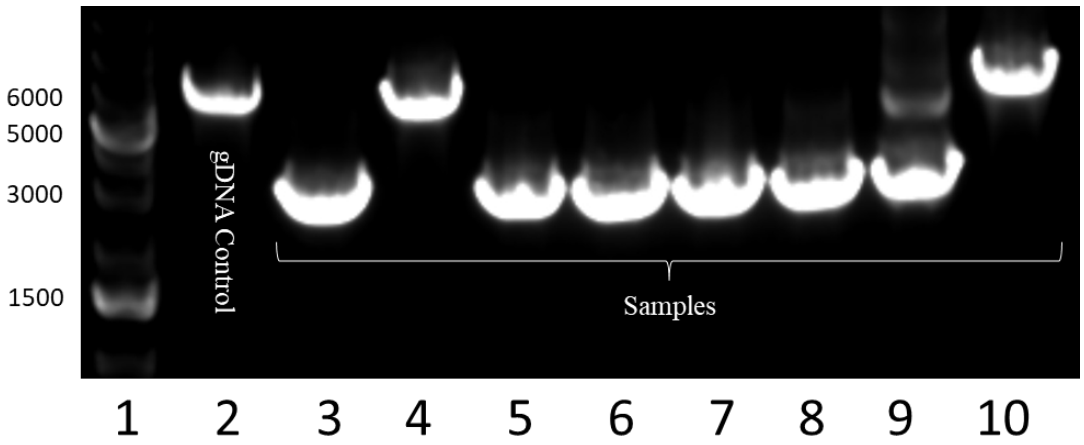

Figure 61 - Screening E. limosum NG-6849 for the deletion of orfD by colony PCR using new primers.

A standard OneTaq colony PCR was performed on 8 colonies of E. limosum NG-6849 transformed with TAKOS-orfD. Primers (orfD TAT SCR2 F and orfD TAT SCR2 R) were used to screen for a chromosomal knockout of the orfD (lanes 3 - 10). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 1% agarose gel. wild-type bands (of ~500 b) appeared in lanes 2 and 5, but not in 3 and 4, suggesting that mutants 3 and 4 were indeed pure knockouts.

The improved primers allowed for easier identification of Δ*orfD* mutants. This colony PCR also demonstrates the significant impact that selection bias can have on the apparent efficiency of a mutation, with only 2 out of 7 colonies being successful based on the screening shown in Figure 59A, and 6 out of 8 being successful based solely on random chance as shown in Figure 61. Even with this caveat, the efficiency was much better than what was observed with CRISPR-Cas9 (4.3.3), and the mutants obtained more rapidly. Within a week of the introduction of the plasmid, pure deletions were identified and isolated.

Due to the vector lacking a Gram-positive replicon, plasmid loss is all but guaranteed. Indeed, no cells that had undergone induction (successfully mutating or otherwise) were able to grow on antibiotic selection plates, indicating that the frequency of escape from toxin counterselection was low. Once it was established that the method worked well and results were repeatable, work began to knockout targets of interest. It is worth noting that when TAKOS-*orfD* was sequenced, a coding change in the anti-

toxin gene was noted, yet function was still retained, suggesting some robustness in the system to mutations within functional regions.

4.3.4.2 Re-generation of *E. limosum* NG-6849 CRISPR-Cas9 mutants using the toxin-antitoxin system

As detailed previously the CRISPR-Cas9 system was not fit for purpose. The issues of plasmid retention and genomic reversion made CRISPR-Cas9 mutants unreliable for phenotype assessment. Furthermore, the inability to generate multiple knockout backgrounds with CRISPR-Cas9 (4.3.2.8) necessitated the recreation of mutants for which multiple mutant backgrounds were desired, such as the Δ*act* mutant, with the toxin-antitoxin system. Knockouts were attempted as detailed previously (2.6.5.2).

4.3.4.2.1 Re-generation of *E. limosum* NG-6849 Δ*act* using the toxin-antitoxin system

To confirm the Δ*act* knockout two rounds of colony PCR screening was required. The initial screening (Figure 62A) displayed a single pure mutant, which was further confirmed as being pure by a restreak to single colonies and rescreening (Figure 62B), resulting in the identification of *act* knockout mutants. With the confirmation of deletion, cells were stocked (2.1.2.5). This knockout was to be used as a background for further knockouts; as such, competent cells were made using these cells (2.3).

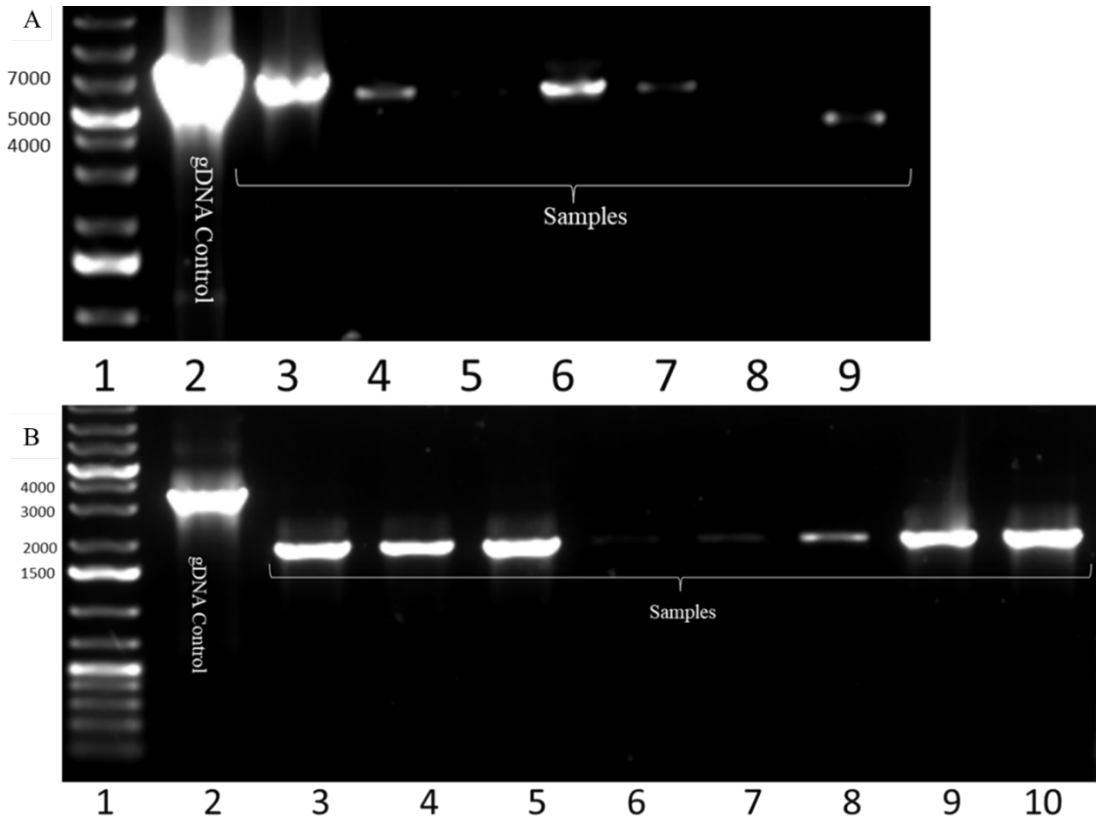

Figure 62 - Screening E. limosum NG-6849 for the deletion of act by colony PCR

A standard DreamTaq colony PCR was performed on E. limosum NG-6849 transformed with TAKOS-act. Primers (act TAT SCR F and act TAT SCR R) were used to screen for a chromosomal knockout of the act in gel 1 and primers (act TAT SCR2 F and act TAT SCR2 R) were used for gel 2. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 0.8% agarose gel.

- *A. 7 colonies (3 - 9) were screened for the knockout of act. wild-type bands (of ~6 kb) appeared in all screened colonies but lane 9 where a smaller band, of the expected deletion size (~4.4 kb) was observed. A single Δact pure knockout was identified in the initial round of screening. However due to the poor-quality image the colony was streaked out to single colonies on plates and single colonies were rescreened to confirm the presence of mutation (Figure 62B).*
- *B. 8 colonies (3 - 10) were rescreened for the knockout of act bands of the expected deletion size (~2 kb) were observed in all lanes but the control which displayed a wild-type band (of ~3.5 kb).*

4.3.4.2.2 Re-generation of *E. limosum* NG-6849 Δ*cop* using the toxin-antitoxin system

The initial knockout attempt performed using TAKOS-*cop* was unsuccessful in generating any Δ*cop* mutants (Figure 63A). As a result, knockouts were attempted again (2.6.5.2) using an alternative vector, TAKOS-*cop*-Short, which used shorter homology arms (0.5 kb instead of the 1.5 kb used previously). The rationale behind this was that the longer homology arms may have contained open reading frames that caused issues. The homology arms were treated as detailed previously (2.6.2) and then introduced into *E. limosum* (2.3), followed by inducement of transformed cells and screening (Figure 63B) with cop TAT SCR F and cop TAT SCR R.

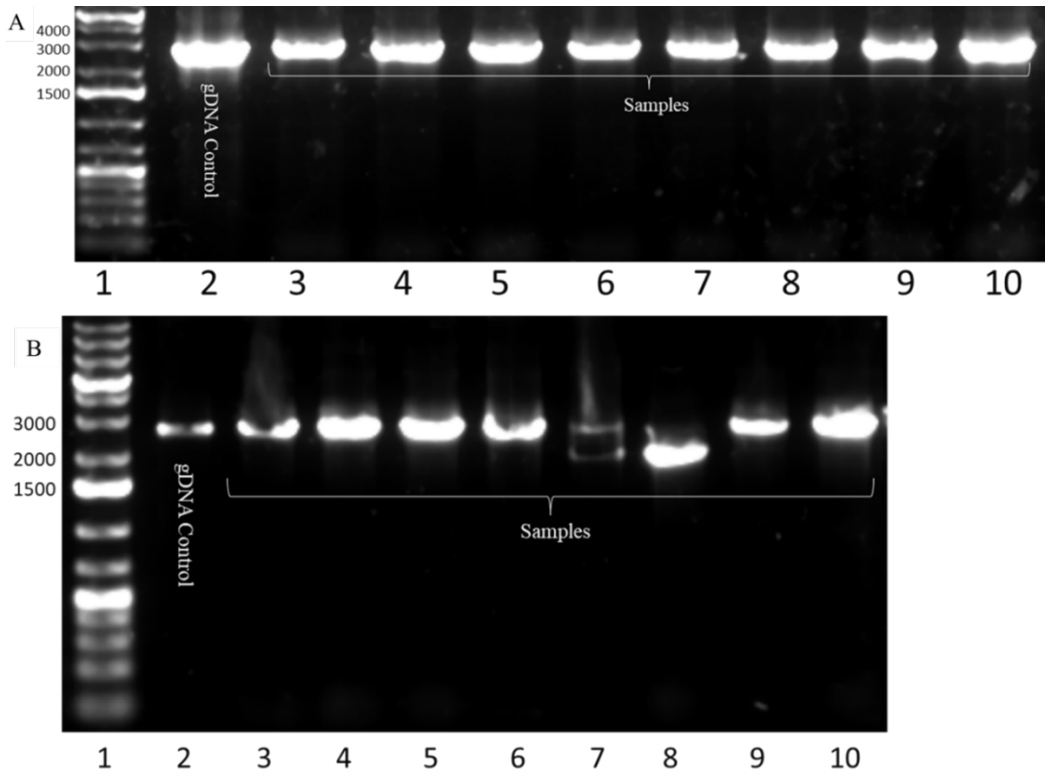

Figure 63 - Screening E. limosum NG-6849 for the deletion of cop by colony PCR

A standard DreamTaq colony PCR was performed on E. limosum NG-6849 transformed with TAKOS-cop (A) and TAKOS-cop-Short (Figure 63B). Primers (cop TAT SCR F and cop TAT SCR R) were used to screen for a chromosomal knockout of the cop. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 0.8% agarose gel.

- ***A.*** *8 colonies (3 - 10) were screened for knockout of cop. All cells displayed a wild-type genotype. Colonies were screened using colony PCR (2.5.2.3.2) with the screening primers cop TAT SCR2 F and cop TAT SCR2 R.*
- ***B.*** *8 colonies (3 - 10) were screened for knockout of cop. The cells in lane 8 displayed a mutant band, the rest displayed a wild-type genotype.*

Screening identified a single mutant of Δ*cop*. With the confirmation of the successful knockout generation, cells were stocked (2.1.2.5), and starter cultures were prepared for assay (4.3.4.9). The fact that smaller homology arms worked where a larger one did not may suggest that open reading frames are causing issues. However, in silico analysis of the homology arms did not suggest any toxic products from any open reading frames. This suggests that the increased transformational efficiency, of a smaller vector, may have been the cause of the difference seen.

4.3.4.2.3 Re-generation of *E. limosum* NG-6849 Δ*mti* using the toxin-antitoxin system

Generation of Δ*mti* using the toxin-antitoxin system was highly successful, with initial colony PCR screening (Figure 64), with the primers mti TAT SCR2 F and mti TAT SC2R R, of cells transformed with TAKOS-mti displaying many pure mutants. Successful pure mutants (Figure 64, lanes 3 and 5) were stocked (2.1.2.5) to await future work (4.3.4.9).

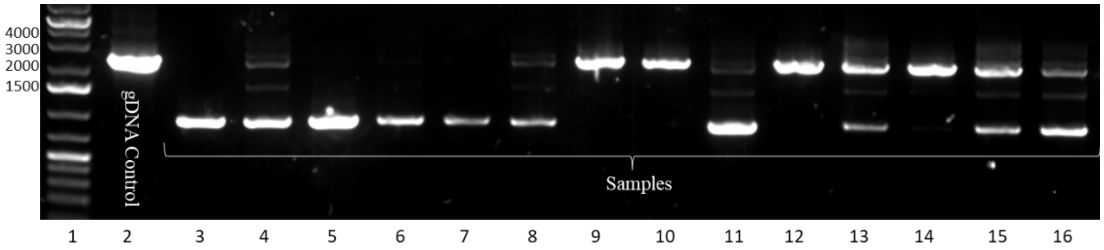

Figure 64 - Screening E. limosum NG-6849 for the deletion of mti by colony PCR

A standard OneTaq colony PCR was performed on 14 colonies of E. limosum NG-6849 transformed with TAKOS-mti. Primers (mti TAT SCR2 F and mti TAT SC2R R) were used to screen for a chromosomal knockout of the mti (lanes 3 - 16). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 0.8% agarose gel. wild-type bands appeared in ~50% screened colonies, with most others containing two bands. Lanes 3 and 5 contained no wild-type band (when over exposed to reveal faint bands), only the mutant band and thus were pure knockouts, the difference in band size corresponds to the size of the deleted gene (~1.3 kb).

4.3.4.2.4 Re-generation of *E. limosum* NG-6849 Δ*mtiI* using the toxin-antitoxin system

The initial attempt to produce Δ*mtiI* mutants using TAKOS-*mtiI* was unsuccessful (Figure 65A). As a result, an alternative vector (2.6.5.2), TAKOS-*mtiI*-Short, which had shorter homology arms, was created, and used to attempt the knockout (2.6.5.2). Mutant screening using colony PCR (Figure 65B), with the primers mtiI TAT SCR2 F and mtiI TAT SCR2 R, identified five mutant Δ*mtiI* colonies. Successful Δ*mtiI* knockout cells were stocked (2.1.2.5), and starter cultures were prepared for phenotype analysis (4.3.4.9).

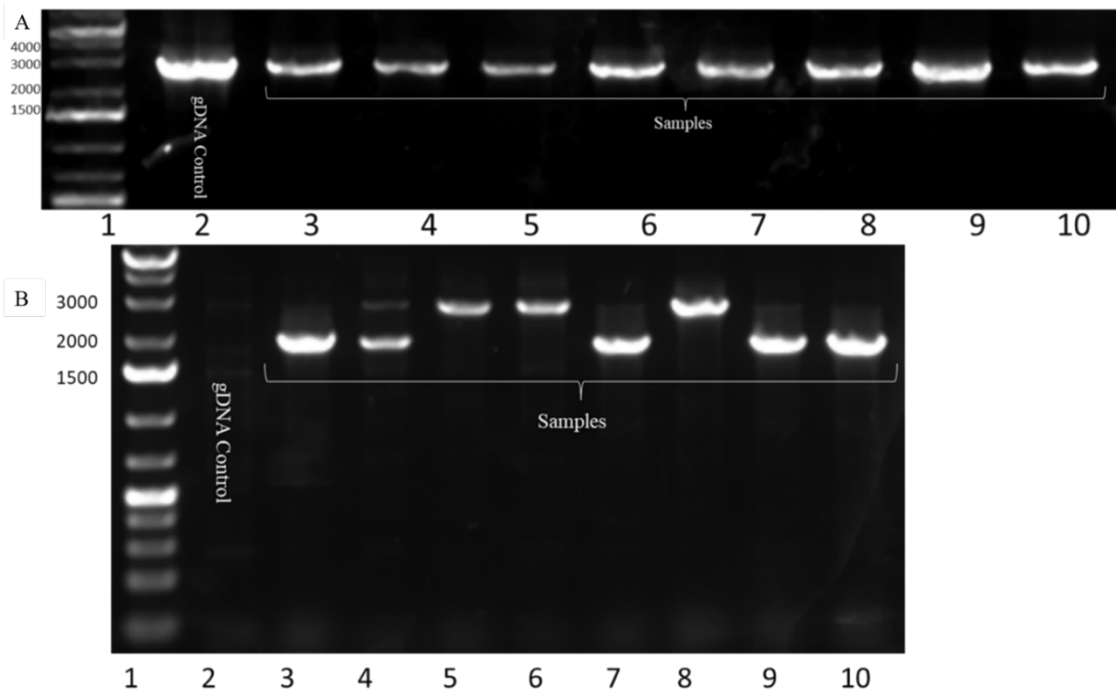

Figure 65 - Screening E. limosum NG-6849 for the deletion of mtiI by colony PCR

A standard DreamTaq colony PCR was performed on E. limosum NG-6849 transformed with TAKOS-mtiI (A) and TAKOS-mtiI-Short (B). Primers (mtiI TAT SCR2 F and mtiI TAT SCR2 R) were used to screen for a chromosomal knockout of the mtiI. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 0.8% agarose gel.

- ***A.*** *8 colonies (lanes 3 - 10) were screened for a knockout of mtiI. All cells displayed a wild-type band.*
- ***B.*** *8 colonies (lanes 3 - 10) were screened for a knockout of mtiI. The gDNA lane did not amplify, this is most likely due to a loading error as the PCRs for all samples worked. Lanes 3, 7, 9 and 10 displayed pure knockouts, with lane 4 displaying dual banding or a mixed genotype.*

4.3.4.2.5 Re-generation of *E. limosum* NG-6849 Δ*pbt1* using the toxin-antitoxin system

Colony PCR screening, using the primers pbt1 TAT SCR2 F and pbt1 TAT SC2R R, of cells transformed with the TAKOS-pbt1 plasmid (Figure 66) isolated two pure knockout Δ*pbt1* mutants. These colonies were stored (2.1.2.5) for future work (4.3.4.7).

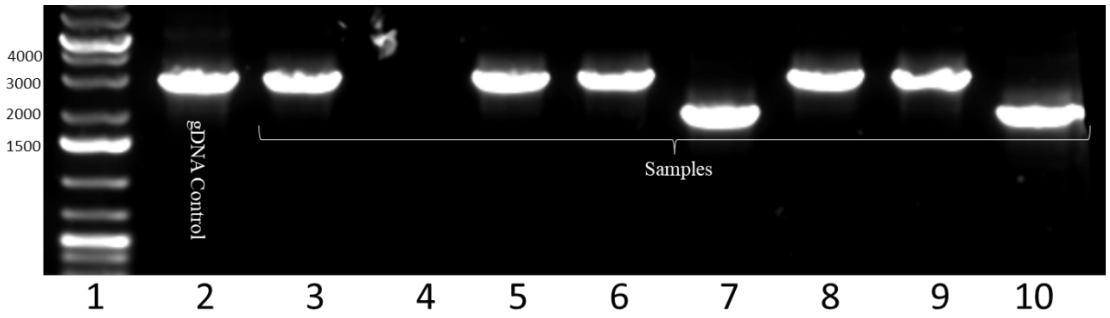

Figure 66 - Screening E. limosum NG-6849 for the deletion of pbt1 by colony PCR

A standard OneTaq colony PCR was performed on 8 colonies of E. limosum NG-6849 transformed with TAKOS-pbt1. Primers (pbt1 TAT SCR2 F and pbt1 TAT SC2R R) were used to screen for a chromosomal knockout of the pbt1 (lanes 3 - 10). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 0.8% agarose gel. wild-type bands (~3.2 kb) appeared in half of the colonies. Lanes 4, 5 8 and 10 no wild-type band, only the mutant band (~2.1 kb) and thus were pure knockouts.

4.3.4.2.6 Re-generation of *E. limosum* NG-6849 Δ*metVmetF* using the toxin-antitoxin system

Despite prior success in the creation of a Δ*metVmetF* mutant (4.3.2.7) multiple rounds of screening of cells transformed with TAKOS vectors designed to knockout *metVmetF* failed to identify any Δ*metVmetF* mutants (Figure 67). The failure to knock this gene out occurred with many different homology arm lengths; a single gel is displayed as all cPCRs were identical in outcome. Given the success of this system with other deletions, it seemed that this gene is essential.

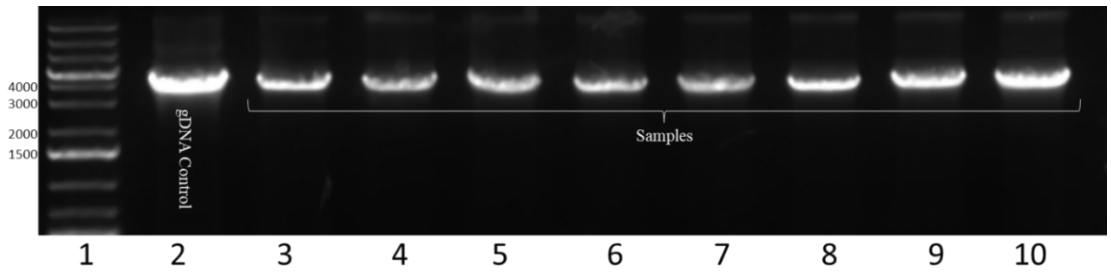

Figure 67 - Screening E. limosum NG-6849 for the deletion of metVmetF by colony PCR

A standard OneTaq colony PCR was performed on 8 colonies of E. limosum NG-6849 transformed with TAKOS-metVmetF. Primers (metVmetF TAT SCR2 F and metVmetF TAT SCR2 R) were used to screen for a knockout of metVmetF (lanes 3 - 10). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder. All cells displayed a wild-type genotype.

Essentiality was not unexpected, as a collaborator (Dr. Helge Dietrich, Goethe University Frankfurt, personal communication) suggested that this knockout should not be possible based on the current understanding of the *E. limosum* Wood-Ljungdahl pathway. Indeed, the failure to create a chromosomal deletion of this gene reinforces the previously mentioned hypothesis regarding plasmid retention with CRISPR-Cas9 being significant and lends weight to the current understanding and model of the Wood-Ljungdahl pathway in *E. limosum*. Essentiality was further confirmed by TraDIS (Dr. Elizabeth Redfern, University of Nottingham, personal communication) data. These data suggested that while the *metV* gene was not essential its partner gene, *metF*, was, which would explain the inability to knock both out. This data was received too late in this project to lead to any lab work.

4.3.4.2.7 Re-generation of *E. limosum* NG-6849 Δ*metF1* using the toxin-antitoxin system

Colony PCR, using the primers metF1 TAT SCR2 F and metF1 TAT SC2R R, was performed on cells transformed with TAKOS-metF1 mutants (Figure 68). Successful pure mutants were identified (Figure 64, lanes 4, 5, 8, and 10) and were stocked (2.1.2.5) to await future work (4.3.4.9).

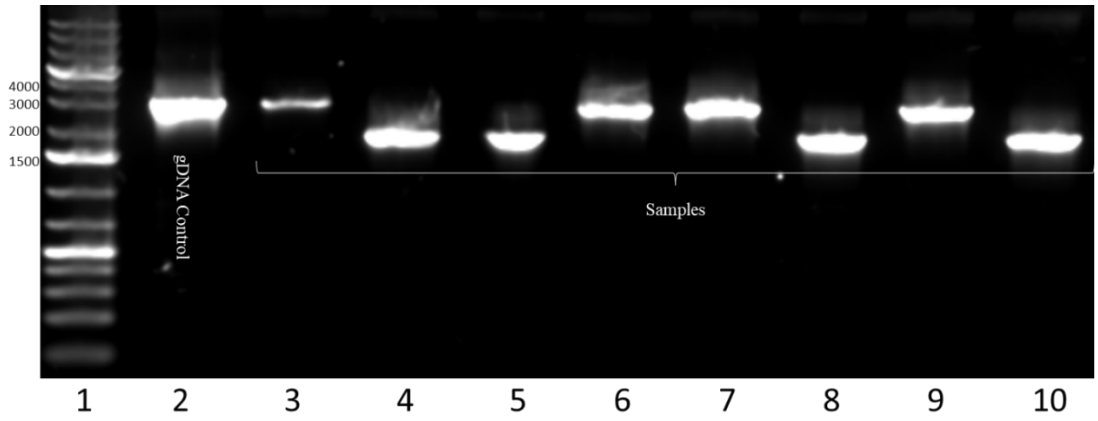

Figure 68 - Screening E. limosum NG-6849 for the deletion of metF1 by colony PCR

A standard OneTaq colony PCR was performed on 8 colonies of E. limosum NG-6849 transformed with TAKOS-metF1. Primers (metF1 TAT SCR2 F and metF1 TAT SC2R R) were used to screen for a chromosomal knockout of the metF1 (lanes 3 - 10). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder. wild-type bands (of ~3 kb) appeared in ~ 50% of screened colonies the others displaying the mutant band (~2.1kb).

4.3.4.3 Generation of *E. limosum* NG-6849 Δ*pbt2* mutants using toxin-antitoxin system

The Δ*pbt2* mutant was not successfully created using CRISPR-Cas9 (4.3.2.8). Therefore, TAKOS was used to attempt to create the mutant. Knockouts were attempted as detailed previously (2.6.5.2) using the plasmid TAKOS-pbt2. Colony PCR screening (Figure 69) with the primers pbt2 TAT SCR2 F and pbt2 TAT SC2R R, identified seven pure knockout mutants; these were stocked (2.1.2.5) for future work (4.3.4.7). The TAKOS-pbt2 vector was then used in a Δ*act* background generated previously (4.3.4.2.1) to perform a double knockout (4.3.4.5).

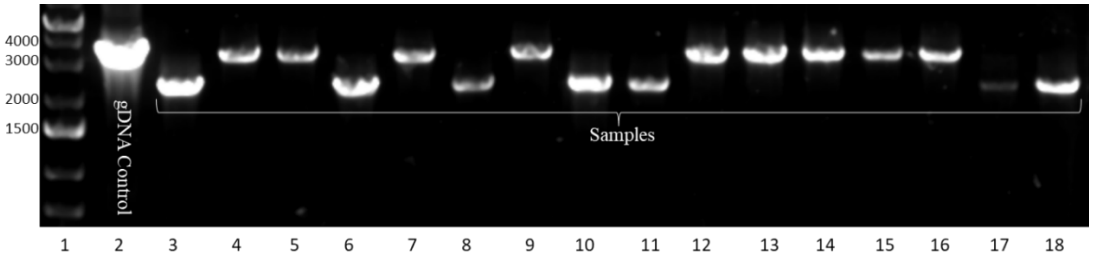

Figure 69 - Screening E. limosum NG-6849 for the deletion of pbt2 by colony PCR

A standard OneTaq colony PCR was performed on 16 colonies of E. limosum NG-6849 transformed with TAKOS-pbt2. Primers (pbt2 TAT SCR2 F and pbt2 TAT SC2R R) were used to screen for a chromosomal knockout of the pbt2 (lanes 3 - 18). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder. wild-type bands (of ~3.2 kb) appeared in ~ 50% of screened colonies the others displaying the mutant band (~2.1kb).

4.3.4.4 Generation of *E. limosum* NG-6849 Δ*act*Δ*pbt1* mutants using toxin-antitoxin system

As discussed previously (4.3.1, 4.3.7, and 4.3.8), elimination of butyrate production was seen as desirable. The elimination of production had been previously unsuccessful. A triple mutant eliminating *act*, *pbt1*, and *pbt2*, was sought, in the hope that deleting all known genes associated with butyrate production would prevent production. Multiple knockout backgrounds were unsuccessful using the CRISPR-Cas9 system (4.3.2.8), likely due to plasmid retention (4.3.3). Both Δ*act*Δ*pbt1* and Δ*act*Δ*pbt2* deletions were attempted in parallel. Having these mutants may help to elucidate whether a triple mutant was necessary. TAKOS-*pbt1* was constructed previously for the creation of the single Δ*pbt1* knockout (4.3.4.2.5). TAKOS-*pbt1* was used to perform a knockout in an *E. limosum* NG-6849 Δ*act* mutant background (2.6.5.2). Colonies were screened via colony PCR (2.6.6) with the primers pbt1 TAT SCR F and pbt1 TAT SCR R (Figure 70).

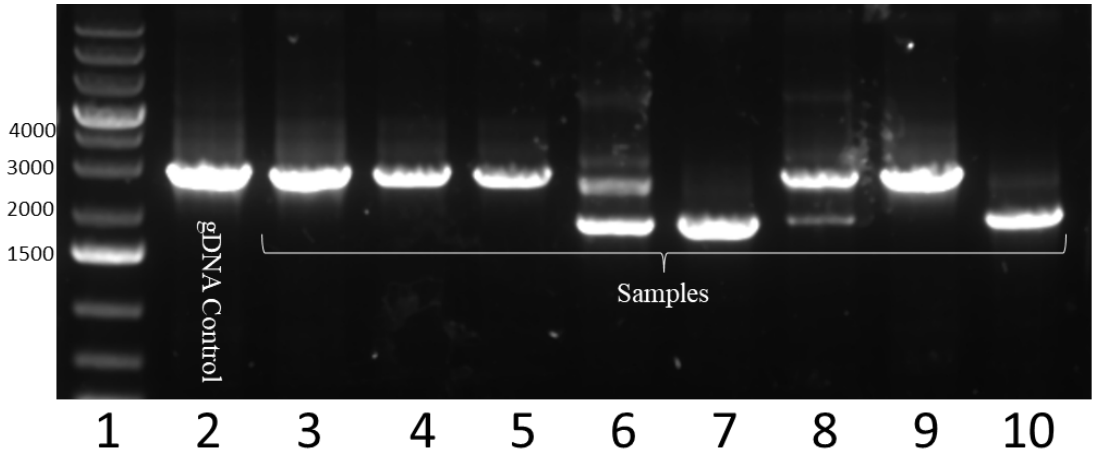

Figure 70 - Screening E. limosum NG-6849 for the deletion of pbt1 by colony PCR in a Δact background

A standard OneTaq colony PCR was performed on 8 colonies of Δact E. limosum NG-6849 transformed with pbt1short-TAKOS. Primers (pbt1 TAT SCR2 F and pbt1 TAT SC2R R) were used to screen for a chromosomal knockout of the pbt1 (lanes 3 - 10). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 0.8% agarose gel. wild-type bands (~3.2 kb) appeared in all but 2 colonies. Lanes 7 and 10 no wild-type band, only the mutant band (~2.1 kb) and thus were pure knockouts.

Screening identified both mixed mutants and pure chromosomal knockouts of *pbt1* (Figure 70). This may suggest that cells within a colony can survive without the

plasmid by growing on top of resistant cells or growing once antibiotic is expended. This would result in both mutant and wild type cells being present within a single colony. This is due to the inability for plasmids to replicate within cells which ensures that many daughter cells will contain no plasmid and so have no resistance to antibiotic.

4.3.4.5 Generation of *E. limosum* NG-6849 Δ*act*Δ*pbt2* mutants using toxin-antitoxin system

The TAKOS-*pbt2* vector was constructed previously for the creation of the single Δ*pbt2* knockout (4.3.4.2.1). and was used to perform a knockout in a *E. limosum* NG-6849 Δ*act* mutant background (2.6.5.2). Colonies were screened (2.6.6) using the screening primers pbt2 TAT SCR F and pbt2 TAT SCR R (Figure 71).

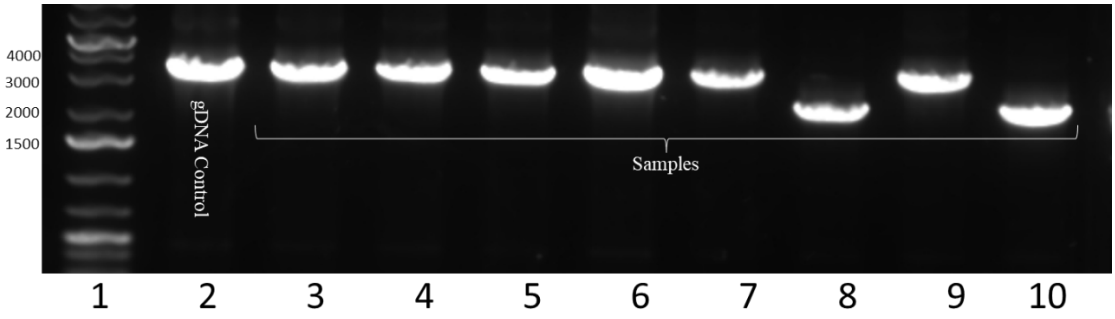

Figure 71 - Screening E. limosum NG-6849 for the deletion of pbt2 by colony PCR in a Δact background

A standard OneTaq colony PCR was performed on 8 colonies of Δact E. limosum NG-6849 transformed with TAKOS-pbt2. Primers (pbt2 TAT SCR2 F and pbt2 TAT SC2R R) were used to screen for a chromosomal knockout of the pbt2 (lanes 3 - 10). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder. wild-type bands (~3.2 kb) appeared in all but 2 colonies. Lanes 8 and 10 no wild-type band, only the mutant band (~2.1 kb) and thus were pure knockouts.

Unfortunately, when attempting a triple knockout, using cells prepared from the initial Δ*act*Δ*pbt2* knockout, isolated colonies did not form on selection plates; rather, a contiguous layer of cell mass was observed. This suggests that antibiotic resistance was present in the cells prior to the transformation of the TAKOS-*pbt1*-short plasmid. Patch plating of the Δ*act*Δ*pbt2* knockout single colonies was performed, with cells growing readily on selection. This implied antibiotic resistance was conferred to the

cells. All cells transformed with TAKOS-*pbt2* were similarly tested, but none, other than the initial Δ*act*Δ*pbt2* knockout, displayed this phenotype.

A vector with smaller homology arms, in line with many of the other mutants, was created. The rationale behind this was that the longer homology arms may have allowed for integration of the resistance cassette into the genome. This vector, TAKOS-*pbt2*short, was introduced into a Δ*act* background (4.3.4.2.1) to see if this phenotype was replicated. This transformation proceeded as normal, with distinct colonies of cells forming. Cells were then screened by colony PCR to determine the success of chromosomal knockout (Figure 72).

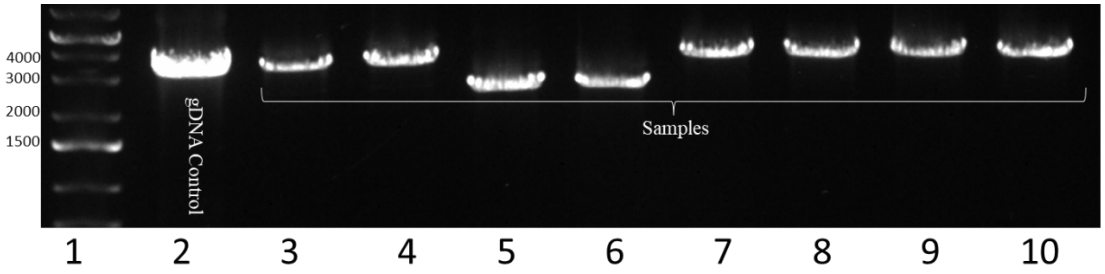

Figure 72 - Screening E. limosum NG-6849 for the deletion of pbt2 by colony PCR in a Δact background

A standard OneTaq colony PCR (2.5.2.3, 2.5.2.3.3) was performed on 8 colonies of Δact NG-6849 E. limosum transformed with TAKOS-pbt2short. Primers (pbt2 TAT SCR2 F and pbt2 TAT SC2R R) were used to screen for a knockout of the pbt2 (lanes 3 - 10). A control of wild-type NG-6849 E. limosum genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 0.8% agarose gel. wild-type bands (~3.2 kb) appeared in all but 2 colonies. Lanes 5 and 6 no wild-type band, only the mutant band (~2.1 kb) and so were pure knockouts.

Screening (Figure 72) identified two chromosomal knockouts of *pbt2* in the Δ*act* background. Successful colonies were streaked onto selective and non-selective medium to ascertain if they retained the ability to grow on selection. Both colonies, and their daughter cells, were unable to grow on selection, which may suggest that the differing homology arm lengths contributed to the previously observed resistance. Prior to the creation of competent cells (2.3) from the double knockout mutants, successful knockouts of *pbt1* (4.3.4.9) and *pbt2* in a Δ*act* background were screened (2.6.6) to ensure the Δ*act* genotype was maintained (Figure 73).

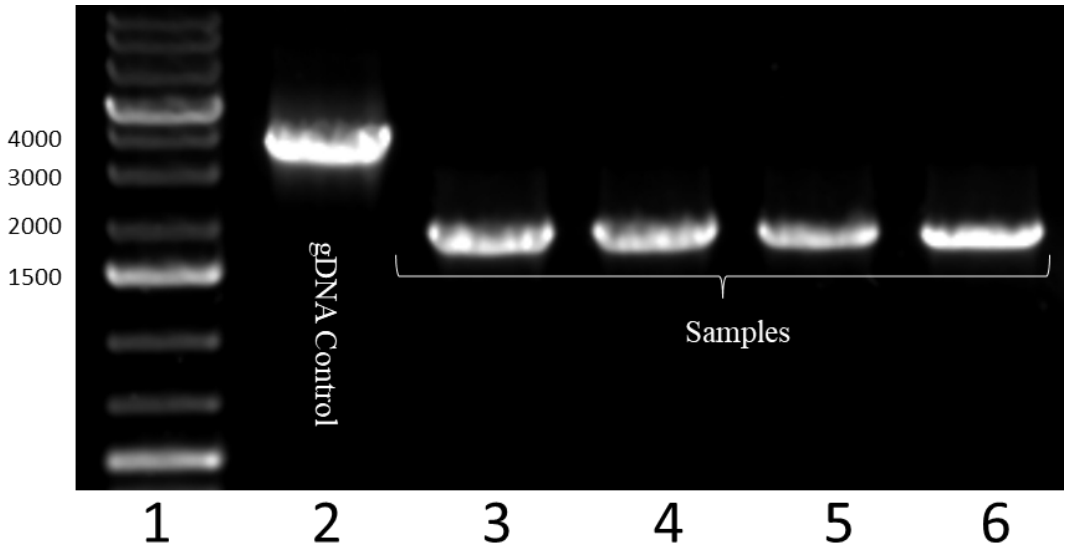

Figure 73 - Screening E. limosum NG-6849 for the deletion of act in a purported ΔactΔpbt1 and ΔactΔpbt2 background by colony PCR

A standard OneTaq colony PCR was performed on 4 colonies of ⊿act E. limosum NG-6849 which had successfully undergone deletion of pbt1 (lanes 3 and 4) and pbt2 (lanes 5 and 6). Primers (act TAT SCR2 F and act TAT SCR2 R) were used to screen for a chromosomal knockout of the act (lanes 3 - 6). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 0.8% agarose gel. Bands of the expected deletion size (~2 kb) were observed in all lanes but the control which displayed a wild-type band (of ~3.5 kb).

Screening (Figure 73) showed that all colonies which had undergone chromosomal knockout of *pbt1* and all which had undergone chromosomal knockout of *pbt2* also retained prior chromosomal knockout of *act*. Successful colonies were stocked (2.1.2.5) and were used to make competent cells (2.3) to allow for the transformation of the other *pbt* knockout vector.

4.3.4.6 Generation of *E. limosum* NG-6849 Δ*act*Δ*pbt1*Δ*pbt2* mutants using toxin-antitoxin system

With the double deletion genotype confirmed, competent cells (2.3) were created and transformed. Δ*act*Δ*pbt1* was initially the selected background for transformation, the outcome would be genetically identical in any case. TAKOS-*pbt2*-short was constructed previously for the creation of the Δ*atc*Δ*pbt2* knockout (4.3.2.5). It was used to perform a Δ*pbt2* knockout in an *E. limosum* NG-6849 Δ*act*Δ*pbt1* mutant background (2.6.5.2). Colonies were screened (2.6.6) with the screening primers *pbt2*

TAT SCR F and *pbt2* TAT SCR R (Figure 74A) for the knockout of *pbt2*. Successful colonies were then screened for retention of the prior two knockouts Δ*act* and Δ*pbt1* (Figure 74C and Figure 74B respectively) using the primers act TAT SCR2 F and act TAT SC2R R and pbt1 TAT SCR2 F and pbt1 TAT SC2R R, respectively.

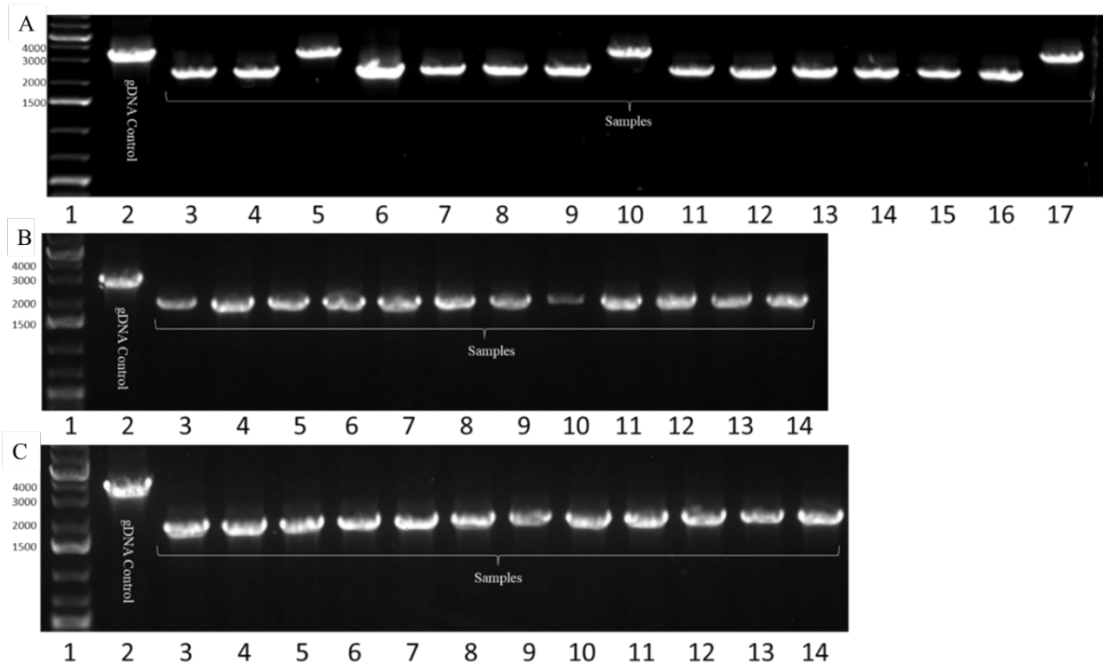

Figure 74 - Screening of E. limosum NG-6849 ΔactΔpbt1 for the deletion of pbt2 (A) and confirmation of previous knockouts for pbt1 (B) and act (C), respectively

A standard OneTaq colony PCR was performed on E. limosum NG-6849 ΔactΔpbt1 transformed with TAKOS-pbt2short. A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, all samples were run on 0.8% agarose gels.

- **A.** *15 colonies (lanes 3 - 17) were screened for a chromosomal knockout of the pbt2 using the primers (pbt2 TAT SCR2 F and pbt2 TAT SC2R R). All screened colonies but those in lanes 5, 10, and 17 were pure knockouts. To ensure colonies had retained the prior two deletions colony PCRs was performed using the same cells mass (Figure 74B and Figure 74C).*
- **B.** *12 colonies with successful chromosomal knockout of pbt2 (Figure 74A) were screened with primers (pbt1 TAT SCR2 F and pbt1 TAT SC2R R) to ensure retention of a chromosomal knockout of pbt1 (lanes 3 - 14). All screened colonies retained chromosomal knockout of pbt1.*
- **C.** *12 colonies with successful chromosomal knockout of pbt1 (Figure 74A) were screened with primers (act TAT SCR2 F and act TAT SC2R R) to ensure retention of a chromosomal knockout of act (lanes 3 - 14). All screened colonies retained chromosomal knockout of act.*

Only those colonies which maintained the prior deletions of *pbt1* (Figure 74B) and *act* (Figure 74C) and attained the new deletion of *pbt2* (Figure 74A) were stocked (2.1.2.5) and underwent phenotype analysis (4.3.4.7).

4.3.4.7 Butyrate production phenotype assessment

The rationale for the creation of Δ*act* (4.3.4.2.1), Δ*pbt1* (4.3.4.5), Δ*pbt2* (4.3.4.6), Δ*act*Δ*pbt1* (4.3.4.9), Δ*act*Δ*pbt2* (4.3.4.10), and Δ*act*Δ*pbt1*Δ*pbt2* (4.3.4.11) mutants was to perturb butyrate production. To test for altered phenotype, mutant butyrate production levels were compared with wild-type cells. Growth curves were performed in serum bottles (2.1.2.1), with phenotypes being assessed both on glucose (Figure 75) and methanol (Figure 76) as feedstock, with methanol being likely to produce more butyrate, based on previous HPLC data (Figure 23).

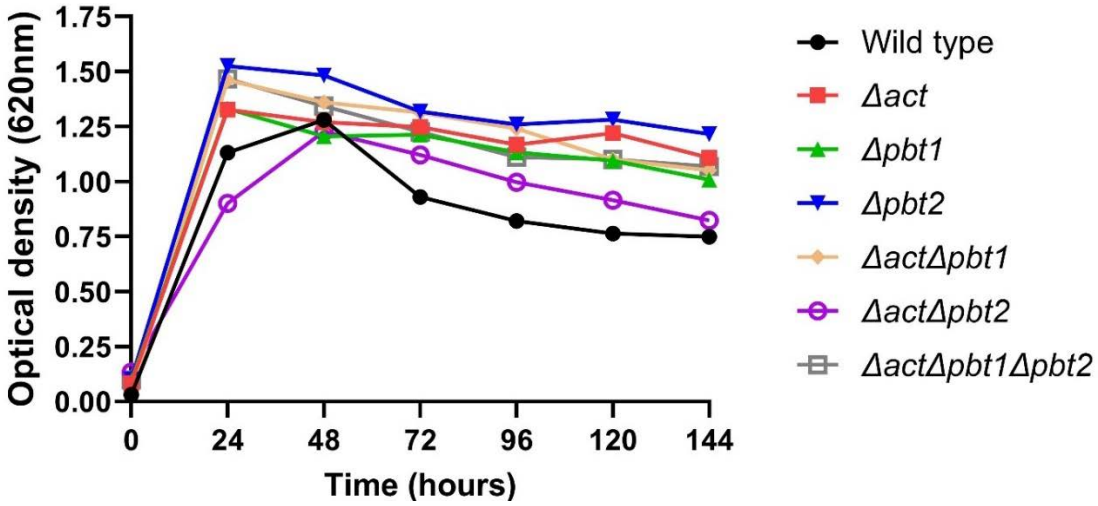

Figure 75 - Growth of putative butyrate production mutants and wild-type E. limosum NG-6849 on glucose

Δact, Δpbt1, Δpbt2, ΔactΔpbt1, ΔactΔpbt2 and ΔactΔpbt1Δpbt2 mutants and a wild-type control were grown in Pacaud medium without yeast extract with 10 mM glucose as the carbon source. Cells were grown in 30 mL of medium in serum bottles with a standard anaerobic gas headspace. All cultures were grown as singlet.

Growth of mutants on glucose (Figure 75) was roughly comparable to that of the wild-type *E. limosum* NG-6849 strain, suggesting no negative growth effects associated with these deletions.

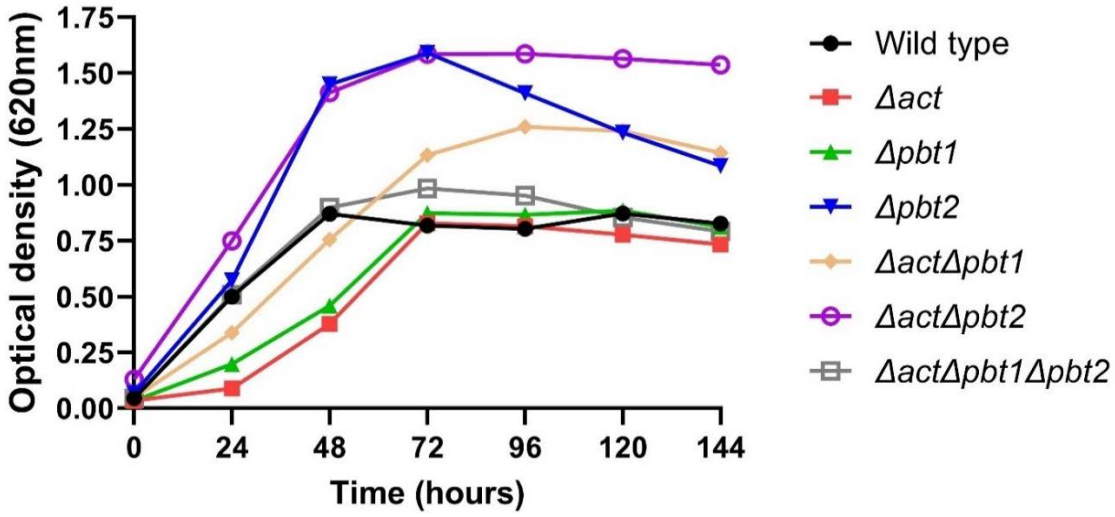

Figure 76 - Growth of putative butyrate production mutants and wild-type E. limosum NG-6849 on methanol.

Δact, Δpbt1, Δpbt2, ΔactΔpbt1, ΔactΔpbt2, ΔactΔpbt1Δpbt2 and a wild-type control were grown in Pacaud medium without yeast extract with 200 mM methanol as the carbon source. Cells were grown in 30 mL of medium in serum bottles with a standard anaerobic gas headspace. All cultures were grown as singlet.

Methanol growth curves (Figure 76) displayed no negative growth phenotype because of knockouts. In the case of the Δ*pbt2* strains, growth was increased compared to the wild type. The reason for this is unclear. If *pbt2* was disadvantageous to growth, then cells would likely down regulate the gene. The *pbt2* gene encodes a phosphate butyryltransferase and should not confer any growth advantage, based on the current understanding of *E. limosum* metabolism.

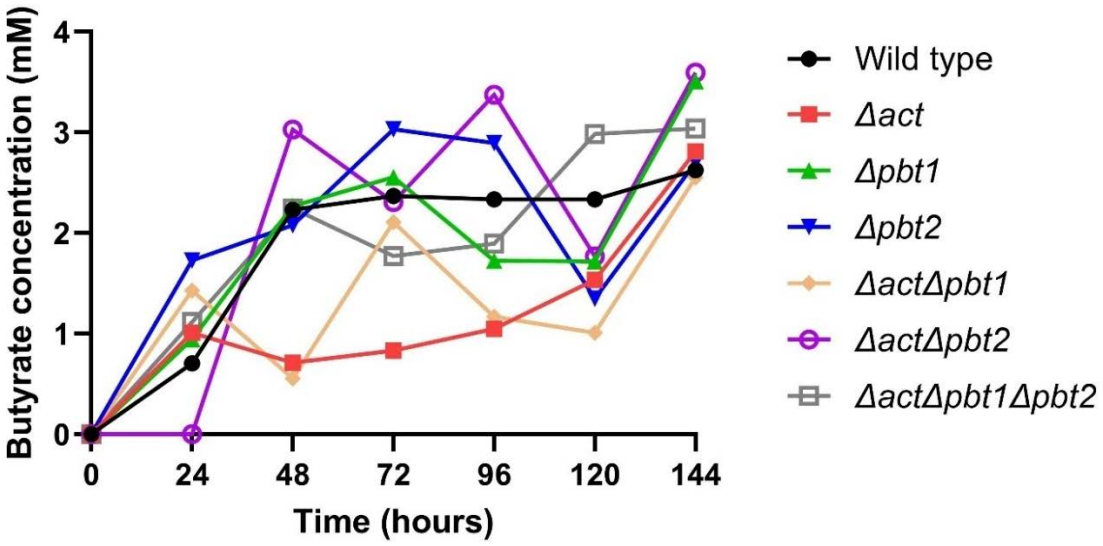

Figure 77 - Butyrate concentration putative butyrate production mutants and wild-type E. limosum NG-6849 growing on glucose.

Δact, Δpbt1, Δpbt2, ΔactΔpbt1, ΔactΔpbt2, ΔactΔpbt1Δpbt2 and a wild-type NG-6849 control was grown in Pacaud medium without yeast extract with 10 mM glucose as the carbon source. Cells were grown in 30 mL of medium in serum bottles with a standard anaerobic gas headspace.

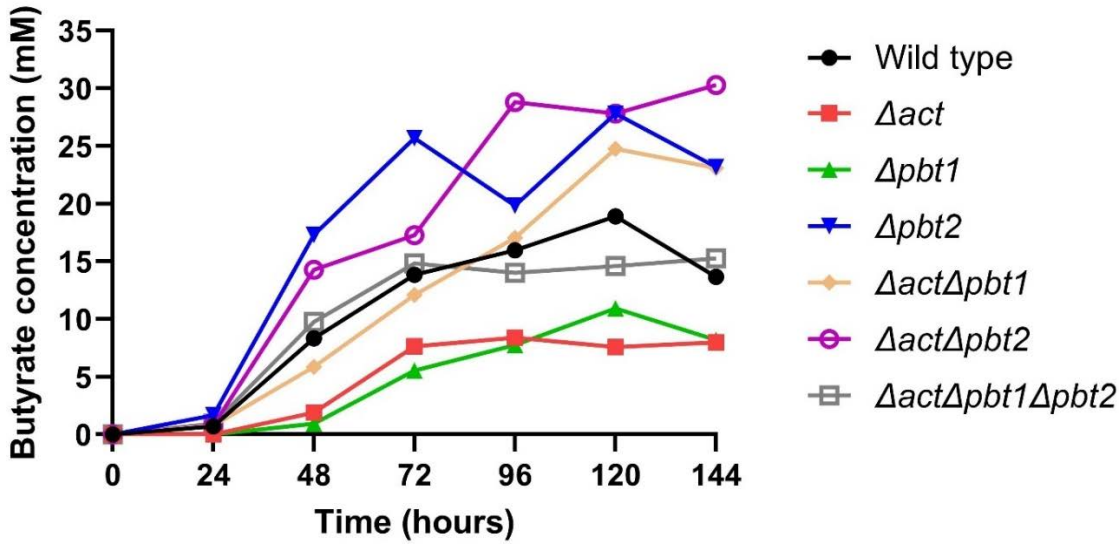

Figure 78 - Butyrate concentration in putative butyrate production mutants and wild-type E. limosum NG-6849 on methanol.

Δact, Δpbt1, Δpbt2, ΔactΔpbt1, ΔactΔpbt2, ΔactΔpbt1Δpbt2 and a wild-type control were grown in Pacaud medium without yeast extract with 200 mM methanol as the carbon source. Cells were grown in 30 mL of Pacaud medium in serum bottles with a standard anaerobic gas headspace.

HPLC data for butyrate concentration were assessed on both glucose (Figure 77) and methanol (Figure 78) growth conditions. In all cases, butyrate was produced, and in the case of methanol growth, more butyrate was produced than in glucose growth. The results of methanol growth (Figure 78) were unexpected, with an Δ*act* mutant

appearing to produce less butyrate than a double deletion background with Δ*actpbt1* and Δ*actpbt2* producing more than the wild type. However, this is not odd when growth data on methanol (Figure 77) is accounted for, with Δ*act* having the lowest maximal OD_{620}. To better visualize the data, graphs were constructed with normalization of butyrate concentration using OD_{620}.

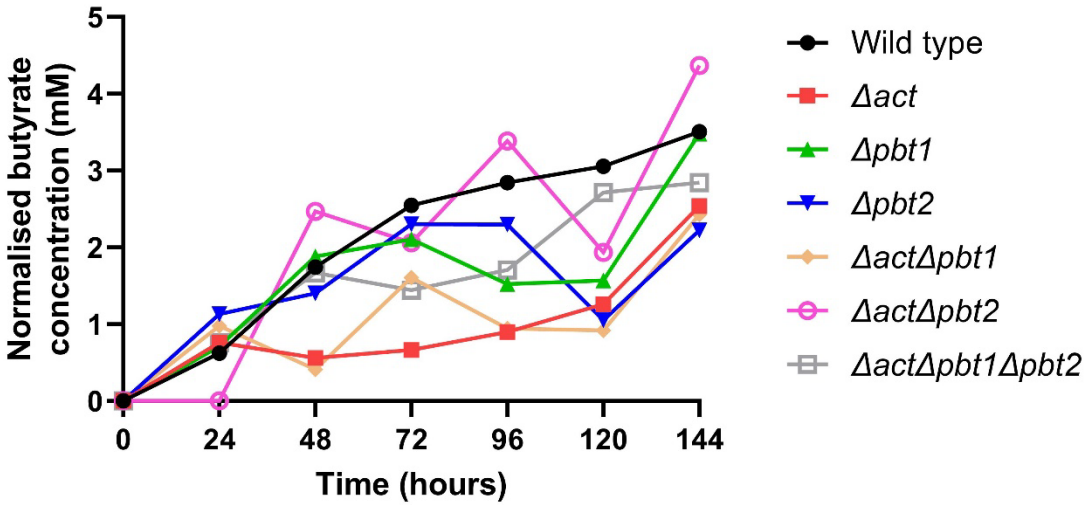

Figure 79 - Normalised butyrate concentration in glucose-grown cultures of putative butyrate production mutants and wild-type E. limosum NG-6849.

Δact, Δpbt1, Δpbt2, ΔactΔpbt1, ΔactΔpbt2, ΔactΔpbt1Δpbt2 and a wild-type control were grown in Pacaud medium without yeast extract with 10 mM glucose as the carbon source. Cells were grown in 30 mL of Pacaud medium in serum bottles with a standard anaerobic gas headspace. Butyrate concentration was normalised with optical density values.

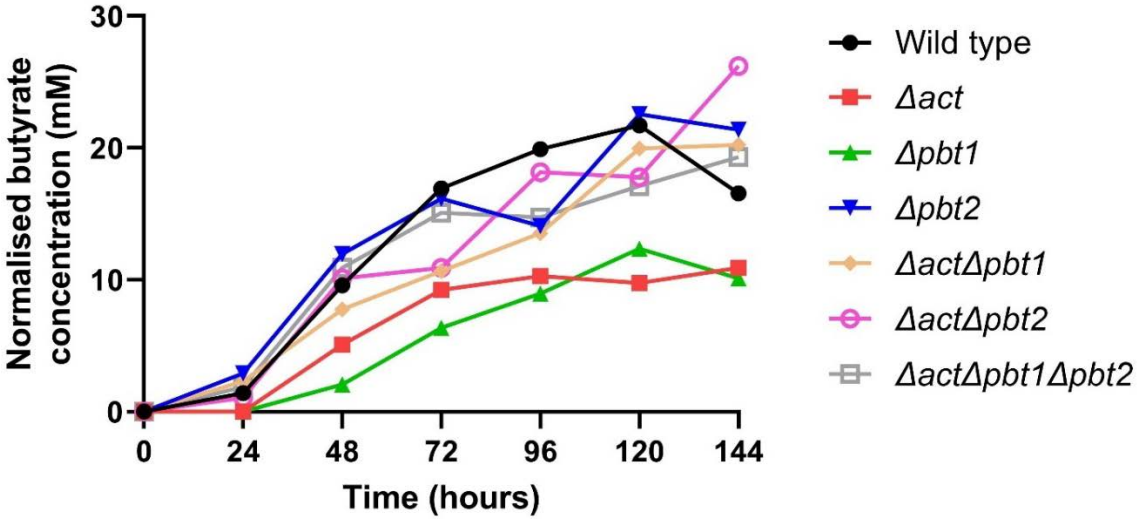

Figure 80 - Normalised butyrate concentration in methanol-grown cultures of putative butyrate production mutants and wild-type E. limosum NG-6849.

Δact, Δpbt1, Δpbt2, ΔactΔpbt1, ΔactΔpbt2, ΔactΔpbt1Δpbt2 and a wild-type control were grown in Pacaud medium without yeast extract with 200 mM methanol as the carbon source. Cells were grown in 30 mL of Pacaud medium in serum bottles with a standard anaerobic gas headspace. Butyrate concentration was normalised with optical density values.

Once the butyrate production data was normalized with growth data for glucose (Figure 79) and methanol (Figure 80) datasets, it became clear that all mutants were still producing butyrate, with maximal levels (Table 24) being sufficiently high to confirm that none of these mutants were able to eliminate butyrate production.

Table 24 - Maximal butyrate titres for all NG-6849 strains assessed, adjusted for OD_{620}, under methanol and glucose growth conditions

Strain	**Maximal Butyrate Titre, Glucose Growth (mM)**	**Maximal Butyrate Titre, Methanol Growth (mM)**
Wild type	3.5	21.7
Δ*act*	2.5	10.9
Δ*pbt1*	3.4	12.4
Δ*pbt2*	2.3	22.6
Δ*act*Δ*pbt1*	2.4	20.2
Δ*act*Δ*pbt2*	4.3	26.2
Δ*act*Δ*pbt1*Δ*pbt2*	2.8	19.3

Since all strains displayed production of butyrate, it suggests that either the mutations were not causing the expected decrease in butyrate titres, or the colony PCR results were not reliable, thus further confirmation was necessary.

4.3.4.7.1 Further mutant confirmation

To exclude the possibility of wild-type alleles being retained after gene knockout, internal screening of all knockout targets was performed to test for the retention of chromosomal knockout genes (Figure 81). One microliter of the stocked sample was taken and resuspended in 30 µl dH_2O to allow for colony PCR to be performed. Cells were screened with the relevant internal primers relative to their purported genotype (*pbt1* int F and *pbt1* int R for *pbt1* internal screening; with an expected size of ~820 bp, *pbt2* int F and *pbt2* int R for *pbt2* internal screening; with an expected size of ~850 bp, and *act* int F and *act* int R for *act* internal screening; with an expected size of 1.5 kB). In addition, a known external primer pair (cop TAT SCR2 F and cop TAT SCR2 R).was used as a positive control to ensure colony PCR was working for each sample. In addition, a control of wild type genomic DNA was used.

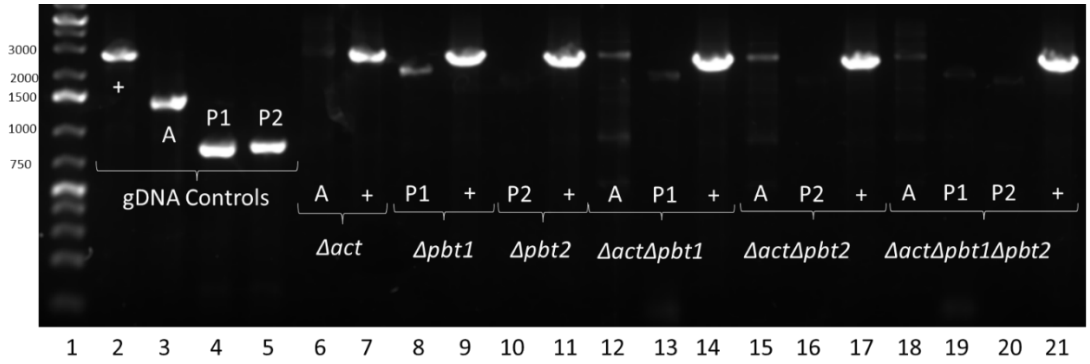

Figure 81 - Screening E. limosum NG-6849 for the deletion of purported butyrate production genes by colony PCR

A standard OneTaq colony PCR was performed on end points of the butyrate mutant glucose growth curve (Figure 75). The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 0.8% agarose gel. A control of wild-type E. limosum NG-6849 genomic DNA was used in lanes 2, 3, 4 and 5, acting as a positive control and screening for act, pbt1 and pbt2 respectively. Lanes 6 and 7 were screening a Δact mutant. Lanes 8 and 9 were screening a Δpbt1 mutant. Lanes 10 and 11 were screening for pbt2 a Δpbt2 mutant. 12, 13 and 14 were screening a ΔactΔpbt1 mutant. 15, 16 and 17 were screening a ΔactΔpbt2 mutant. 12, 13 and 14 were screening a ΔactΔpbt1Δpbt2 mutant.

- *"+" refers to a positive control lane amplifying using the primers cop TAT SCR2 F and cop TAT SCR2 R,*
- *"A" refers to an internal screening of act using the primers act int F and act int R,*
- *"P1" refers to an internal screening of pbt1 using the primers pbt1 int F and pbt1 int R,*
- *"P2" refers to an internal screening of pbt2 using the primers pbt2 int F and pbt2 int R.*

All controls worked as expected (Figure 81), confirming primer function and PCR conditions. All mutants displayed strong bands with the positive control primers, suggesting a good quality template. While some non-specific banding occurred in

some mutants, no previously confirmed mutants displayed any genetic retention of the purportedly deleted gene.

To further confirm this result, genomic sequencing was performed on colonies to look for knockout of the butyrate-associated genes. Single colonies were isolated after several restreakings to ensure an absence of contamination. Amplification of 16S rRNA gene sequences from biomass was performed using colony PCR (2.5.2.3.1), using the primers 16S 8F and 16S 1492R, and sent for Sanger sequencing (2.5.3) to ensure the cells were *E. limosum*. Upon species confirmation, genomic extractions were performed (2.5), and after quality control they were sent for Illumina sequencing (2.7.1). Paired reads were received, and genetic analysis was performed using CLC genomics workbench 21 (2.7). Sequences were imported using the Illumina import function, and paired reads were selected, as read data was in said format. Alignments were performed against the SBRC ATCC 8486 reference sequence used (and verified) in Chapter 3 (Figure 82), with reports confirming good quality reads. Reads aligned >99.5% for all 6 mutants and, on average, had over 900,000 reads of lengths 250-254.

The purpose of genomic sequencing was to try to find any traces of the purportedly knocked out genes. Using the "Find low coverage" tool on the main navigation window, regions with low coverage could be identified. This would validate the external screening observed via colony PCR. Regarding reintegration of genes elsewhere in the genome, CLC genomic workbench's alignment tool would align the region but flag it with a yellow colour. This would indicate that the region exists in the sequence provided but is not located in the same position as in the reference genome, indicating the gene's presence elsewhere in the genome.

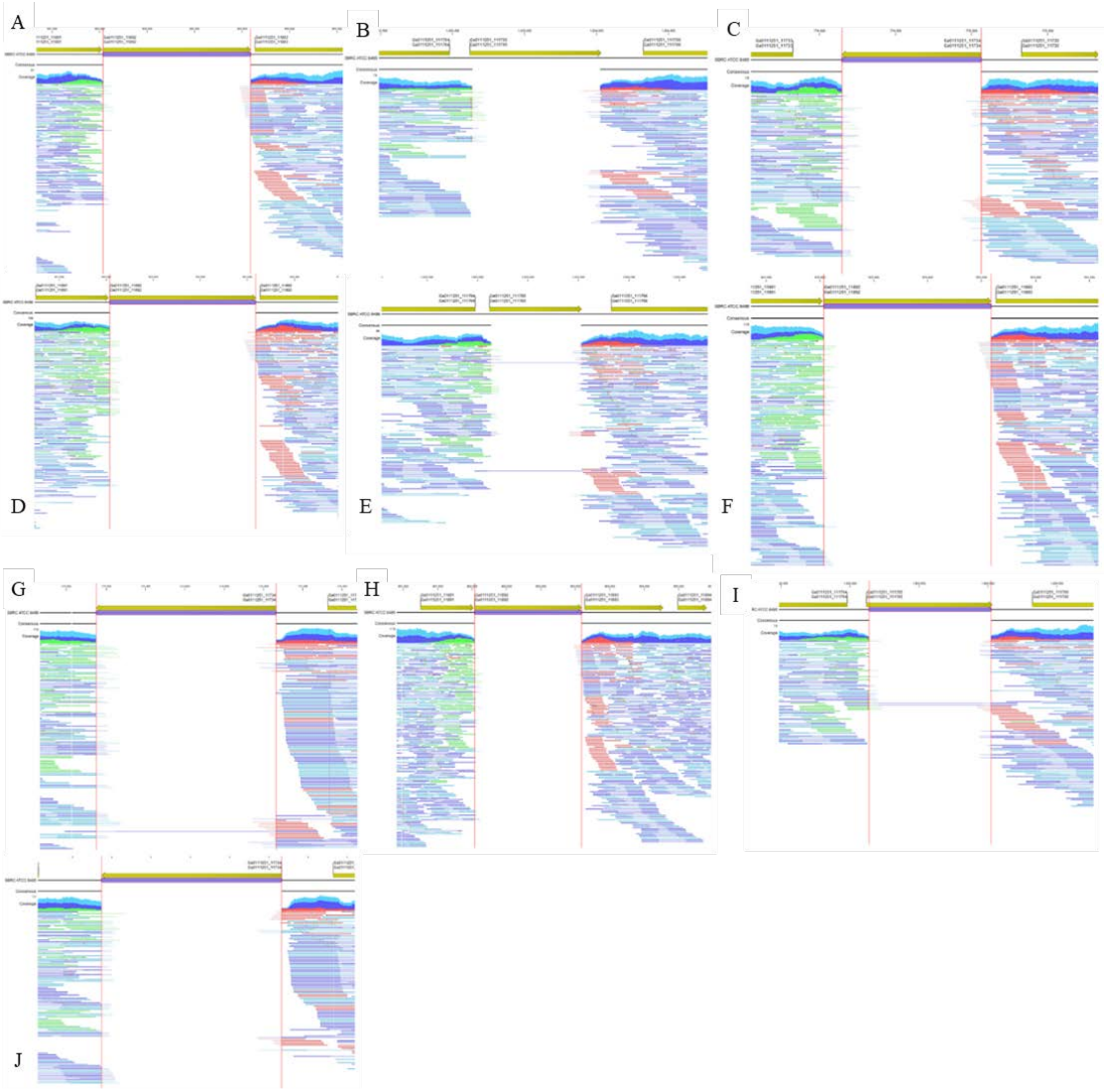

Figure 82 – Navigation area of CLC genomics workbench 21 displaying chromosomal locations of knockout targets in purported Δact, Δpbt1, Δpbt2, ΔactΔpbt1, ΔactΔpbt2 and ΔactΔpbt1Δpbt2 mutants.

Summary of CLC genomics alignment search against knockout targets in purported knockouts displaying no alignments and thus confirming chromosomal knockout

- **A.** *Alignment of act gene against purported Δact mutant (Figure 106)*
- **B.** *Alignment of pbt1 gene against purported Δpbt1 mutant (Figure 107)*
- **C.** *Alignment of pbt2 gene against purported Δpbt2 mutant (Figure 108)*
- **D.** *Alignment of act gene against purported ΔactΔpbt1 mutant (Figure 109)*
- **E.** *Alignment of pbt1 gene against purported ΔactΔpbt1 mutant (Figure 110)*
- **F.** *Alignment of act gene against purported ΔactΔpbt2 mutant (Figure 111)*
- **G.** *Alignment of pbt2 gene against purported ΔactΔpbt2 mutant (Figure 112)*
- **H.** *Alignment of act gene against purported ΔactΔpbt1Δpbt2 mutant (Figure 113)*
- **I.** *Alignment of pbt1 gene against purported ΔactΔpbt1Δpbt2 mutant (Figure 114)*
- **J.** *Alignment of pbt2 gene against purported ΔactΔpbt1Δpbt2 mutant (Figure 116)*

A lack of reads across the wild-type *act* region in all Δ*act* mutants (Figure 106, Figure 109, Figure 111, and Figure 113), combined with the lack of any sequences suggesting the presence of *act* elsewhere in the genome, confirms that the *act* gene

was knocked out cleanly and that there was no reintegration of the wild-type allele elsewhere in the chromosome. While it initially appears that reads exist across *pbt1* regions (Figure 114), manual examination showed no alignment with the region for those reads. Figure 115 had no reads across it, indicating that no *pbt1* is present at the wild-type locations in any of the Δ*pbt1* mutants (Figure 107, Figure 109, and Figure 114). *pbt1* was not present elsewhere in the genome at any point, confirming that all Δ*pbt1* mutants were indeed clean *pbt1* knockouts. *pbt2* was also confirmed as being deleted, with no *pbt2* in the wild-type regions or elsewhere in the genome (Figure 108, Figure 112, and Figure 116), such that all Δ*pbt2* mutants were clean knockouts. The genetic sequencing therefore confirmed the deletion of the target genes. Given that production of butyrate is still occurring, other enzymes must be involved.

4.3.4.8 Generation of *E. limosum* NG-6849 Δ*bcact* mutants using toxin-antitoxin system

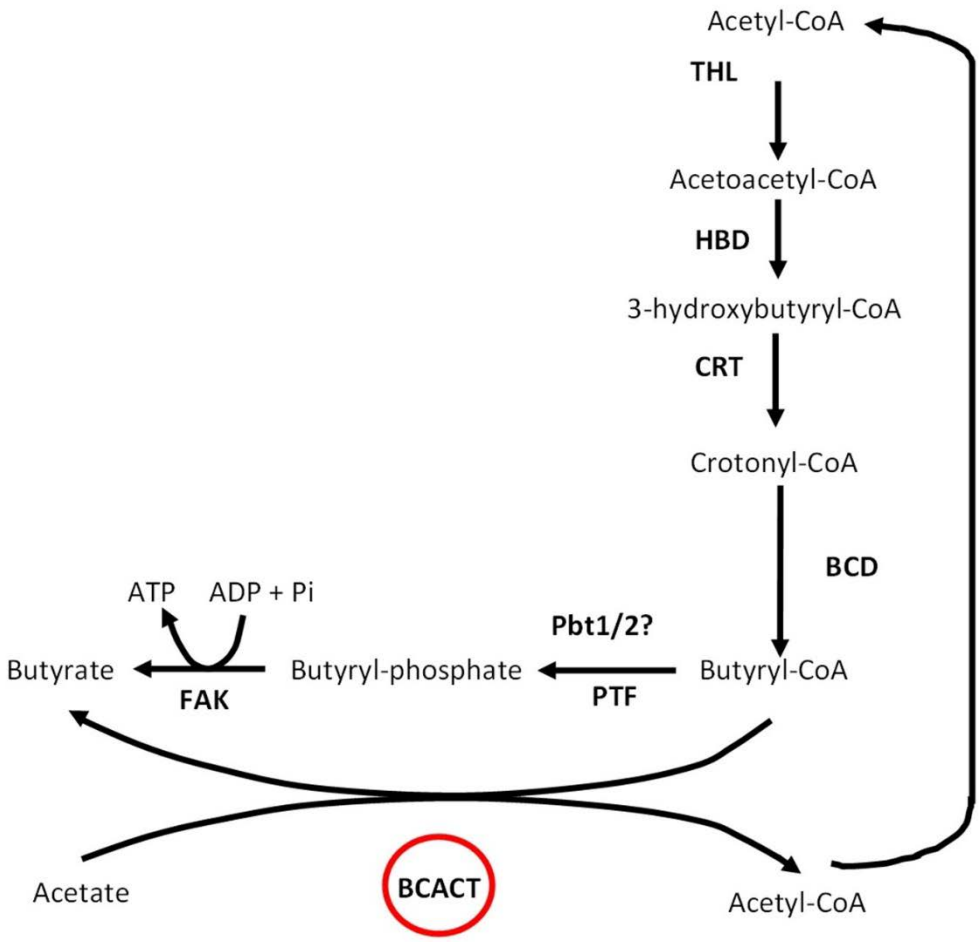

Figure 83 - Putative reaction catalysed by butyryl CoA:acetate CoA transferase.

BCACT butyryl CoA:acetate CoA transferase; CRT, crotonase; FAK, fatty acid kinase; HBD, 3-hydroxybutyryl-CoA dehydrogenase; BCD, butyryl-CoA dehydrogenase; PTF, phosphotransferase; THL, thiolase

Butyryl CoA:acetate CoA transferase (EC 2.8.3.8) is a family I CoA-transferase that transfers CoA from butyryl-CoA to an acetyl group, forming acetyl-CoA. The overall reaction takes place through a “ping-pong” kinetic mechanism (326,327). This mechanism is non-sequential, meaning that one product is formed and released before the binding of a second substrate. Deletion of this gene was hypothesised to reduce or prevent the production of butyrate due to the inability to reclaim CoA from butyryl-CoA, due to the lack of butyryl CoA:acetate CoA transferase activity. The purpose of this knockout was to analyse the effect of this gene on butyrate production in *E. limosum.*

A TAKOS vector, TAKOS-*bcact*, was designed and created as previously described (2.6.2). Upon reviewing TraDIS data (Dr. Elizabeth Redfern, University of Nottingham, personal communication; Figure 117) it seemed likely that this deletion would present issues. Therefore, initial work was all performed in wild-type *E. limosum* NG-6849. Transformations were done in triplicate and double the number of colonies were screened to maximise the chances of identifying a mutant. However, for all three transformations, all cells screened (Figure 84) were wild-type, which suggests either an issue with the vector or that the TraDIS prediction for essentiality (Dr. Elizabeth Redfern, University of Nottingham, personal communication; Figure 117) was correct.

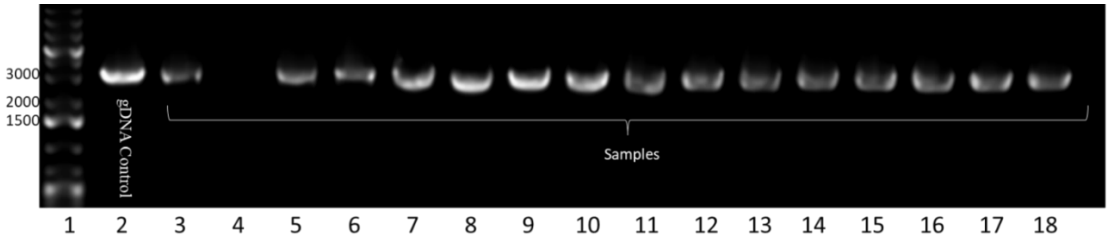

Figure 84 - Screening E. limosum NG-6849 for the deletion of bcact by colony PCR

A standard OneTaq colony PCR was performed on 16 colonies of Δpbt1Δpbt2 E. limosum NG-6849 transformed with TAKOS-bcact. Primers (bcact TAT SCR F and bcact TAT SCR R) were used to screen for a chromosomal knockout of the bcact (lanes 3 - 18). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, all samples were run on a 0.8% agarose gel. All screened colonies were wild type.

No further attempts were made to create this mutant. However, this plasmid, along with the TAKOS plasmids for *pbt1* and *pbt2*, was passed on to Philippe Soucaille, a collaborator on the Era CoBiotech BioMetChem project. At the time of this work ended, a new strain had been developed, and they wished to try to replicate the triple knockout with *bcact*, believing that the adaptations accrued in their strain would overcome the essentiality of *bcact* in the NG-6849 strain.

4.3.4.9 Phenotype assessment of toxin-antitoxin derived Δ*cop*, Δ*mti*, Δ*mtiI*, and Δ*metF1* mutants

The phenotypes of the methanol-related mutants Δ*cop*, Δ*mti*, Δ*mtiI*, and Δ*metF1* needed to be assessed to determine if their function was linked to growth on methanol. The labelling of the *metF1* gene as a "*metF*" implies that it is a methylenetetrahydrofolate reductase (328), if *metF1*, like the regular *metF*, is essential to methanol metabolism, then growth on methanol would be affected if it was deleted. If it is extraneous, it would not be. Δ*metF1* and Δ*mti* mutants were generated first and were, therefore, assessed first.

Growth curves (Figure 85) were performed using methanol feedstock in serum bottles (2.1.2.1) with starter cultures grown on glucose. This was done as Δ*mti* and Δ*metF1* should be unable to or less able to grow on methanol. Insufficient biomass accumulation in the starter culture would prevent the growth in serum bottles. To minimise carryover of glucose, biomass for inoculation was washed several times in inert medium prior to inoculation in methanol containing Pacaud medium.

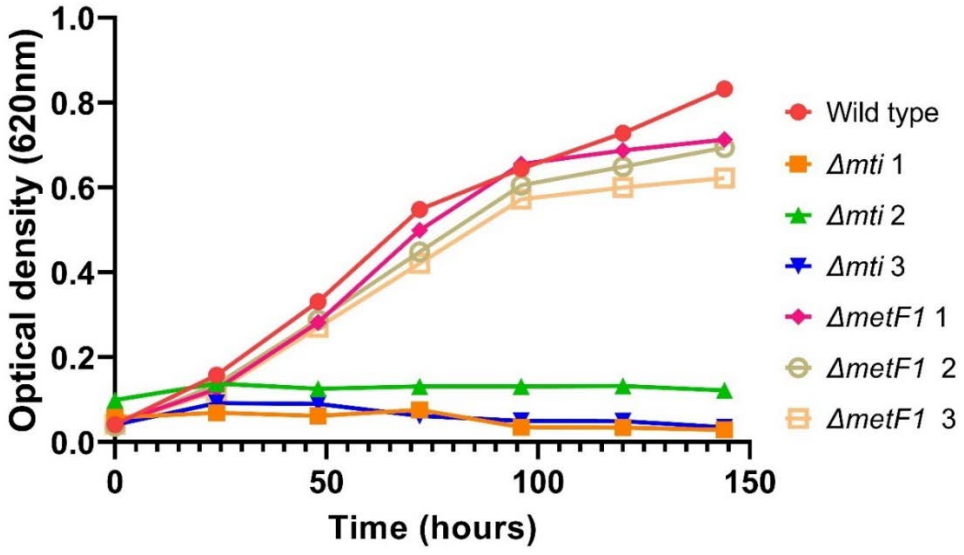

Figure 85 - Growth Curve displaying the growth of ΔmetF1 and Δmti E. limosum NG-6849 mutants on methanol in Pacaud medium.

Wild-type, ΔmetF1 and Δmti E. limosum NG-6849 strain were grown in batch culture in serum bottles with 200 mM methanol as the substrate and an anaerobic gas headspace (~8% hydrogen, ~77% N_2, ~15% CO_2) of 200 mL. Spectrophotometer readings for absorbance at 620 nm were taken every 24 hours. Wild type E. limosum NG-6849 growth was done as singlet. ΔmetF1 and Δmti were grown as triplicates.

The long lag phase observed in the wild-type control indicates good removal of glucose-containing medium from the inoculum. The long lag phase is indicative of a switch of carbon source from glucose to methanol in an abrupt fashion, rather than of mixotrophic growth on glucose and methanol in *E. limosum*. Δ*mti* cells did not revert to a native phenotype at any point of growth and were unable to grow on methanol, as would be expected given the putative genetic function. Δ*metF1* behaved essentially as the wild-type, able to grow as well on methanol, suggesting that if the gene is involved in any way in methanol metabolism, it is redundant and may have been misannotated as a methylenetetrahydrofolate reductase. This is further compounded by TraDIS data (Dr. Elizabeth Redfern, University of Nottingham, personal communication), suggesting that *metFV* is essential, unlike *metF1*. If *metF1* were a second copy of a methylenetetrahydrofolate reductase, *metFV* should be able to recover the phenotype sufficiently to allow at least some growth in a Δ*metF1* background, thus making the knockout possible.

The Δ*cop* and Δ*mtiI* mutants (4.3.4.3 and 4.3.4.5, respectively) were generated later in the project and thus phenotypic assessment was not performed concurrently with that of Δ*mti* and Δ*metF1*. Based on TraDIS data (Dr. Elizabeth Redfern, University of Nottingham, personal communication), these genes should not be essential for methanol growth. Presumably other proteins can act to replace their function. Growth curves (Figure 86) were performed using methanol in serum bottles (2.1.2.1) and as previously noted with Δ*metF1* and Δ*mti* mutants, with starter cultures performed in glucose without growth on methanol prior to the serum bottles, with the concern being regarding biomass if they did not grow. As with the prior growth curve, the wild-type control displayed a significant lag, again indicative of a lack of glucose carryover.

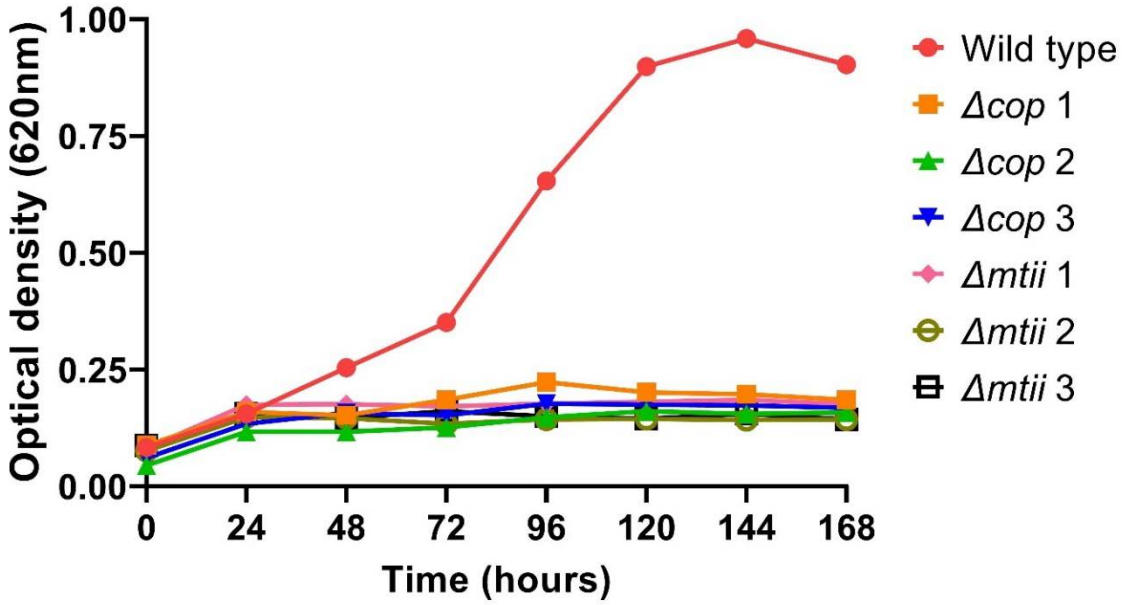

Figure 86 - Growth Curve displaying the growth of Δcop and ΔmtiI E. limosum NG-6849 strain mutants on methanol in Pacaud medium.

Wild-type, Δcop and ΔmtiI E. limosum NG-6849 strain were grown in batch culture in serum bottles with 200 mM methanol as the substrate and an anaerobic gas headspace (~8% hydrogen, ~77% N_2, ~15% CO_2) of 200 mL. Spectrophotometer readings for absorbance at 620 nm were taken every 24 hours. Wild type E. limosum NG-6849 growth was done as singlet. Δcop and ΔmtiI were grown as triplicates.

The *Δcop* and *ΔmtiI* E. limosum NG-6849 mutants were unable to grow effectively using methanol as a carbon source. While this result was expected based on the original understanding of methanol usage (106), it is perhaps incongruent with TraDIS data (Dr. Elizabeth Redfern, University of Nottingham, personal communication), which predicted that these genes were not essential for growth on methanol. Perhaps the slight increase in growth (a doubling from initial biomass) suggests that although some limited growth is possible, the deletion of these genes is highly disadvantageous for growth on methanol.

4.3.4.10 Complementation of methanol genes; *cop*, *mti* and *mtiI*

Given the altered phenotype observed in (4.3.4.9), it should follow that the knockouts are responsible for such a change in phenotype. Complementation is a good method to ensure this is the case. If a knockout causes a phenotypical change, reversing the knockout (knocking in the deleted gene) should restore phenotype to that of the wild

type. It is possible that the observed phenotype occurred due to off-target activity or other unknown effects of the system. This toxin-antitoxin system had not been extensively used in *E. limosum* prior to this work, and any off-site effects or issues it may cause were not known.

As reliable knockout vectors existed, they were used as a base for knock-in vectors (Figure 87), with no changes other than the insertion of cargo, i.e., the wild-type gene for the respective knockouts, between the homology arms using NEBuilder® HiFi DNA assembly (2.2.9).

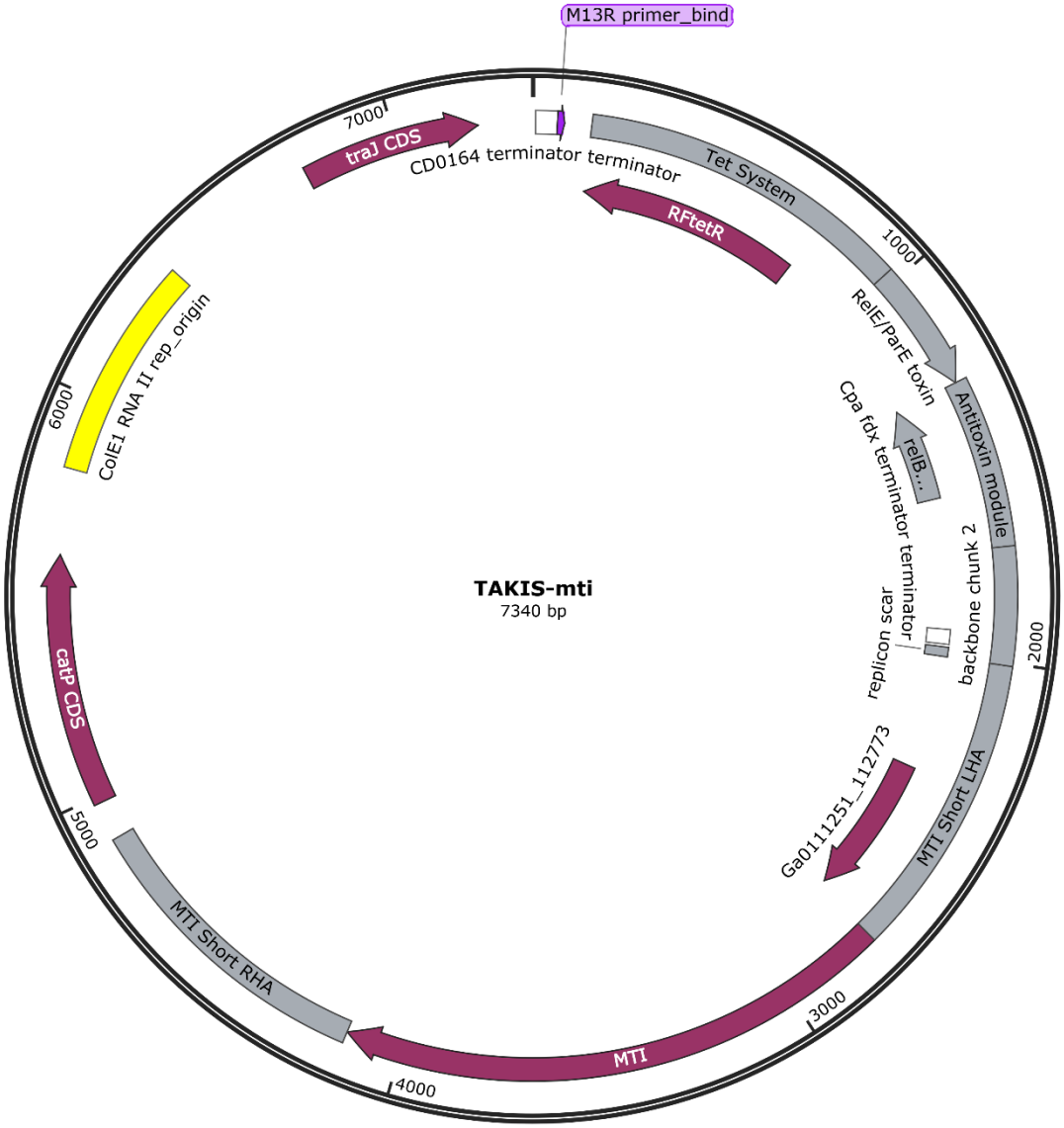

Figure 87 - TAKIS-mti a type 1 toxin-antitoxin based plasmid to knock in mti into a Δmti mutant background

Once plasmids were created (2.2.9) and confirmed via PCR (2.5.2.1), they were then introduced into *E. limosum* to restore genotype to that of the wild type (2.6.5.2). Colonies were screened (Figure 88 and Figure 89) using colony PCR (2.5.2.3.2) with the screening primers of each knockout. A "wild-type" product would be expected for a successful knock-in.

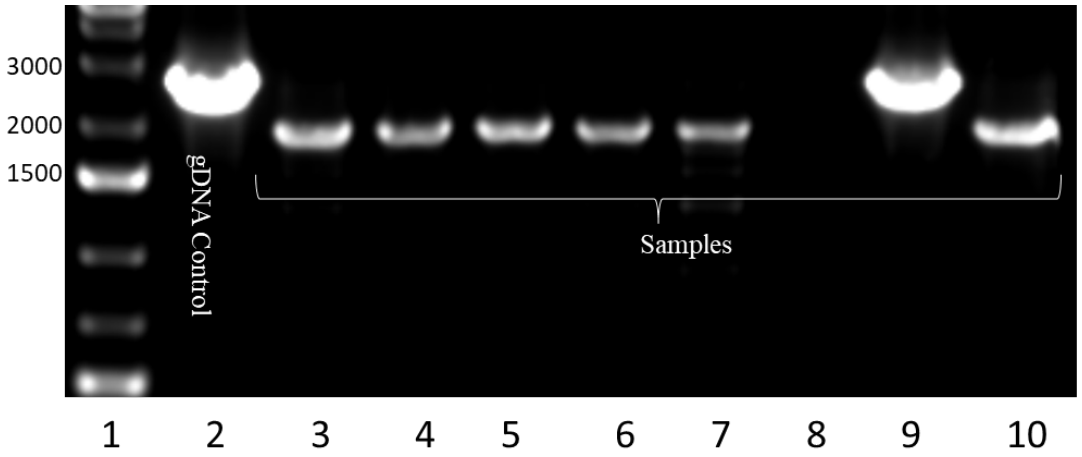

Figure 88 - Screening E. limosum NG-6849 for the integration of cop in complemented mutants by colony PCR screening

A standard OneTaq colony PCR was performed on 8 colonies of E. limosum NG-6849 transformed with cop-TAKIS-Short. Primers (cop TAT SCR2 F and cop TAT SCR2 R) were used to screen for a knockout of cop (lanes 3 - 10). A control of wild-type E. limosum NG-6849 genomic DNA was used in lane 2. The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 0.8% agarose gel. Only cells in lane 9 had correctly integrated the wild-type gene and restored genotype to that of the wild-type.

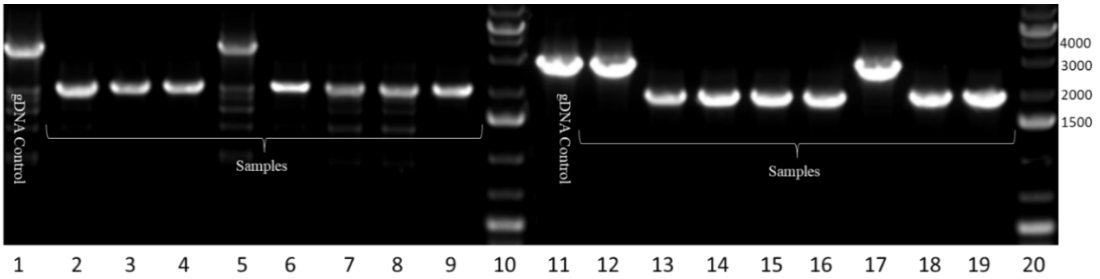

Figure 89 - Screening E. limosum NG-6849 for the integration of mti and mtiI in complemented mutants by colony PCR screening

A standard OneTaq colony PCR was performed on 8 colonies of E. limosum NG-6849 transformed with mti-TAKIS and mtiI-TAKIS, respectively. Primers (mti TAT SCR2 F and mti TAT SCR2 R) were used to screen for a knockout of mti (lanes 2 - 9) and (mtiI TAT SCR2 F and mtiI TAT SCR2 R) mtiI (lanes 11 - 19). A control of wild-type E. limosum NG-6849 genomic DNA was used in lanes 1 and 11. The ladder used (in lanes 10 and 20) was GeneRuler 1 kb plus DNA plus Ladder, samples were run on a 0.8% agarose gel. Cells in lanes 5, 12 and 17 had correctly integrated the wild-type gene and restored genotype to that of the wild-type.

Screening (Figure 88 and Figure 89) identified successful complementation of gene knockouts in several colonies. Successful colonies underwent the same phenotype assessment as their knockout mutant lineage (4.3.4.9). As previously these were performed as triplicate as *cop*, *mti*, and *mtiI* 1 – 3.

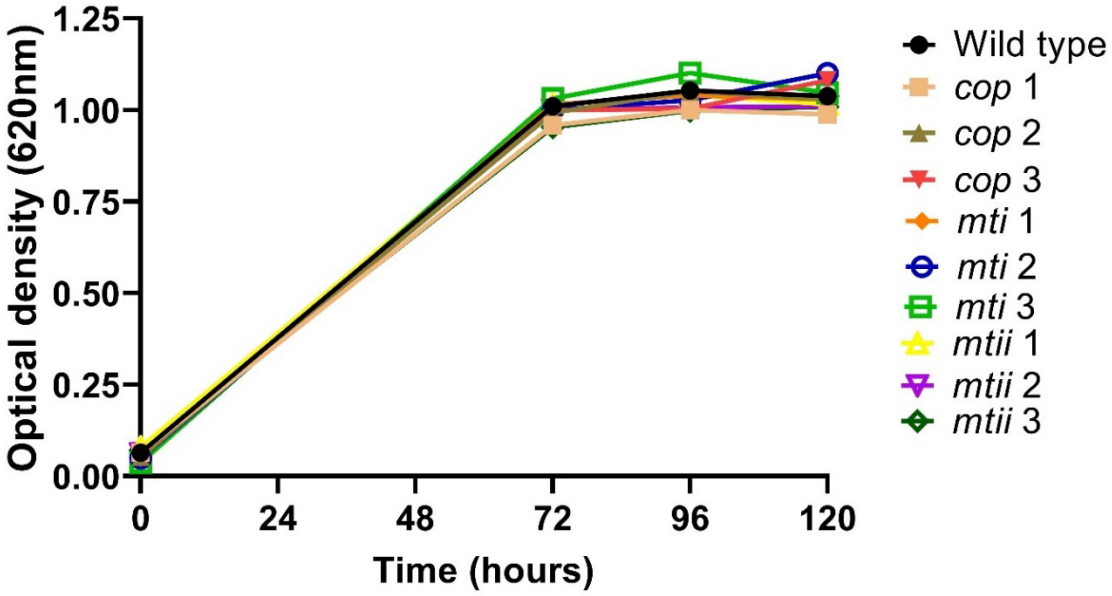

Figure 90 - Growth on methanol is restored in complemented E. limosum NG-6849 Δcop, Δmti and ΔmtiI mutants

Δcop restored to wild-type (cop 1 - 3), Δmti restored to wild-type (mti 1 - 3) and ΔmtiI restored to wild-type complemented mutants (mtiI 1 - 3) and a wild-type control were grown in Pacaud medium (2.1.1.2) without yeast extract with 200 mM methanol as the carbon source. Cells were grown in 30 mL of Pacaud medium in serum bottles with a standard anaerobic gas headspace. Wild type E. limosum NG-6849 was grown as a single flask, Δcop, Δmti and ΔmtiI mutants were grown as biological replicates.

In all cases, restored colonies behaved essentially as wild type, which is the expected outcome if *cop*, *mti*, and *mtiI* are involved in methanol metabolism. These data suggest that they are.

4.4 Discussion

A robust genetic toolbox is desirable for work with any organism; unfortunately, none was available for *E. limosum* at the start of this work. As this work progressed, improved tools were developed, with the toxin-antitoxin system (227) becoming available in the last six months of this work, allowing for the rapid creation of scarless knockouts within a week of the creation of a plasmid. A summary of all mutants attempted using both CRISPR-Cas9 and the TAKOS is shown in **Table 25**. The table lists the success in initial mutant creation and whether the mutants were stable over time or under selective stress conditions whereby the wild-type genotype would be advantageous. The area of phenotype of the mutations was expected to affect e.g., butyrate metabolism. Also listed is whether phenotypical analysis corroborates the respective hypothesis regarding function; phenotype analysis could not be performed on mutant backgrounds which could not successfully be created.

Table 25 - Genes targeted for deletion with CRISPR-Cas9 and TAKOS

Target Gene(s)	**Success**	**Stability**	**Method**	**Phenotype**	
				Expected	**Confirmed**
act	Yes	Unstable	CRISPR	Butyrate metabolism	No
act	Yes	Stable	TAKOS	Butyrate metabolism	No
act, metVmetF	No	N/A	CRISPR	Butyrate & methanol metabolism	N/A
act, pbt2	Yes	Stable	TAKOS	Butyrate metabolism	No
act ,pbt1	Yes	Stable	TAKOS	Butyrate metabolism	No
act, pbt1, pbt2	Yes	Stable	TAKOS	Butyrate metabolism	No
bcact	No	N/A	TAKOS	Butyrate metabolism	N/A
cop	Yes	Unstable	CRISPR	Methanol metabolism	Yes
cop	Yes	Stable	TAKOS	Methanol metabolism	Yes
hisI	Yes	Unstable	CRISPR	Histidine synthesis	Yes
ilvE	No*	N/A	TAKOS	Isoleucine, leucine & valine synthesis	N/A
glyA	No	N/A	TAKOS	Glycine synthesis	N/A
metF1	Yes	Unstable	CRISPR	Methanol metabolism	Yes
metF1	Yes	Stable	TAKOS	Methanol metabolism	Yes
metVmetF	Yes**	Unstable**	CRISPR	Methanol metabolism	N/A
metVmetF	No	N/A	TAKOS	Methanol metabolism	N/A
mti	Yes	Unstable	CRISPR	Methanol metabolism	Yes
mti	Yes	Stable	TAKOS	Methanol metabolism	Yes
mtiI	Yes	Unstable	CRISPR	Methanol metabolism	Yes
mtiI	Yes	Stable	TAKOS	Methanol metabolism	Yes
pbt1	Yes	Unstable	CRISPR	Butyrate metabolism	No
pbt1	Yes	Stable	TAKOS	Butyrate metabolism	No

pbt2	No	Unstable	CRISPR	Butyrate metabolism	N/A
pbt2	Yes	Stable	TAKOS	Butyrate metabolism	No

**Later analysis of gels suggests this mutation did occur however due to the discarding of the cell line and the lack of further verification it is marked as being unsuccessful.*
***this mutation was only possible due to maintenance of the gene on a plasmid, without plasmid maintenance of the gene genomic deletion was impossible.*

Regarding the CRISPR-Cas9 system, many challenges arose. Despite the widespread use of SpCas9 in other anaerobes within the SBRC, its application in *E. limosum* proved challenging. Initially, the system did not work at all in *E. limosum* NG-6849 (4.1). A strong promoter to control expression of the sgRNA proved essential for the system to function, albeit poorly. The system exhibited very low efficiency and required multiple weeks to create some mutants (4.3.2) the usefulness of which was compromised by retention of the knockout plasmid (4.3.3). This is in contrast to what has been observed in organisms such as *C. acetobutylicum*, in which a similar system has been employed (329). The cause of plasmid retention is unknown. A plasmid containing a gene that is essential could allow for genomic reintegration of a native, hitherto deleted, gene via allelic exchange (164–166), leading to a reversion to native genotype from the mutant.

However, a plasmid could avoid loss and be retained if the sgRNA or Cas9 regions mutate, negating counter-selection. Of these possibilities, mutations in the Cas9 seem the most likely. The *S. pyogenes* Cas9 used as part of the constitutive system (330) had already mutated previously on a plasmid while hosted in *E. coli*. As such, there were previously documented mutations in this Cas9. Given this and given the high degree of genetic regulatory homology between *E. coli* and *E. limosum*, it stands to reason that a minority of *E. limosum* cells may contain a further mutated version of Cas9 on a plasmid. If indeed it was a minority retaining a modified plasmid, then it would explain the growth patterns observed. If a significant majority of a population fails to grow on selection, then it is possible that the remaining cells may not grow to sufficient biomass in the allotted to timeframe to be readily identified. This would also explain the long lag in the reversion to the native genotype observed, with a minority of the modified plasmid-bearing cells undergoing allelic exchange of the plasmid-borne deletion target gene to the native position, resulting in a reversion to

the native genotype under selective pressure. If such a situation were occurring, then, due to the small starting population, it may take many doubling times for the mass of the reverted cell lines to predominate in the population and display the native phenotype. The success of a suicide plasmid to eliminate retained plasmid from the population further corroborates the Cas9 mutation theory.

Ideally, mutation in Cas9 would have been assessed by plasmid extraction and sequencing across the Cas9 gene. However, concentrations of extracted plasmids from *E. limosum* were poor and below the thresholds accepted for the sequencing services used as part of this work. This was somewhat mitigated by transforming the extracted plasmid into *E. coli*, and then extracting from *E. coli* which resulted in an increased plasmid yield. This was attempted in the Δ*cop* background. Plasmid was extracted from the Δ*cop E. limosum* using the same kit used for *E. coli* (2.2.2). The extracted plasmid was then introduced into *E. coli* (2.2.1.1) before being extracted. However, there were issues getting good sequencing reads. This issue was never overcome, and a good quality sequencing of plasmids extracted from *E. limosum* mutants, whether directly or using *E. coli* as an intermediary, was never achieved.

Retention of the Cas9 vector may explain the failure of the attempted double and triple knockouts described here. In addition to generating Δ*act*Δ*metVmetF* (4.3.2.8) triple knockouts, both Δ*act*Δ*pbt1* and Δ*act*Δ*pbt2* double knockouts were attempted with CRISPR-Cas9, but none were successful. In the case of Δ*act*Δ*metVmetF*, as observed for the toxin anti-toxin system (4.3.4.8), the *metVmetF* knockout itself is likely not viable. However, in the former cases, both double knockouts were observed when utilising the toxin anti-toxin system (4.3.4.4 and 4.3.4.6), and so the knockouts were feasible.

While the suicide-based plasmids were able rescue cell lines (4.3.3.1), it was a remedial solution at best and did not alleviate the many other issues with this CRISPR-Cas9 system. Compounded, they further reinforced the necessity of an alternate system. Alternative CRISPR-Cas9 systems do exist and, in some cases, have been demonstrated to work in *E. limosum*. The KAIST CRISPR-Cas9 system (315) creates successful knockouts in a rapid fashion. It, however, has the significant

drawback of inserting a selection marker into the genome, limiting the potential for double or triple knockouts, potentially causing unintended effects upon the cell, and limiting non-laboratory uses of cells. Of note Jongoh Shin and co-workers have made use of CRISPR to both knock genes out and to modulate gene expression using CRISPRi (315).

The RiboCas CRISPR-Cas9 system (218) is an alternative system that has had much success in many other organisms, but despite many attempts, function was never observed in *E. limosum*. The failure of RiboCas in *E. limosum* is likely due to the system making use of riboswitches which had not been assayed for their activity in *E. limosum* at the time of writing. While such work may allow this system to be functional, it was pursued as part of this study, though its success in other organisms may warrant future research into its function in *E. limosum*.

Attempts to employ the native CRISPR-Cas3 system of *E. limosum* were not successful. This suggests that more than a surface-level analysis would need to be performed to identify the steps needed to get this system working in *E. limosum*. Were the issues to have been resolved, however, this CRISPR-Cas3 system has issues associated with guide generation, as no algorithm (as exists for CRISPR-Cas9) to design high-quality guides currently exists. The selection of guides would be essentially arbitrary. Development of such an algorithm is far beyond the scope of this study.

With the failure of CRISPR-Cas based systems, an alternative counter-selection system was investigated (227). The toxin-antitoxin system is an efficient, rapid, and agile system that renders CRISPR obsolete in *E. limosum* in terms of generating gene knockouts. Indeed, mutant generation in *E. limosum* via the toxin-antitoxin method is more limited by the speed of plasmid construction than by the process of mutant generation, allowing mutants to be rapidly made in a short period of time. All the mutants in this study created with the toxin-antitoxin system were generated in less than 6 months.

While the rapid nature of mutant generation with this toxin-antitoxin system is a marked improvement from the CRISPR-Cas9 system, it highlighted a glaring issue: the lack of a complete list of essential genes in *E. limosum*. The TraDIS system (1.5.3.4) would allow for a determination of gene essentiality for the whole genome, and it was hoped that this work (performed by other researchers) would have been complete long before this work was near its end. The lack of it demonstrates the value that such a system would have on a piece of research. The failures demonstrated in creating Δ*metVmetF* are likely a result of gene essentiality, based upon the TAKOS effectiveness with other targets; this target would have been ignored had their essentiality been known beforehand. While on the surface, this may seem like a failure, it, in fact, demonstrates one of the systems unforeseen advantages over CRISPR-Cas9, that being an inability to generate "false" knockouts, with reintegration into the genome elsewhere or by maintenance on a plasmid conferring function.

The issues surrounding the *act/bcact* (4.3.4.2.1/4.3.4.8) highlight the essentiality of good genome annotation but, moreover, show the usefulness of TraDIS. At the point this issue was encountered, work by other researchers had proceeded such that gene essentiality was somewhat understood. In the case of NG-6849, M23_RS20600, *bcact*, TraDIS data (Figure 117) showed no insertions on either methanol or glucose growth conditions, which suggested gene essentiality. If no inserts within a gene are possible, then it is likely essential for growth grow, or insertion causes a lethal product.

4.5 Chapter conclusions and future work

This chapter identified significant flaws with the CRISPR-Cas9 system used. This is of key importance as regards future work with *E. limosum*. While this CRISPR-Cas9 system could be improved to function with the efficiency and stability of comparable CRISPR-Cas9 systems present in other organisms, at its present state, it is not fit for purpose. Alternative counter-selection systems are better employed for genetic modification of *E. limosum*.

The efficacy of the toxin-antitoxin system has been demonstrated in this chapter, generating stable clean mutants. The system is mature, and its protocol rapid and effective, allowing for easy creation of mutants in parallel in a short time frame. Rapid plasmid loss, due to the use of suicide plasmids, allows for shorter times between the creation of additional mutations in the same background. However, there is still scope for improvement, with multiplexing being a desirable feature that has not been demonstrated in this system. The success with anhydrotetracycline induction (4.3.4) would suggest that it, rather than lactose, should be employed as the inducer in future work when any new inducible systems are developed or adapted for use in *E. limosum*. Its inability to be utilised as a carbon source is a further advantage in addition to its proven function.

Many mutants were generated in this chapter, including Δ*act*, Δ*act*Δ*pbt1*, Δ*act*Δ*pbt2*, Δ*act*Δ*pbt1*Δ*pbt2*, Δ*cop*, Δ*hisI*, Δ*metF1*, Δ*mti*, Δ*mtiI*, Δ*pbt1*, and Δ*pbt2*. These mutant backgrounds were able to confirm the expected function of the *cop*, *mti*, and *mtiI* genes, with all being involved in the uptake of methanol into the Wood-Ljungdahl pathway, as was predicted from their homologues in *A. woodii*. These mutants also resulted in the rejection of *act*, *pbt1*, and *pbt2* as being involve in butyrate synthesis with no mutant background generated being sufficient to perturb butyrate production. The inability to create a strain that does not produce butyrate is perhaps not unexpected given that said route results in ATP production. However, it does raise the question of whether ATP-producing reactions could be introduced further up the Wood-Ljungdahl pathway to make said phenotype possible in an aim to direct more inputted carbon towards desired products. As this work's original goal surrounded the

production of high-value chemicals, using this genetic modification systems to introduce production pathways was the natural next step. Designing synthetic pathways or inserting pathways that exist in other organism could extend the product spectra beyond that of the native phenotype. Indeed, acetone production was achieved as part of this work (Chapter 5).

Chapter 5
Creation of a chromosomal integrant acetone producing strain of *E. limosum* NG-6849

5.1 Introduction

In an attempt to meet the stated goals of this work, primarily the production of value-added chemicals, and to demonstrate the various genetic modification tools that had been developed for *E. limosum*, genomic integration of genes to produce a non-native product, was performed. *E. limosum* had very limited free energy in the form of excess ATP (75,121), which could be used to produce non-native products, even assuming excess carbon, nitrogen, etc. Therefore, any introduced pathway would need to either produce ATP (as the native butyrate pathway does) or be ATP neutral. Furthermore, a starting chemical in the Wood-Ljungdahl pathway would need to be identified that could be turned into the desired chemical. *E. limosum* lacks a native acetone production pathway but given its ATP neutral cost and the native presence of acetoacetyl-CoA, an essential foundation for acetone production (Figure 91), as part of the *E. limosum* Wood-Ljungdahl pathway, the introduction of exogenous genes to allow for its production seemed like a viable task.

Acetone is an important chemical both as a solvent and as a chemical precursor. It had previously been produced from anaerobic fermentation (24,331), so was a viable target chemical an acetogenic bacterium (332). Acetone-butanol-ethanol fermentation using *Clostridium acetobutylicum* was previously developed and commercialised by Chaim Weizmann (247), and while fermentation on molasses is no longer financially viable, acetone production from fermentation is still desirable. The existence of acetone-producing acetogenic bacteria, such as *C. acetobutylicum* (Figure 91), would allow for exogenous genes to be harvested and integrated into the *E. limosum* genome to allow for production (Figure 92). By using *E. limosum* as a base organism, it may be possible to produce acetone while growing on a mix of CO_2 gas and methanol or on a mix gas feed of CO/CO_2. The *E. limosum* NG-6849 strain was used as the basis for the acetone work. Its ability to grow on a mix of CO and CO_2, on methanol, and on glucose offers greater feedstock adaptability, which could have advantages in an industrial setting.

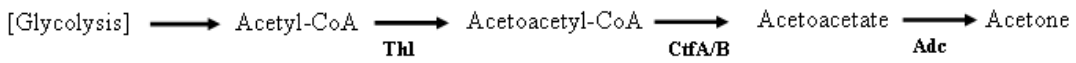

Figure 91 - Acetone pathway in C. acetobutylicum

A simplified pathway displaying the enzymes at each step in the production of acetone, based on Li et al 2011 (333). *Thl; Thiolase, CtfA/B; coenzyme A transferase, Adc; acetoacetate decarboxylase.*

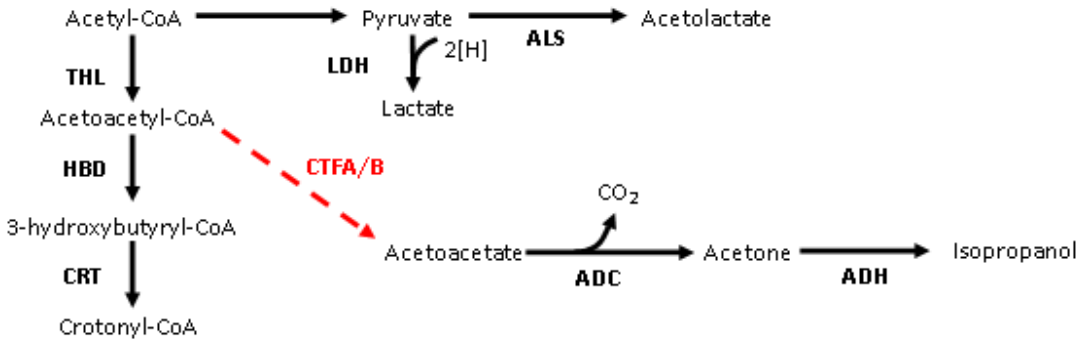

Figure 92 - Native acetone pathway in E. limosum

The native, incomplete, pathway present in E. limosum, broken at acetoacetyl-CoA due to lacking homologues for cfta and cftB. ADC, acetone decarboxylase; ADH, alcohol dehydrogenase; ALS, acetolactate synthase; CRT, crotonase; CTFA/B, coenzyme A/B transferase; HBD, 3-hydroxybutyryl-CoA dehydrogenase; LDH, lactate dehydrogenase; THL, thiolase Figure adapted from Humphreys & Minton (109).

A BLASTp search against the *E. limosum* genome with the acetone operon proteins from *C. acetobutylicum* showed that *E. limosum* appears to lack a key enzyme required for acetone production, namely for the conversion of acetoacetyl-CoA to acetoacetate, as indicated by the lack of homologues encoding its subunits *cfta* and *cftB* (Figure 92). The generalised Wood-Ljungdahl pathway presented earlier can be extended to enable the production of acetone and isopropanol (Figure 92).

5.2 Aims

The aim of this chapter was to introduce a set of genes to allow for the completion of the truncated native acetone production pathway into the genome of *E. limosum*. This would allow for the expansion of the native product spectrum whilst also making use of the previously developed tools and auxotrophic mutants, allowing for further validation of the prior work. In addition to the production of acetate, the native presence of a secondary alcohol dehydrogenase would likely result in isopropanol production.

5.3 Integrating acetone production into *E. limosum*

In previous work at the University of Nottingham, *Acetobacterium woodii* was engineered to produce acetone (Dr. Jon Baker & Dr. James Millard, University of Nottingham, personal communication). As part of that work, a plasmid, pMTL84151-acetone-thl (24-1), containing the *ctfA*, *cftB*, *thl*, and *adc CDS* genes from *C. acetobutylicum* was developed (Figure 93). Once transformed, this plasmid enabled *A. woodii* to express the *C. acetobutylicum* genes and produce acetone. Due to *A. woodii* having a high degree of physiological similarity to *E. limosum*, it was hoped that these genes would work in *E. limosum* to produce acetone.

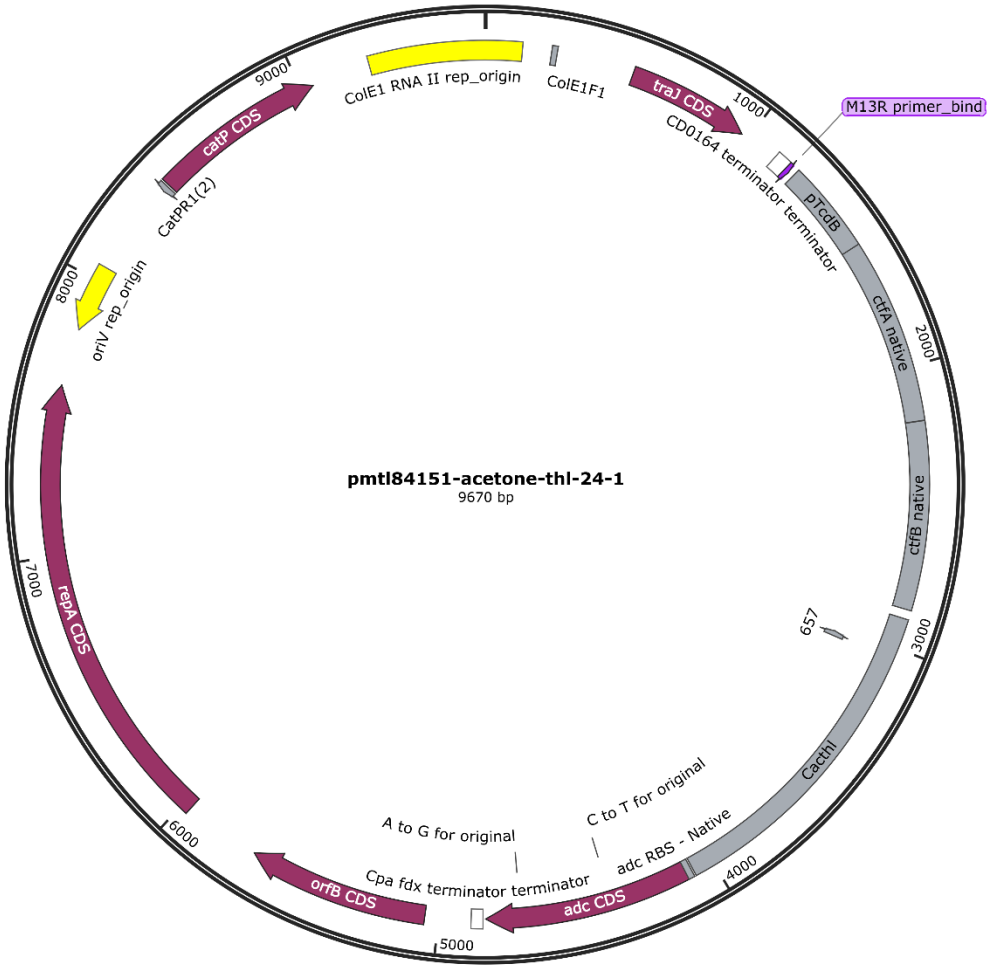

Figure 93 - pMTL84151-acetone-thl (24-1)

Acetone production plasmid developed by Dr. Jonathan Baker at the University of Nottingham for use in A. woodii (unpublished). Contains the acetone production genes ctfA, ctfB, a thiolase gene (thlCac) and an acetone decarboxylase (adc). pTCdB is a tcdr based promoter, this necessitates the use of a background strain able to make use of a tcdr system. The plasmid confers resistance to chloramphenicol and thiamphenicol via the catP gene.

Genes from *C. acetobutylicum* could be inserted at the *hisI* locus using allele-coupled exchange (ACE), in which restoration of histidine prototrophy would be coupled to insertion of the desired genes (4.3.2). This required a background strain with a truncated *hisI* gene. The functionality of the acetone genes and their associated regulatory elements in *E. limosum* was unknown, and their level of performance would determine acetone production levels. Initial testing of the acetone production genes on a plasmid was, therefore, seen as a prudent piece of preliminary work prior to the creation of an ACE integrant. pMTL84151-acetone-thl (24-1), Figure 93, however, would not work as-is in *E. limosum*. It contained an incompatible replicon, so could not propagate within *E. limosum*. The plasmid also made use of a *tcdR* based promoter system which would not work in a wild-type *E. limosum* background. While a *tcdR* mutant could have been made, it was be easier to alter the plasmid to work in the available strains. Two elements needed to be changed in this plasmid, the promoter for the acetone genes, and the replicon for maintenance in the anaerobe. However, a simpler approach was first attempted. The acetone genes were to be digested out of the vector and inserted into a new one with the correct promoter and replicon. A vector that had previously been used successfully was selected, pSBRCas9-Cpf-*act*.

5.3.1 pSBRCas9-Cpf-*act*-based acetone vector

Using pSBRCas9-Cpf-*act* as the starting construct required removal of the Cas9 and Cas9-associated regions. This was done using restriction digestion (2.2.3). The acetone production genes were removed from pMTL8415184151-acetone-thl (24-1) using restriction digest (using NdeI and AatII) and ligated into the modified pSBRCas9-Cpf-*act* vector (2.2.3). Ligations were then introduced into *E. coli* (2.2.1), extracted (2.2.2), and sequenced (2.2.7) using the sequencing primers Seq Primer Acetone Step 4 Primer 1, Seq Primer Acetone Step 4 Primer 2, Seq Primer Acetone Step 4 Primer 3, Seq Primer Acetone Step 4 Primer 4, Seq Primer Acetone Step 6 Primer 1, and Seq Primer Acetone Step 6 Primer 2 (Table 26).

Sequencing coverage was poor, only primers that read into the ends of the operon produced good reads (Seq Primer Acetone Step 2 Primer 1 and Seq Primer Acetone Step 6 Primer 1). The assumption was made that if both ends had inserted correctly into the plasmid, and that the rest had gone in. Analytical digests (2.2.3) confirmed the expected product size for the insert. The new complete plasmid was designated pSBR-CpfAcetone (Figure 94).

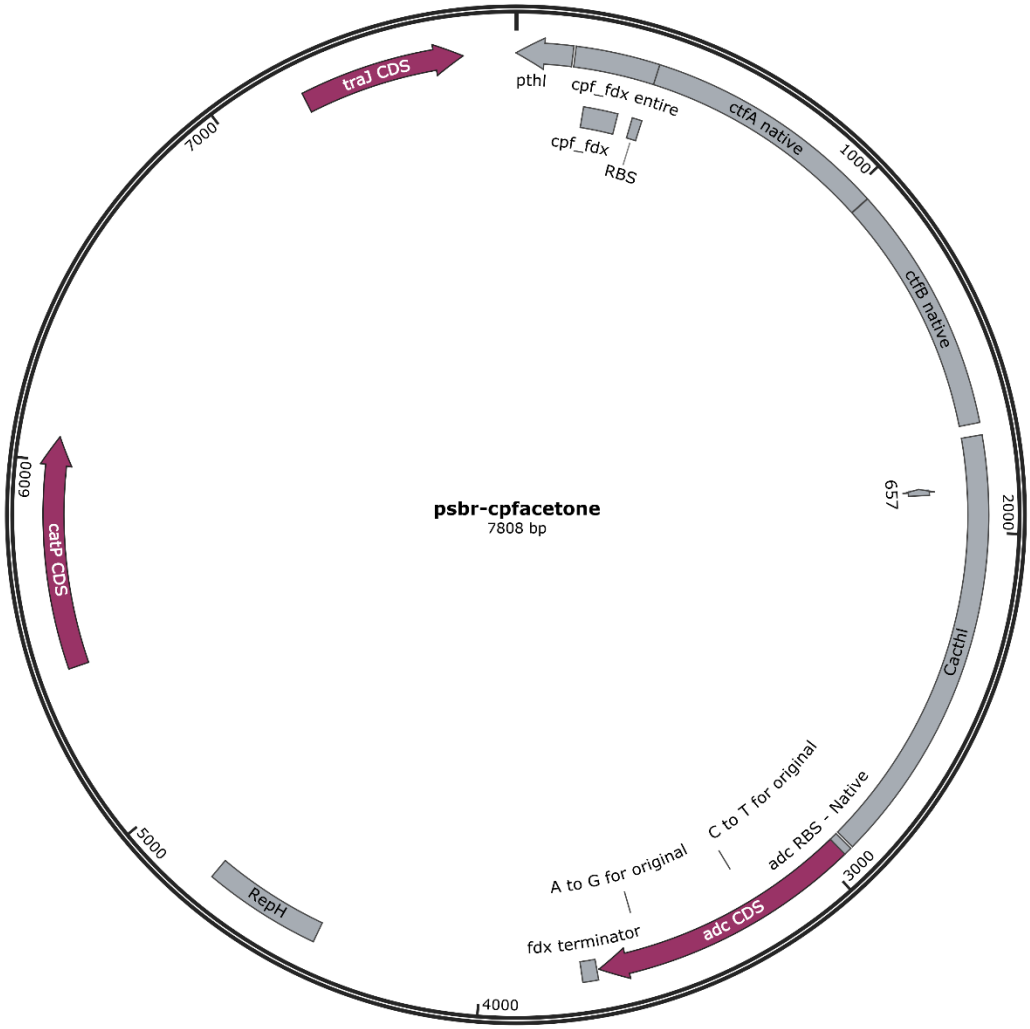

Figure 94 – pSBR-CpfAcetone

The complete acetone test vector, with the acetone production genes, ctfA/b, thlCac and adc, driven by the Pfdx promoter. The residual thiolase promoter (Pthl), has no function in this design; it was not removed due to the presence of an XbaI site in the thlCac gene. The plasmid confers resistance to chloramphenicol and thiamphenicol via the catP gene.

5.3.2 Testing pSBR-CpfAcetone functionality in *E. limosum*

The purpose of putting these genes onto a plasmid where they would be constitutively expressed was to ensure the operon functioned as expected prior to any integration work. The sequenced plasmids were transformed into wild-type *E. limosum* NG-6849. Cells were grown on plates, and then a single colony was picked and grown in a starter culture of Pacaud medium (2.1.1.2) with 30 mM glucose and 2 g/L yeast extract. Once the cell mass was sufficiently high, three serum bottles, biological replicates, were set up (2.1.2.1) and cells inoculated into each. 100 mL of Pacaud medium (2.1.1.2) with 100 mM glucose, and 2 g/L yeast extract was used. Antibiotic selection of thiamphenicol was maintained to ensure maintenance of the plasmid. Cells were sampled every 24 hours for spectrophotometry analysis of optical density (OD_{600}), and samples to perform HPLC were taken in tandem (Figure 95).

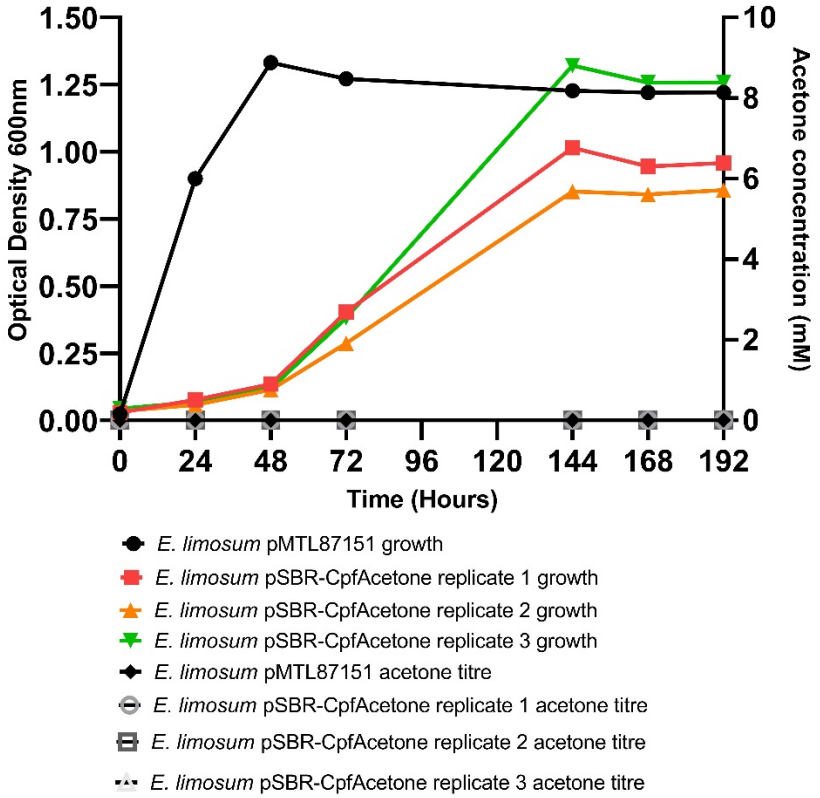

Figure 95 - Growth of E. limosum NG-6849 bearing pMTL87151 and E. limosum NG-6849 bearing pSBR-CpfAcetone.

E. limosum NG-6849 bearing pMTL87151 and E. limosum NG-6849 bearing pSBR-CpfAcetone were grown in Pacaud medium on 30 mM glucose in batch culture in serum bottles with 2 g/L yeast extract, mM thiamphenicol and an anaerobic gas headspace (~8% hydrogen, ~77% N_2, ~15% CO_2) of 200 mL. Spectrophotometer readings for absorbance at 620 nm and samples for HLPC were taken every 24 hours. E. limosum NG-8869 bearing pMTL87151 was performed as a singlet control. E. limosum NG-6849 bearing pSBR-CpfAcetone was grown as three biological replicates.

Differences in growth profiles between *E. limosum* NG-6849 bearing pMTL87151 and *E. limosum* NG-6849 bearing pSBR-CpfAcetone suggest a greater burden for the maintenance of pSBR-CpfAcetone over pMTL87151, likely due to the constitutively expressed genes on the pSBR-CpfAcetone plasmid. However, HPLC samples showed no production of acetone, raising concerns about the functionality of the *C. acetobutylicum* template used to create this plasmid, since the acetone production genes had been successfully been used in *E. limosum* previously (117). It was hypothesised that errors in either the plasmid or the operon might be the cause, so new primers were designed (Acetone_full_forward_1, Acetone_full_reverse_1, Acetone_full_forward_2, and Acetone_full_reverse_2) and sequencing (2.2.7) was performed on pSBR-CpfAcetone to ensure no mutations or other differences from the in silico plasmid map. Although, sequencing data was incomplete the successfully sequenced regions appeared to be as expected, suggesting an issue with the operon itself.

5.3.3 Codon-optimised *C. acetobutylicum* acetone production genes

Given the lack of any apparent sequence issues in pSBR-CpfAcetone, the acetone production genes' coding sequences from pMTL84151-acetone-thl (24-1) were examined further. It is possible that the *C. acetobutylicum* genes did not function as expected in *E. limosum*, perhaps due to codon usage differences between the organisms. Although codon usage in *E. limosum* is not well-characterised, a codon usage report (Codon OptimWiz®, previously unused data, Dr. James Millard, University of Nottingham, personal communication) suggested that the arginine codons AGG and CGA were not utilised by *E. limosum*. These codons were present three times within this acetone functional region, all within *cfta*.

A synthetic version of the functional region was created (2.2.10), with *ctfA* being created using reverse translation using *E. coli*-optimised-codons (all of which seem to be used in *E. limosum*). The other acetone production genes contained structures that may have hindered synthesis, such as hairpins. This required the substitution of

codons to reduce secondary structure; this was combined with the substitution of theoretically unused codons (Figure 96).

Figure 96 - Alignment between synthetic codon optimised ctfA and native ctfA genes

The misalignments are noted in by the dots in place of lines, with alignment shown by lines; the amino acid sequence however is identical despite the nucleotide differences.

The synthetic operon was inserted into the pMTL87151 vector using NEBuilder® HiFi DNA Assembly (2.2.9). pMTL87151 was chosen as the backbone vector due to its high stability in *E. limosum*, as it is based on the p19 replicon derived from *C. carboxidivorans* (Dr. James Millard, University of Nottingham, personal communication). The use of a highly stable plasmid was crucial due to the difficulty of growing *E. limosum* on methanol with antibiotic selection. Despite the provision of resistance genes, it was not possible to grow *E. limosum* on either thiamphenicol or ampicillin selection with methanol as the carbon source; this issue was also encountered by BioMetChem partners (Dr. Maximilian Flaiz, Ulm University, personal communication). A highly stable plasmid can be maintained without selection for a sufficient period to allow for characterisation.

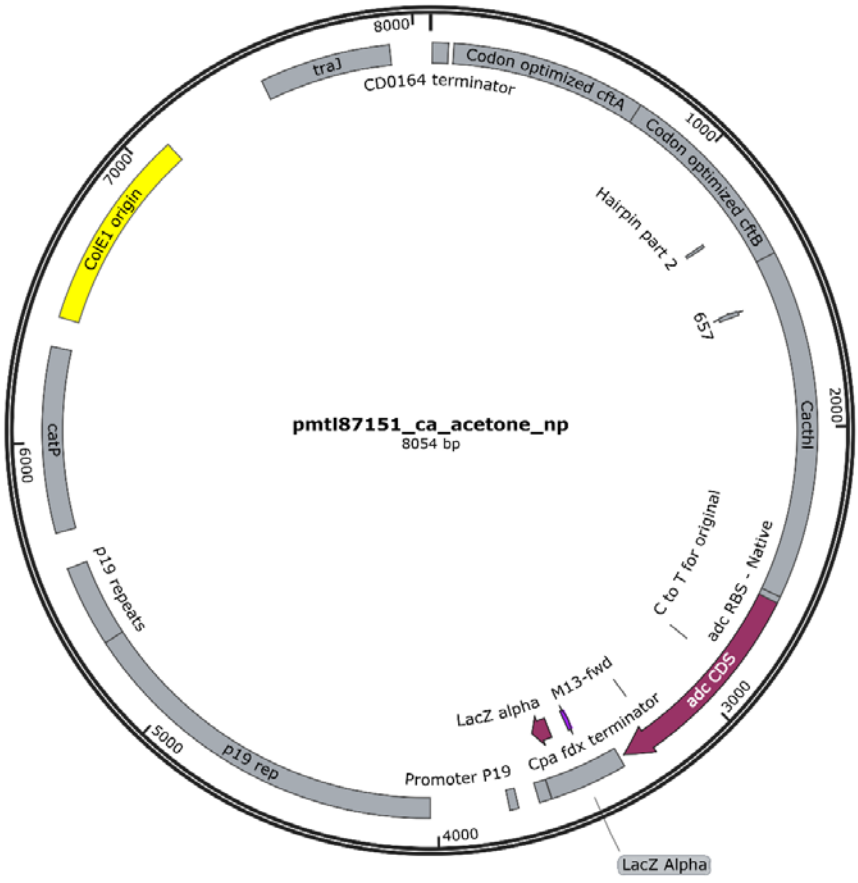

Figure 97 - pMTL87151_CA_acetone_NP, a maintenance plasmid containing the synthetically derived codon optimised acetone production genes from C. acetobutylicum.

The base plasmid was pMTL87151. The codon optimised operon from C. acetobutylicum was synthesised and inserted by HiFi with the codon optimised ctfA/B (encoding the CoA transferase A and B subunits) in red and purple, respectively. The acetone production genes thlCac and adc, encoding a thiolase and acetoacetate decarboxylase, respectively, followed the cfta/b. The plasmid confers resistance to chloramphenicol and thiamphenicol via the catP gene.

Initially, a pMTL87151 maintenance plasmid without promoters expressing the operon was constructed (Figure 97). This was done to ensure that the plasmid was not causing any issues unrelated to the promoters in *E. limosum*. Once assembled and confirmed as being sequence correct, one of the strongest promoters available, the *C. perfringens* P_{fdx} promoter, was amplified (using the primers P_{fdx}_Acetone_fwd and P_{fdx}_Acetone_rev) and inserted into the vector using NEBuilder® HiFi DNA assembly (2.2.9). The rationale for using this promoter was that it had previously worked with CRISPR-Cas9 (4.1). Due to the promotor being of high strength it was hoped that it would maximise expression and maximise yields. This resulted in the creation of a new plasmid, pMTL87151_CA_acetone_P_{fdx}, as show in Figure 98.

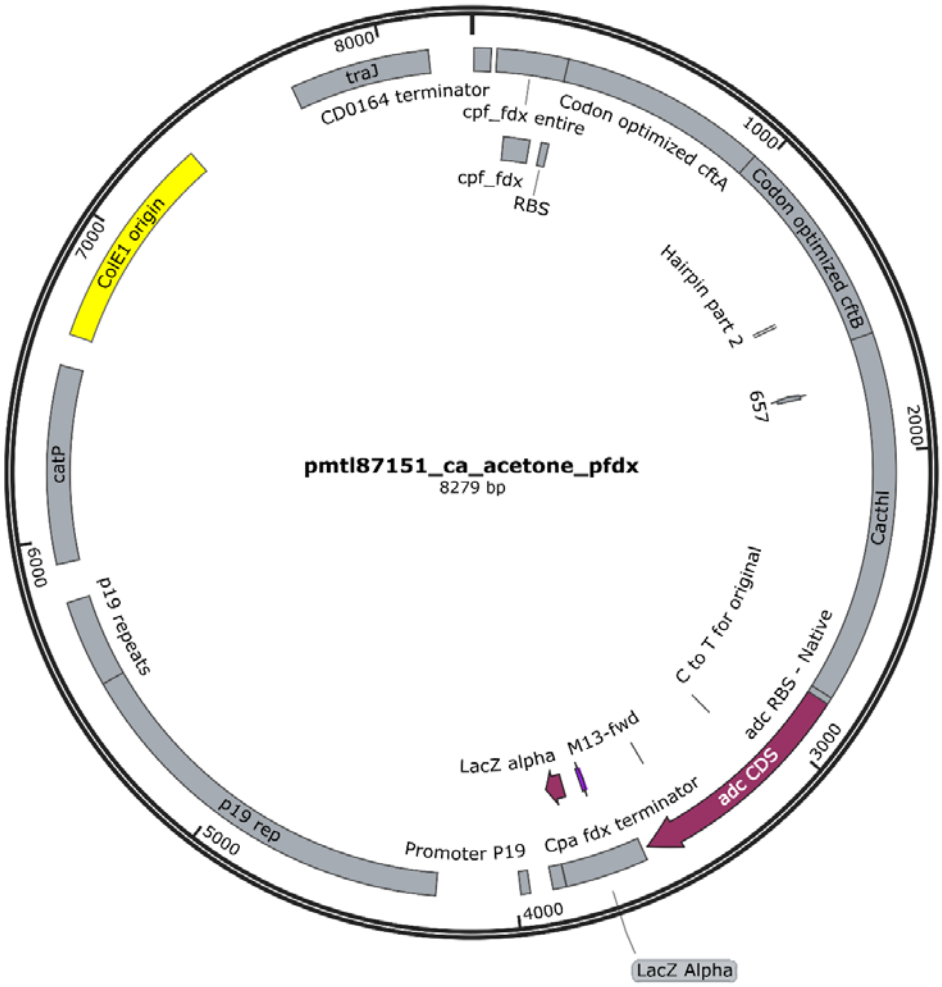

Figure 98 - pMTL87151_CA_acetone_P_{fdx}, a P_{fdx}-driven acetone production genes containing plasmid.

The base plasmid was pMTL87151. The codon optimised operon from C. acetobutylicum was synthesised and inserted by HiFi. with the codon optimised ctfA/B in red and purple, respectively. The acetone production genes thlCac and adc, encoding a thiolase and acetoacetate decarboxylase, respectively, followed the cfta/b. The plasmid confers resistance to chloramphenicol via the catP gene.

5.3.4 Testing of P_{fdx}-driven acetone vectors

Once the sequence of pMTL87151_CA_acetone_Pfdx was confirmed (2.2.7), it was introduced into *E. limosum* NG-6849 (2.3). Serum bottle work was then performed in triplicate using glucose as a feedstock, as standard (2.1.2.1) to test the functionality of the plasmid. Unfortunately, pMTL87151_CA_acetone_Pfdx did not result in acetone production (data not shown). Therefore, existing vectors that other researchers had made were tested to investigate the cause of the failure.

The pMTL8J15 Acetone 77 thl and pMTL8J15 Isopropanol 77 thl vectors had previously produced acetone in *E. limosum* ATCC 8486 (117), although such work had not been performed in the NG-6849 strain. These plasmids were similar to pMTL87171_CA_acetone_Pfdx with the primary difference being that the *ctfA* and *ctfB* (primarily *ctfA*) in pMTL87171_CA_acetone_Pfdx were codon-optimised, and there was a significant difference in the nucleotide sequence (Figure 96), although the amino acid sequence was identical. The pMTL8J15 Isopropanol 77 contained a secondary alcohol dehydrogenase in addition to the acetone production genes, but all other elements were the same. The pMTL8J15 Acetone 77 thl and pMTL8J15 Isopropanol 77 thl plasmids were introduced into wild-type *E. limosum* NG-6849 strain (2.3), and growth curves were performed as standard (2.1.2.1) on glucose with thiamphenicol selection (2.1.1.4) to ensure plasmid maintenance, with time points (Figure 99) and samples for HPLC analysis (2.1.2.4.2) taken every 24 hours and at the 0-hour point (Figure 99).

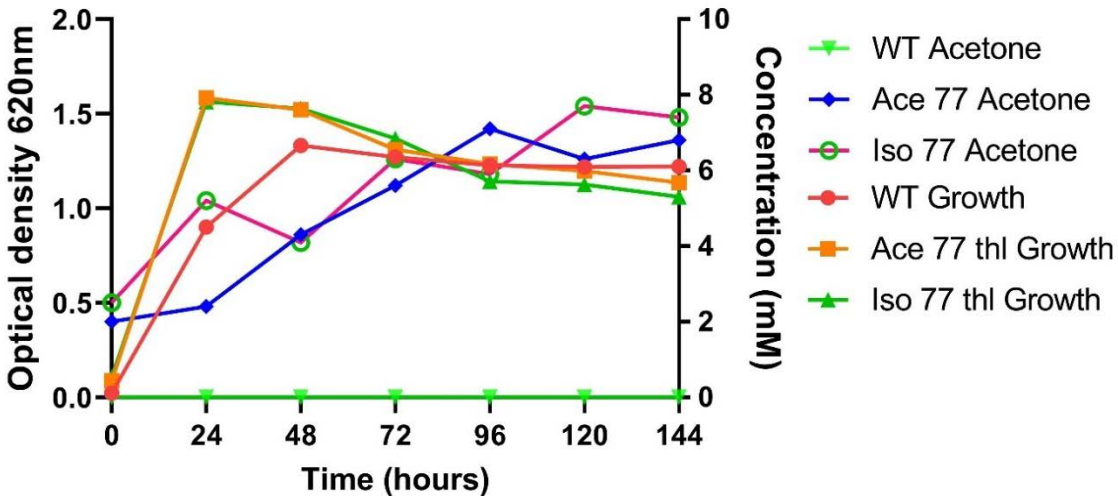

Figure 99 - Growth of E. limosum NG-6849 wild-type and NG-6849 maintaining pMTL8J15 Acetone 77 thl and pMTL8J15 Isopropanol 77 thl.

Wild-type E. limosum NG-6849 and strains maintaining pMTL8J15 Acetone 77 thl and pMTL8J15 Isopropanol 77 thl were grown in batch culture on Pacaud medium in serum bottles with 15 mM glucose as the substrate and an anaerobic gas headspace (~8% hydrogen, ~77% N_2, ~15% CO_2) of 200 mL. Spectrophotometer readings for absorbance at 620 nm and samples for HLPC were taken every 24 hours. Growth was performed with antibiotic selection. E. limosum NG-6849 growth was done as singlet. E. limosum NG-6849 maintaining pMTL8J15 Acetone 77 thl and E. limosum NG-6849 maintaining pMTL8J15 Isopropanol 77 thl were grown as triplicate with mean of values shown.

The growth of the pMTL8J15 Acetone 77 thl and pMTL8J15 Isopropanol 77 thl plasmid-maintaining *E. limosum* NG-6849 strain and that of the wild-type *E. limosum* NG-6849 strain were mostly comparable (Figure 99), with no growth disadvantage conveyed to cells by plasmid maintenance. Both plasmid maintaining strains produced acetone (Figure 99). The initial concentration of acetone was unexpectedly above 0 mM, likely due to carryover from the precultures. Although acetone production was lower (~7 mM) than that observed in *A. woodii* (~15-20 mM), where the operon had been used previously (332), it indicated that the operon was functional, with wild-type cells producing no acetone. Once it was clear that the operon was functional in *E. limosum*, integration into the genome was performed to allow for comparison with the plasmid-based operon, with the expectation that there would be less production in the integrant due to the lower gene copy number relative to the multi-copy plasmid.

5.3.5 Integration of the acetone production cluster into the *E. limosum* NG-6849 chromosome

The maintenance of plasmids using antibiotic selection is not desirable for industrial strains, nor would it be viable for growth on alternative feedstocks such as methanol. Integration of the acetone production genes into the genome would negate the need for antibiotic selection to maintain the non-native pathway. pACE_*hisI*_Acetone_KI (Figure 100) was designed, which would integrate the functional acetone production operon under the control of *C. sporogenes* P_{fdx} at the *hisI* locus, using ACE (2.6.3) (159,164).

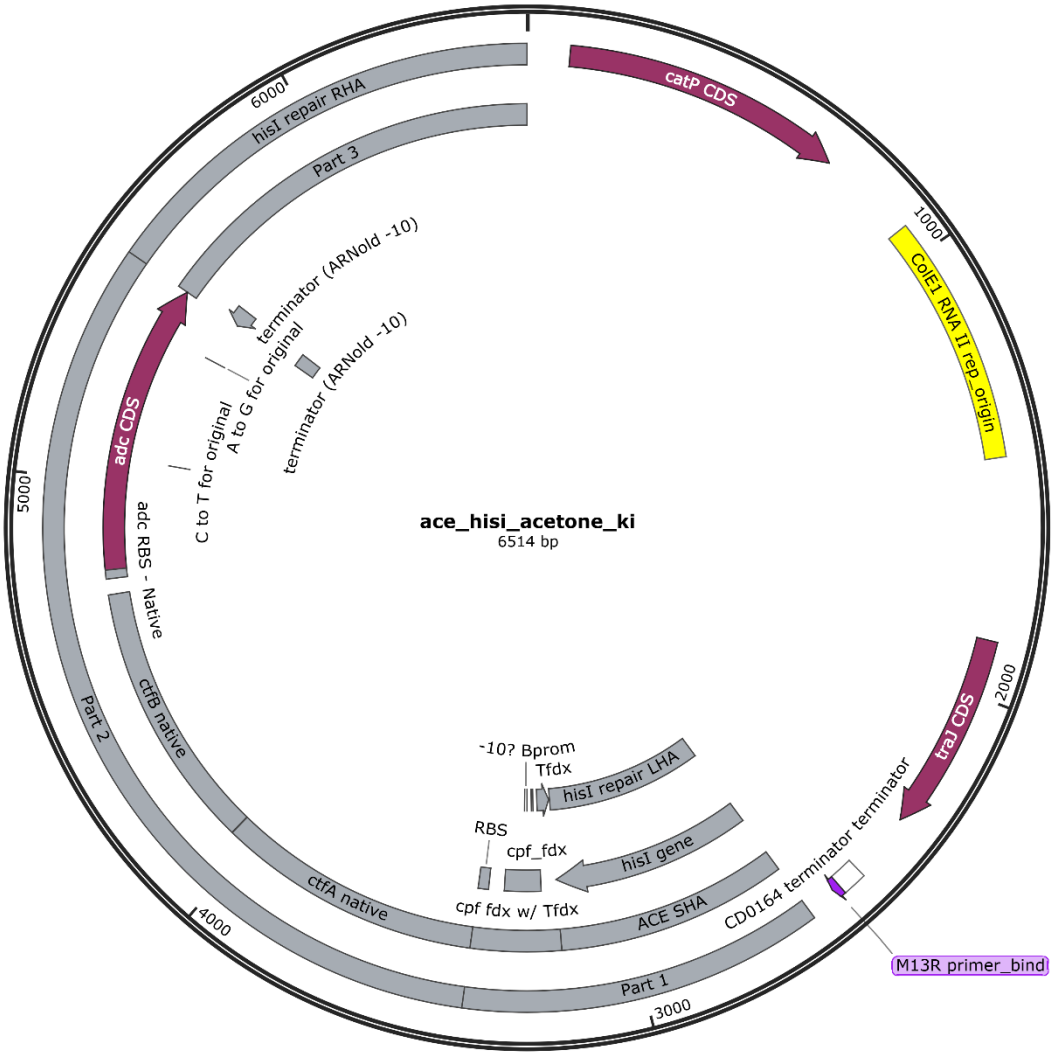

Figure 100 - pACE_hisI_Acetone_KI, an ACE vector to integrate the acetone production genes from C. acetobutylicum at hisI in an auxotrophic ΔhisI mutant of E. limosum.

hisI repair sections (part 1 and part 3) allow for restoration of hisI gene, the size disparity to ensure correct order of crossover. Part 2 contains the ctfA/B and adc (acetone decarboxylase), the thiolase was not included due to E. limosum possessing a native thiolase.

pACE_*hisI*_Acetone_KI was introduced into *E. limosum NG-6849* Δ*hisI* (2.3), and several transformants were recovered. The emergent colonies were streaked onto histidine-deficient medium and allowed to grow. Colonies that arose on the deficient medium were screened for integration of the acetone pathway genes using colony PCR (Figure 101).

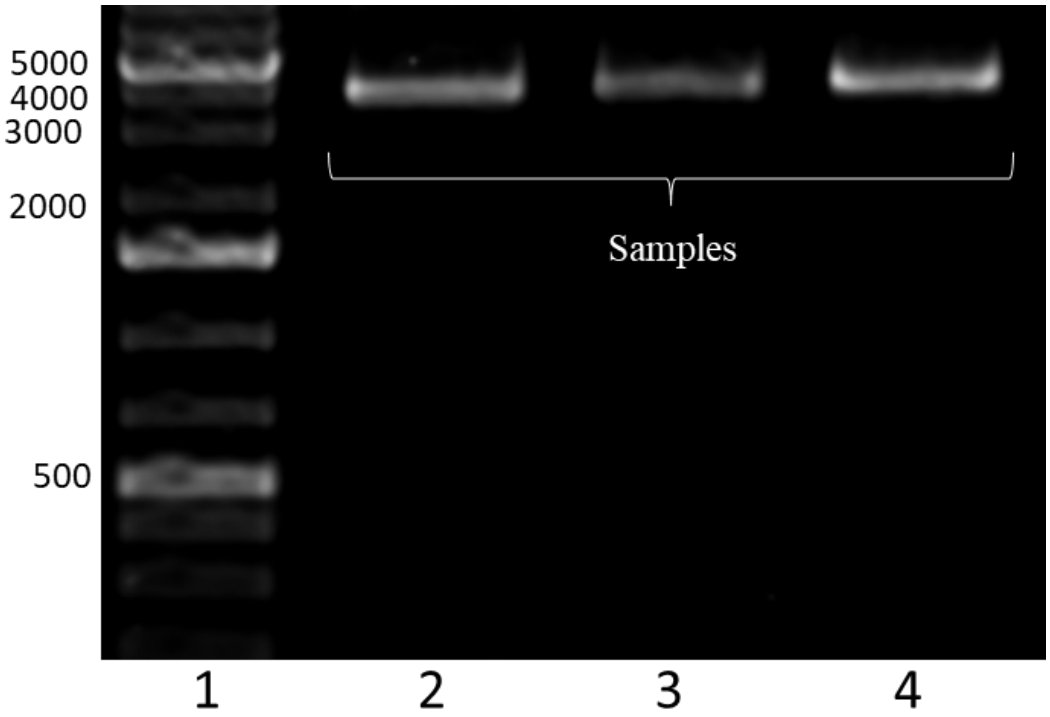

Figure 101 - Screening for E. limosum for the insertion of acetone pathway genes at the hisI locus by colony PCR screening

A standard OneTaq colony PCR was performed on 3 colonies of ΔhisI E. limosum NG-6849 transformed with ACE_hisI_Acetone_KI. Primers (hisI integrant R and hisI integrant R) were used to screen for the integration of the acetone production genes (lanes 2 - 6). A control was not used,as the primers bind outside of the repair regions (to prevent plasmids producing products). The ladder used (in lane 1) was GeneRuler 1 kb plus DNA Ladder, samples were run on a 0.8% agarose gel. Colonies 2 - 4 appeared to have integrated the region correctly with a product of the appropriate size.

All cell lines had regained the native phenotype of histidine prototrophy. As such, the high effective efficiency of integration was as expected since Δ*hisI* mutants cannot grow without the supplementation of histidine into the growth media. The only cells that would survive are those that had regained *hisI*, which could only occur through the proper integration of the acetone production genes. Once colonies were isolated, and restoration of phenotype confirmed, growth assays were performed. An *E. limosum* integrant mutant had been created with acetone pathway genes that had worked when expressed from a plasmid. This strain was known as *E. limosum*

Ace_KI (acetone knock-in). Growth assays were performed to assess the functionality of these genes when expressed from the chromosome. These were performed as detailed previously with triplicate growth of the NG-6849 *E. limosum* Ace_KI strain and a control of wild type NG-6849 *E. limosum.*

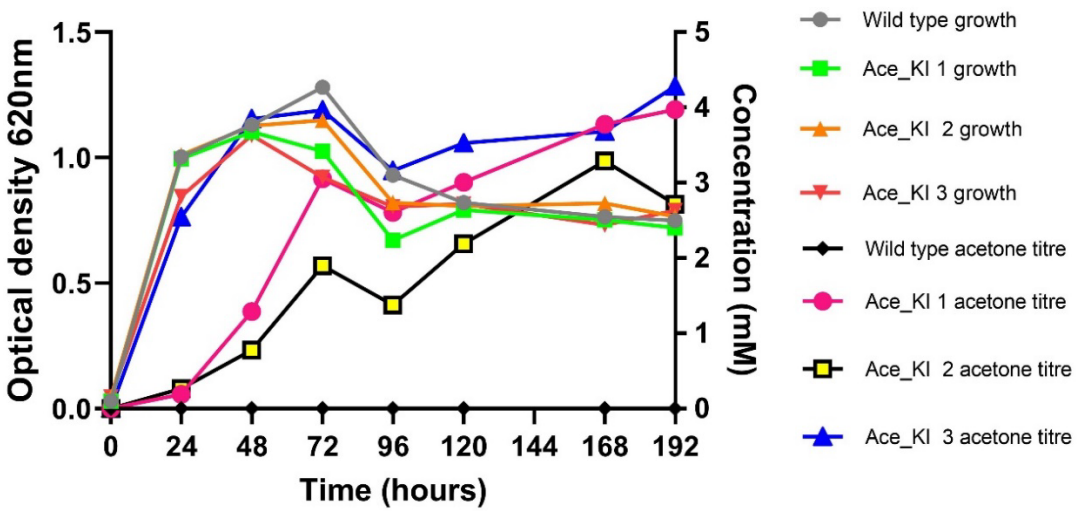

Figure 102 - Growth of E. limosum NG-6849 wild-type and NG-6849 Ace_KI E. limosum NG-6849 strain on glucose.

E. limosum NG-6849 wild-type and Ace_KI were grown in Pacaud medium in batch culture in serum bottles with 15 mM glucose as the substrate and an anaerobic gas headspace (~8% hydrogen, ~77% N_2, ~15% CO_2) of 200 mL. Spectrophotometer readings for absorbance at 620 nm. Samples for HLPC were taken every 24 hours. NG-6849 E. limosum Ace_KI was grown in triplicate with a control of wild type NG-6849 E. limosum grown as a single flask.

The growth of the integrated strains was nearly identical to that of the wild-type *E. limosum* (Figure 102), indicating no burden conferred by the integration of the genes or the production of any acetone. Acetone production (Figure 102) was detected in all integrated replicates, with none detected in the wild type. The titres were low (~3 - ~4 mM) but detectable.

5.4 Discussion and future work

Initially (5.1.3) acetone production was attempted using a plasmid maintaining a synthetically derived *C. acetobutylicum* acetone operon. This operon was created using IDT (Coralville, IA, USA) gBlocks™, as previously used for assembling of CRISPR-Cas9 vectors in this study (4.2). The native acetone production genes from wild-type *C. acetobutylicum* were too complex to synthesise, so silent modifications were made to reduce structures that were problematic for gene synthesis. This synthesis sequence optimisation allowed for the synthesis of these genes and allowed for plasmids with correct sequences to be created. However, the resulting plasmid, pMTL87171_CA_acetone_Pfdx, did not result in any detectable acetone production.

Previous work by other researchers at SBRC used acetone production genes expressed from plasmids and resulted in the production of plasmids that could be used as templates for this work (334). These operons had not previously been integrated into the *E. limosum* NG-6849 strain. Integration of the functional acetone production genes, *ctfA/B,* and *adc*, under the transcriptional control of the strong P_{fdx} promoter using ACE was successful, with maximum acetone concentrations of approximately 4 mM while growing on glucose.

Acetone production levels varied between plasmid-maintaining strains and integrant strains, with the plasmid-maintaining strain having approximately double the titres of the integrant strain. (5.3.4). Integration is likely to lead to a lower copy number of expressed genes since only one copy can exist per integrant cell, where plasmids can be present at greater than one copy per cell. It would, therefore, be expected that production would be lower in integration strains when compared to those maintaining genes on plasmids. This is what the data generated as part of this study shows and is, therefore, consistent with the literature. If the inverse were the case, it is likely that the burden of plasmid maintenance, more than one copy per cell, or some element of the plasmid itself, would be responsible for lower titres. Titres were poor in both cases, with industrial strains of *C. acetobutylicum* producing ~5 g/L of acetone (335). Titres in the plasmid strain were maximally 8 mM in 30 mL, or approximately 0.46 g/L.

Growth assays on glucose were performed as part of this study, however, additional assays on methanol or on a CO and CO_2 gas mix were not conducted due to time constraints. Such data would be useful, not only due to potentially altered yields (butyrate yields were higher on methanol (4.3.4.7)), but also because the interest in *E. limosum* is its ability to grow on methanol. Other acetone producing organisms, which have much greater degrees of development and potential (331,332,335), can grow well on glucose. Methanol growth conditions would also be of value to test the viability of using a high stability p19 plasmid without selection to produce a target chemical, given the issues with antibiotic selection on methanol growth. Performing growth assays in fermenters would be of value to both assess if fermentation conditions improve titres and whether adaptive or fed batch fermentations improve titres. Fermentation would also allow for the assessment of downstream processing techniques such as gas stripping (307–309). Such techniques can allow for higher titres by minimising the buildup of toxic products (336) and can maintain cells in acetogenesis or solventogenesis, the latter being preferable for acetone production.

Of concern is the stability of generated integrant mutants. It is common (337) for industrial strains to suffer from genetic drift. This may result in decreased product titres and require inoculation with a cell line closer to that of the original. The long-term genetic instability of *E. limosum* is not known as it is not currently used industrially. Research would, therefore, need to be done in this area, with continuous fermentations performed for many thousands of hours. If stability is good, then it bolsters the organism's appeal as an industrial fermenter. If not, then it is likely a serious detraction, and may limit fermentation to batch or shorter continuous fermentations only. Given the efficacy of adaptive fermentation and the seeming plasticity of the genome (Chapter 3) it is likely that *E. limosum* would not be stable. However, given the significant genomic rearrangement which occurred without any clearly observed phenotype this may not in fact be the case.

If acetone production, or any plasmid-based product formation, were to be pursued in *E. limosum*, the stability of the p19 plasmid may offer the advantage of higher copy number and theoretically higher product concentrations in the absence of antibiotic selection. Alternatively, the use of a histidine auxotroph allows for the possibility of

maintaining plasmids (304,305) or using a toxin-antitoxin system for plasmid maintenance (306,307). Development of the plasmids and codon optimisation of the acetone production operon may prove possible and may improve yields (338). However, the data presented here cannot definitively show that the failure of the codon-optimised acetone production genes contained on plasmids was due to the codon-optimised operon itself. Determining this would require the integration of the codon optimised acetone production genes and expression of said operon from the successful plasmid system.

Promoters are likely to play a part in optimisation, and the high-strength promoters used as part of this work may have been detrimental due to the high expression of non-rate limiting proteins. An improved operon could be designed that could make use of multiple promoters (with terminators between genes to prevent read through) to reduce excess expression of non-limiting enzymes and to maximize production of rate-limiting proteins. This approach is not often considered but given the very limited free energy of acetogenic bacteria (75,121), even small energy savings can be essential for viability. It should be noted that this expression optimisation would be a significant piece of work, on par with or greater than the totality of this study. Such work would have value for other products, as acetone is likely not a worthwhile industrial use for *E. limosum*. Further the removal of genes for which *E. limosum* has homologues for and instead focusing on the *ctfa/b* genes may aid titres by reducing the size of plasmids and limiting the high strength promoter to only synthesising genes which are missing.

5.5 Chapter conclusions

The work presented in this chapter is a proof of concept for the tools now available in *E. limosum*. However, it is likely not indicative of the direction that further work with the organism should proceed. Acetone is already produced industrially using *C. acetobutylicum*, and there is little advantage in trying to compete with this well-established and developed organism for such a low-value product. However, this work does show that large pathways can be inserted and can function in *E. limosum* with the current toolset available. Future work should focus on the selection of a viable product and increasing knowledge of *E. limosum,* with an emphasis on issues that are of interest for industrial fermentation, such as scale-up, mutant stability, product titres and downstream processing. Ideally the product chosen should make the best use of the advantages that *E. limosum* possesses rather than in trying to use what works well elsewhere.

Chapter 6
Conclusions and outlook

This work has employed adaptive fermentation to generate a novel adapted strain of *E. limosum* ATCC 8486 and has attempted to characterise the genomic changes accumulated over a period of bioreactor growth. Further, many different tools have been used to create mutants in *E. limosum*, with mutants confirming or dispelling various hypotheses surrounding genes and their purported functions. Problems with tools were identified and overcome, and a roadmap for development of this was organism created. An enzymatic pathway which draws upon acetyl-CoA produced by the Wood-Ljungdahl pathway, allowing for acetone production, was genomically integrated into *E. limosum* and functioned with detectable levels of acetone produced. This work is a proof of concept as regarding *E. limosum*. Essential groundwork has been laid, and research is in an ideal position for further development towards commercial production from feedstocks such as CO/CO_2 and methanol/CO_2.

There is a good outlook for *E. limosum*, and there is much scope for future work. The development of the toxin-antitoxin system (227) ensures that future work in *E. limosum* will have a reliable, rapid, accurate, and powerful tool for both insertion and deletion of genes, with older technologies like ACE able to insert operons with very high effective efficiency. Key to future work is the selection of a value-added chemical that is viable to produce in sufficient quantities for a financially tenable industrial process to be developed. The biggest hurdle now for *E. limosum* is increasing product yields and scaling up to industrial fermentation. This will require much in the way of process development, which would need to be coordinated in line with the desired product.

Redox balance, the balance reduction and oxidation reactions within an organism, is important in all organisms for cellular homeostasis (339) and so is an important consideration in metabolic engineering. A more comprehensive understanding of redox balance in *E. limosum* would be beneficial, especially as it relates to the Wood-Ljungdahl pathway and methanol/CO/CO_2 usage (340) as redox reactions are essential for energy production. Energy conservation is of paramount importance in acetogens, given the limited amount of free energy (341). This limits products and may limit titres either by limiting cell mass or by being a limiting factor at an ATP-dependent reaction prior to that of the desired product. The literature (341) suggests

several routes to address energy conservation issues in acetogens, and these could be investigated in *E. limosum*; these would vary depending on the product of interest. If the product was ethanol, a native product of *E. limosum*, and an H_2/CO_2 feedstock were used, it may be possible to improve titres by the identification or introduction (if not natively present) of an aldehyde ferredoxin oxidoreductase. This may result in the product (overall) yielding ATP instead of costing ATP, though ferredoxin is required (341). Energy conservation is a direction of research that is likely to yield iterative improvements to an industrial fermentation. This may result in *E. limosum* becoming commercially viable for traditional acetogen products and may allow chemicals of interest to BioMetChem collaborators, such as 1,4-butanediol (1.8), to be produced in *E. limosum*.

Of note, however, is the native production of vitamin B12 in *E. limosum* (312). Vitamin B12 is essential for most life on Earth and cannot be natively produced by most organisms. Vitamin B12 production may be more profitable than acetone production and may be of greater commercial interest, as other organisms such as *C. acetobutylicum* are already well-established producers of acetone. If vitamin B12 was the product of interest, then process development would need to consider the cobalt complementation necessary for B12 production (342). If cobalt toxicity was an issue the toxin-antitoxin system could be used to increase the tolerance of cells to cobalt, if known genes or mechanisms are involved in tolerance or susceptibility to cobalt (343,344). If no known compatible genes are known to aid in cobalt tolerance, then adaptive fermentation, ideally of a TraDIS library, could be performed in a similar fashion to that done in this study with CO, with increasing cobalt concentrating being the selection pressure. Either of these approaches may lead to increased yields of vitamin B12. However, these processes would need to compete with the currently existing industrial exploitation of *Propionibacterium shermanii* (345) and *Pseudomonas denitrificans* (346), an anaerobe and aerobe respectively. While neither of the existing processes can be performed while transforming methanol, it is unclear whether *E. limosum* could produce sufficient B12 to be competitive. Further, as new technologies emerge, the existing industry is making full use of them, with *Pseudomonas denitrificans* being genetically engineered to improve vitamin B12 yields (347). It is, however, worth investigating as the current biosynthesised vitamin

B12 industry is worth over $200 million, with predictions of a doubling of value by 2030 (348). Being able to compete (even marginally) in this industry could finance further work into the development of *E. limosum*. Either further engineering *E. limosum* to enhance vitamin B12 production or co-production with another high-value chemical is likely essential for the financial viability and long-term usefulness of *E. limosum* as compared to other acetogenic bacteria. Of some concern for acetogen researchers, work has been done in aerobic bacteria to try to introduce genes to allow for robust growth on methanol, with *E. coli* being the current leader in this regard (349). While growth is relatively comparable to *E. limosum*, the well-developed nature of *E. coli* in terms of molecular tools and industrial usage may present significant competition to acetogens. Methylotrophy is also present in some species of yeast natively (350). In addition, engineered strains have also been developed which are able to grow on methanol and include *Pichia pastoris* (351) and *Saccharomyces cerevisiae* (351,352). Yeasts are especially interesting from the point of view of industrial fermentation given the extensive availability of fermentation equipment and expertise in the fermentation of yeast, primarily from the brewing industry. However, aerobic organisms are unable to natively capture all of the carbon, with loss in the form of CO_2 (353), which decreases the conversion efficiency of methylotrophic aerobes relative to their anaerobic counterparts. This is unlike acetogens which are able to fully capture CO_2. This landscape is constantly changing, and the research literature needs to be reevaluated constantly to ensure that work is directed in a valuable direction.

Laboratory evolution using long-term fermentation experiments, perhaps making use of a transposon mutant pool to create a population of mutants to select from, could result in a strain with favourable growth characteristics or tolerance towards products of interest. The simplest way to improve yields is to increase biomass, with the same reactor volume and more cells, the higher the potential productivity. It is worth noting however that toxic product concentrations are limiting and without active removal of product may limit growth and decrease biomass (307). Regarding fermentation, it may be possible to co-culture *E. limosum* either with different strains of *E. limosum* or with another organism that either produces chemicals useful for *E. limosum* or *E.*

limosum produces chemicals utilised by the said organism (354), which may make the process either more efficient or more financially viable.

Process development is key to any industrial process, and fermentation is no different. Process optimisation can take many forms, from the use of an optimised medium (355), to efficient downstream processing (307,308), to improved fermenter technologies and operation efficiency optimisations. The importance of this cannot be overstated, many proofs of concepts exist as lab scale but sadly many remain at said scale due to the difficulty in scaling to viable industrial process (356). Collaboration with engineers, existing bioprocesses (such as the brewing industry) and scientists is likely essential for processes to see commercial success. This commercial success is necessary to supplant fossil fuel-based processes as commercial viability is a measure of the reality of the process. If for instance a commercial biofuel producing process were to require more fossil fuels per litre of produced fuel it may not be a process worth perusing from a climate change mitigation perspective (357). Indeed, sustainability and the concerns of the environment should be put centre stage within the bioprocess industry to ensure that a future civilizational model is not built on poor foundations which will similarly cause a crisis akin to that we face currently with the climate.

The molecular biology landscape more generally seems strong. In the last 20 years, technology has advanced significantly. The development of powerful genetic modification techniques like CRISPR-Cas (218) and toxin-antitoxin-based (227) counter-selection systems not only allow for the rapid production of clean and stable mutants but, when combined with technologies like TraDIS (224,358), could allow for the development of high-throughput workflows (359) where the phenotypical effects of many non-essential genes could be understood in a way which was impossible before. Developments in sequencing are also useful in understanding organisms, with epigenetics coming to the fore in many organisms as it becomes apparent that the regulation of genes, even in prokaryotes, is complex beyond previously understood (304,360,361). Automation technologies (362) and machine learning (363) promise to be powerful tools that can further the field of biotechnology generally which feeds into the C1 field. Genome-scale models (364), combined with

the aforementioned developments, promise the ability to direct research direction in a more rational way than ever before and could pave the way for working environments that allow for more effective research collaborations with researchers from all over the world. There is significant interest in the field not just for the potential to aid with climate change and sustainability (365) but also in understanding (366) and preventing disease (367) and the production of complex pharmaceutical chemicals (368,369) that are otherwise difficult to produce.

Regarding C1 feedstocks in general, the increased interest in alternatives to petrochemicals and the desire for carbon capture means that it is very likely to remain of tremendous interest and growth in the years to come. C1-utilising organisms that make use of CO_2, especially, will be of interest as storage of captured CO_2 from direct air carbon capture (370) remains an issue if significant amounts are to be removed and sequestered. Incorporation into non-degraded products may be a solution (371), and C1-utilising organisms could be used to that end.

E. limosum is poised to be part of the C1 solution, with the main technical hurdles overcome as regards genetic modification, allowing the organism to be developed for either lone use in industrial fermentation, as part of a co-culture with another organism (perhaps one that requires vitamin B12 supplementation), or as part of an integrated facility where waste products and produced products from one organism are used to feed or supplement another, allowing for CO_2, methanol or CO to be converted, stepwise, into progressively higher value products.

7. References

1. Sangeeta, Moka S, Pande M, Rani M, Gakhar R, Sharma M, et al. Alternative fuels: An overview of current trends and scope for future. Renew Sustain Energy Rev. 2014 Apr 1;32:697–712.
2. Abas N, Kalair A, Khan N. Review of fossil fuels and future energy technologies. Futures. 2015 May 1;69:31–49.
3. Noguera-Santaella J. Geopolitics and the oil price. Econ Model. 2016 Jan 1;52:301–9.
4. Oelkers EH, Cole DR. Carbon Dioxide Sequestration A Solution to a Global Problem. Elements. 2008 Oct 1;4(5):305–10.
5. United Nations. Convention on Climate Change: Climate Agreement of Paris. 2015;1–25.
6. Anderson JO, Thundiyil JG, Stolbach A. Clearing the air: a review of the effects of particulate matter air pollution on human health. J Med Toxicol. 2012 Jun;8(2):166–75.
7. Burnett R, Chen H, Szyszkowicz M, Fann N, Hubbell B, Pope CA, et al. Global estimates of mortality associated with longterm exposure to outdoor fine particulate matter. Proc Natl Acad Sci U S A. 2018 Sep 18;115(38):9592–7.
8. Rogelj J, Shindell D, Jiang K, Fifita S, Forster P, Ginzburg V, et al. Mitigation Pathways Compatible With 1.5°C in the Context of Sustainable Development. Glob Warm 15°C An IPCC Spec Rep [.]. 2018;82pp.
9. Handler RM, Shonnard DR, Griffing EM, Lai A, Palou-Rivera I. Life Cycle Assessments of Ethanol Production via Gas Fermentation: Anticipated Greenhouse Gas Emissions for Cellulosic and Waste Gas Feedstocks. Ind Eng Chem Res. 2016 Mar 30;55(12):3253–61.
10. Raza A, Gholami R, Rezaee R, Bing CH, Nagarajan R, Hamid MA. Well selection in depleted oil and gas fields for a safe CO2 storage practice: A case study from Malaysia. Petroleum. 2017 Mar 1;3(1):167–77.
11. Loisel J, Connors JPC, Hugelius G, Harden JW, Morgan CL. Soils can helpmitigate CO2 emissions, despite the challenges. Vol. 116, Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences; 2019. p. 10211–2.

12. Paustian K, Larson E, Kent J, Marx E, Swan A. Soil C Sequestration as a Biological Negative Emission Strategy. Front Clim. 2019 Oct 16;1:8.
13. Bastin J-F, Finegold Y, Garcia C, Mollicone D, Rezende M, Routh D, et al. The global tree restoration potential. Science (80-). 2019 Jul 5;365(6448):76–9.
14. Seetharaman, Moorthy K, Patwa N, Saravanan, Gupta Y. Breaking barriers in deployment of renewable energy. Heliyon. 2019 Jan;5(1):e01166.
15. Forsberg CW. Sustainability by combining nuclear, fossil, and renewable energy sources. Prog Nucl Energy. 2009 Jan 1;51(1):192–200.
16. Agusdinata DB, Liu W, Eakin H, Romero H. Socio-environmental impacts of lithium mineral extraction: towards a research agenda. Environ Res Lett. 2018 Nov 27;13(12):123001.
17. Goldemberg J. The Brazilian biofuels industry. Biotechnol Biofuels. 2008 May 1;1(1):6.
18. Goldemberg J. The ethanol program in Brazil. Vol. 1, Environmental Research Letters. IOP Publishing; 2006. p. 014008.
19. Hassan MH, Kalam MA. An overview of biofuel as a renewable energy source: Development and challenges. In: Procedia Engineering. 2013. p. 39–53.
20. Kesieme U, Pazouki K, Murphy A, Chrysanthou A. Biofuel as an alternative shipping fuel: technological, environmental and economic assessment. Sustain Energy Fuels. 2019 Mar 26;3(4):899–909.
21. Azami MH, Savill M. Comparative study of alternative biofuels on aircraft engine performance. Proc Inst Mech Eng Part G J Aerosp Eng. 2017 Jun 22;231(8):1509–21.
22. Rahmes T, Kinder J, Crenfeldt G, LeDuc G, Abe Y, McCall M, et al. Sustainable Bio-Derived Synthetic Paraffinic Kerosene (Bio-SPK) Jet Fuel Flights and Engine Tests Program Results. In: 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO). Reston, Virigina: American Institute of Aeronautics and Astronautics; 2009.
23. Wu L, Moteki T, Gokhale AA, Flaherty DW, Toste FD. Production of Fuels and Chemicals from Biomass: Condensation Reactions and Beyond. Chem. 2016 Jul 7;1(1):32–58.
24. Jones DT, Woods DR. Acetone-butanol fermentation revisited. Microbiol Rev.

1986 Dec;50(4):484–524.
25. Luyben WL. Design and Control of the Cumene Process. Ind Eng Chem Res. 2010 Jan 20;2(2):719–34.
26. Reiner JM. METABOLIC INTERMEDIATES IN ADAPTIVE FERMENTATION OF GALACTOSE BY YEAST. J Gen Physiol. 1947 Mar 20;30(4):355–65.
27. Borisut P, Nuchitprasittichai A. Methanol Production via CO2 Hydrogenation: Sensitivity Analysis and Simulation—Based Optimization. Front Energy Res. 2019 Sep 6;7:81.
28. Marlin DS, Sarron E, Sigurbjörnsson Ó. Process Advantages of Direct CO2 to Methanol Synthesis. Front Chem. 2018 Sep 27;6:446.
29. Shapouri H, Salassi M. The economic feasibility of ethanol production from sugar in the United States. USDA Rep. 2006;(July):78.
30. McNeil B, Kristiansen B. The Acetone Butanol Fermentation. Adv Appl Microbiol. 1986;31(C):61–92.
31. Eisentraut A. Sustainable Production of Second generation Biofules. IEA Energy Pap. 2010;1–39.
32. Mimura N. Sea-level rise caused by climate change and its implications for society. Proc Jpn Acad Ser B Phys Biol Sci. 2013 Jul 7;89(7):281.
33. Fan M, Shen J, Yuan L, Jiang R, Chen X, Davies WJ, et al. Improving crop productivity and resource use efficiency to ensure food security and environmental quality in China. J Exp Bot. 2012;63(1):13–24.
34. Schütz L, Gattinger A, Meier M, Müller A, Boller T, Mäder P, et al. Improving Crop Yield and Nutrient Use Efficiency via Biofertilization—A Global Meta-analysis. Front Plant Sci. 2018;8(January).
35. Fageria NK, Baligar VC, Li YC. The role of nutrient efficient plants in improving crop yields in the twenty first century. Vol. 31, Journal of Plant Nutrition. Taylor & Francis Group; 2008. p. 1121–57.
36. Key S, Ma JKC, Drake PMW. Genetically modified plants and human health. J R Soc Med. 2008;101(6):290–8.
37. López-Arredondo D, González-Morales SI, Bello-Bello E, Alejo-Jacuinde G, Herrera L. Engineering food crops to grow in harsh environments. F1000Research. 2015;4.

38. Adriansen HK. Land reclamation in Egypt: A study of life in the new lands. Geoforum. 2009;40(4):664–74.
39. Sabo JL, Sinha T, Bowling LC, Schoups GHW, Wallender WW, Campana ME, et al. Reclaiming freshwater sustainability in the Cadillac Desert. Proc Natl Acad Sci. 2010;107(50):21263–9.
40. Mohammadi M, Najafpour GD, Younesi H, Lahijani P, Uzir MH, Mohamed AR. Bioconversion of synthesis gas to second generation biofuels: A review. Renew Sustain Energy Rev. 2011;15(9):4255–73.
41. Pace G, Sheehan SW. Scaling CO2 Capture With Downstream Flow CO2 Conversion to Ethanol. Front Clim. 2021 May 6;3:35.
42. Middleton RS, Bielicki JM. A scalable infrastructure model for carbon capture and storage: SimCCS. Energy Policy. 2009 Mar 1;37(3):1052–60.
43. Gabrielli P, Gazzani M, Mazzotti M. The Role of Carbon Capture and Utilization, Carbon Capture and Storage, and Biomass to Enable a Net-Zero-CO2 Emissions Chemical Industry. Ind Eng Chem Res. 2020 Apr 15;59(15):7033–45.
44. Galadima A, Muraza O. From synthesis gas production to methanol synthesis and potential upgrade to gasoline range hydrocarbons: A review. J Nat Gas Sci Eng. 2015 Jul;25:303–16.
45. Cañete B, Gigola CE, Brignole NB. Synthesis Gas Processes for Methanol Production via CH 4 Reforming with CO 2 , H 2 O, and O 2. Ind Eng Chem Res. 2014;53(17):7103–12.
46. Lundgren J, Asp B, Larsson M, Grip C-E. Methanol Production at an Integrated steel mill. Proc 18th Int Congr Chem Process Eng. 2008;11.
47. Wu H, Liu Y, Ji D, Li Z, Yi G, Yuan D, et al. Renewable and high efficient syngas production from carbon dioxide and water through solar energy assisted electrolysis in eutectic molten salts. J Power Sources. 2017 Sep 15;362:92–104.
48. Phillips JR, Huhnke RL, Atiyeh HK, Phillips JR, Huhnke RL, Atiyeh HK. Syngas Fermentation: A Microbial Conversion Process of Gaseous Substrates to Various Products. Fermentation. 2017 Jun 16;3(2):28.
49. Lamy-Pitara E, Bencharif L, Barbier J. Effect of sulphur on the properties of platinum catalysts as characterized by cyclic voltammetry. Appl Catal. 1985 Sep 1;18(1):117–31.

50. Gandhi HS, Shelef M. Effects of sulphur on noble metal automotive catalysts. Appl Catal. 1991 Oct 15;77(2):175–86.
51. Klasson KT, Ackerson CMD, Clausen EC, Gaddy JL. Biological concersion of synthesis gas into fuels. Int J Hydrogen Energy. 1992;17(4):281–8.
52. Tremblay P-L, Höglund D, Koza A, Bonde I, Zhang T. Adaptation of the autotrophic acetogen Sporomusa ovata to methanol accelerates the conversion of CO2 to organic products. Sci Rep. 2015 Dec 4;5(1):16168.
53. Comer AD, Long MR, Reed JL, Pfleger BF. Flux balance analysis indicates that methane is the lowest cost feedstock for microbial cell factories. Metab Eng Commun. 2017 May;5:26–33.
54. David WIF. Effective hydrogen storage: a strategic chemistry challenge. Faraday Discuss. 2011;151:399.
55. Daniell J, Koepke M, Simpson SD. Commercial Biomass Syngas Fermentation. Energies. 2012 Dec;5(12):5372–417.
56. Olah GA, Prakash GKS, Goeppert A. Anthropogenic Chemical Carbon Cycle for a Sustainable Future. J Am Chem Soc. 2011 Aug 24;133(33):12881–98.
57. Liu W-C, Baek J, Somorjai GA. The Methanol Economy: Methane and Carbon Dioxide Conversion. Top Catal. 2018 Mar 8;1–12.
58. Olah GA. Beyond Oil and Gas : The Methanol Economy. Angew Chemie. 2005;44:2636–9.
59. Hill J, Nelson E, Tilman D, Polasky S, Tiffany D. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proc Natl Acad Sci. 2006;103(30):11206–10.
60. Galindo Cifre P, Badr O. Renewable hydrogen utilisation for the production of methanol. Energy Convers Manag. 2007;48(2):519–27.
61. Parks G, Boyd R, Cornish J, Remick R. Hydrogen Station Compression, Storage, and Dispensing Technical Status and Costs: Systems Integration. 2014;(May).
62. Becherif M, Ramadan HS, Cabaret K, Picard F, Simoncini N, Bethoux O. Hydrogen Energy Storage: New Techno-Economic Emergence Solution Analysis. Energy Procedia. 2015;74(0):371–80.
63. Bromberg L, Cheng WK. Methanol as an alternative transportation fuel in the U.S.: Options for sustainable and energy-secure transportation. Sloan Automot

Lab Massachusetts Inst Technol. 2010;(4000096701):81.
64. Murphy JD, McCarthy K. Ethanol production from energy crops and wastes for use as a transport fuel in Ireland. Appl Energy. 2005;82(2):148–66.
65. Torli M, Geer L, Kontogeorgis GM, Fosbøl PL. Solubility of Syngas Components in Water, Acetic Acid, and Alcohol Using New Standard Fugacity Methodology. Ind Eng Chem Res. 2018 Dec 12;57(49):16958–77.
66. Richter H, Martin M, Angenent L, Richter H, Martin ME, Angenent LT. A Two-Stage Continuous Fermentation System for Conversion of Syngas into Ethanol. Energies. 2013 Aug 7;6(8):3987–4000.
67. Yasin M, Park S, Jeong Y, Lee EY, Lee J, Chang IS. Effect of internal pressure and gas/liquid interface area on the CO mass transfer coefficient using hollow fibre membranes as a high mass transfer gas diffusing system for microbial syngas fermentation. Bioresour Technol. 2014 Oct 1;169:637–43.
68. Tappe W, Tomaschewski C, Rittershaus S, Groeneweg J. Cultivation of nitrifying bacteria in the retentostat, a simple fermenter with internal biomass retention. FEMS Microbiol Ecol. 1996 Jan 17;19(1):47–52.
69. van Verseveld HW, Arbige M, R. Chesbro W. Continuous culture of bacteria with biomass retention. Vol. 2, Trends in Biotechnology. Elsevier Current Trends; 1984. p. 8–12.
70. Pollice A, Laera G, Blonda M. Biomass growth and activity in a membrane bioreactor with complete sludge retention. Water Res. 2004 Apr 1;38(7):1799–808.
71. Lehmann D, Hönicke D, Ehrenreich A, Schmidt M, Weuster-Botz D, Bahl H, et al. Modifying the product pattern of Clostridium acetobutylicum. Appl Microbiol Biotechnol. 2012 May 14;94(3):743–54.
72. Wilding-Steele T, Ramette Q, Jacottin P, Soucaille P. Improved CRISPR/Cas9 Tools for the Rapid Metabolic Engineering of Clostridium acetobutylicum. Int J Mol Sci. 2021 Apr 1;22(7).
73. Drake HL, Gößner AS, Daniel SL, Goessner AS, Daniel SL, Gößner AS, et al. Old acetogens, new light. Wiegel RJ Adams,MWW JM, editor. Ann N Y Acad Sci. 2008 Mar 1;1125(1):100–28.
74. Schiel-Bengelsdorf B, Duerre P. Pathway engineering and synthetic biology using acetogens. FEBS Lett. 2012 Jul 16;586(15):2191–8.

75. Schuchmann K, Müller V. Energetics and Application of Heterotrophy in Acetogenic Bacteria. Stams AJM, editor. Appl Environ Microbiol. 2016 Jul 15;82(14):4056–69.
76. Berg IA. Ecological aspects of the distribution of different autotrophic CO2 fixation pathways. Appl Environ Microbiol. 2011;77(6):1925–36.
77. Ragsdale SW, Pierce E. Acetogenesis and the Wood-Ljungdahl Pathway of CO2 Fixation. Biochim Biophys Acta. 2008 Dec;1784(12):1873.
78. Deamer DW. The first living systems: a bioenergetic perspective. Microbiol Mol Biol Rev. 1997;61(2):239–61.
79. Kaufmann M. On the free energy that drove primordial anabolism. Int J Mol Sci. 2009;10(4):1853–71.
80. Wieringa KT. Over het verdwijnen van waterstof en koolzuur onder anaerobe voorwaarden - Voortzetting van een onderzoek van wijlen Prof. Dr. Ir. N. L. S??hngen. Antonie Van Leeuwenhoek. 1936 Dec;3(4):263–73.
81. Schuchmann K, Müller V. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. Nat Rev Microbiol. 2014 Dec 10;12(12):809–21.
82. Min F, Kopke M, Dennis S. Gas Fermentation for Commercial Biofuels Production. In: Liquid, Gaseous and Solid Biofuels - Conversion Techniques. InTech; 2013.
83. van der Werf W. Advanced Fuels & Chemicals from Waste: Waste in a Circular Economy. AMF + IEA Bioenergy Jt Work Infrastruct Compat Transp Fuels, May 20th 2014, Copenhagen Denmark. 2014;
84. Drake HL, Daniel SL. Physiology of the thermophilic acetogen Moorella thermoacetica. Res Microbiol. 2004 Dec;155(10):869–83.
85. Pierce E, Xie G, Barabote RD, Saunders E, Han CS, Detter JC, et al. The complete genome sequence of Moorella thermoacetica (f. Clostridium thermoaceticum). Environ Microbiol. 2008 Oct;10(10):2550–73.
86. Islam MA, Zengler K, Edwards EA, Mahadevan R, Stephanopoulos G. Investigating Moorella thermoacetica metabolism with a genome-scale constraint-based metabolic model. Integr Biol (Camb). 2015 Aug;7(8):869–82.
87. Iwasaki Y, Kita A, Sakai S, Takaoka K, Yano S, Tajima T, et al. Engineering of a functional thermostable kanamycin resistance marker for use in *Moorella*

thermoacetica ATCC39073. FEMS Microbiol Lett. 2013 Jun 1;343(1):8–12.
88. Byrer DE, Rainey FA, Wiegel J. Novel strains of Moorella thermoacetica form unusually heat-resistant spores. Arch Microbiol. 2000 Nov 20;174(5):334–9.
89. Iwasaki Y, Kita A, Yoshida K, Tajima T, Yano S, Shou T, et al. Homolactic Acid Fermentation by the Genetically Engineered Thermophilic Homoacetogen Moorella thermoacetica ATCC 39073. Appl Environ Microbiol. 2017 Apr 15;83(8):e00247-17.
90. KARITA S, NAKAYAMA K, GOTO M, SAKKA K, KIM W-J, OGAWA S. A Novel Cellulolytic, Anaerobic, and Thermophilic Bacterium, *Moorella* sp. Strain F21. Biosci Biotechnol Biochem. 2003 Jan 22;67(1):183–5.
91. KITA A, IWASAKI Y, YANO S, NAKASHIMADA Y, HOSHINO T, MURAKAMI K. Isolation of Thermophilic Acetogens and Transformation of Them with the pyrF and kanr Genes. Biosci Biotechnol Biochem. 2013;77(2):301–6.
92. Liew F, Henstra AM, Köpke M, Winzer K, Simpson SD, Minton NP. Metabolic engineering of Clostridium autoethanogenum for selective alcohol production. Metab Eng. 2017 Mar 1;40(May 2016):104–14.
93. Nagaraju S, Davies NK, Walker DJF, Köpke M, Simpson SD. Genome editing of Clostridium autoethanogenum using CRISPR/Cas9. Biotechnol Biofuels. 2016 Dec 18;9(1):219.
94. Liou JS-C, Balkwill DL, Drake GR, Tanner RS. Clostridium carboxidivorans sp. nov., a solvent-producing clostridium isolated from an agricultural settling lagoon, and reclassification of the acetogen Clostridium scatologenes strain SL1 as Clostridium drakei sp. nov. Int J Syst Evol Microbiol. 2005 Sep 1;55(5):2085–91.
95. Phillips JR, Atiyeh HK, Tanner RS, Torres JR, Saxena J, Wilkins MR, et al. Butanol and hexanol production in Clostridium carboxidivorans syngas fermentation: Medium development and culture techniques. Bioresour Technol. 2015 Aug 1;190:114–21.
96. Bruant G, Lévesque M-J, Peter C, Guiot SR, Masson L. Genomic analysis of carbon monoxide utilization and butanol production by Clostridium carboxidivorans strain P7. PLoS One. 2010 Sep 27;5(9):e13033.
97. Hagemeier CH, Krer M, Thauer RK, Warkentin E, Ermler U. Insight into the

mechanism of biological methanol activation based on the crystal structure of the methanol-cobalamin methyltransferase complex. Proc Natl Acad Sci U S A. 2006 Dec 12;103(50):18917–22.

98. Ragsdale SW, Pierce E. Acetogenesis and the Wood-Ljungdahl pathway of CO(2) fixation. Biochim Biophys Acta. 2008 Dec;1784(12):1873–98.
99. Bogdanović N, Trifunović D, Sielaff H, Westphal L, Bhushan S, Müller V, et al. The structural features of *Acetobacterium woodii* F- ATP synthase reveal the importance of the unique subunit γ-loop in Na^+ translocation and ATP synthesis. FEBS J. 2019 May 12;286(10):1894–907.
100. Schoelmerich MC, Katsyv A, Sung W, Mijic V, Wiechmann A, Kottenhahn P, et al. Regulation of lactate metabolism in the acetogenic bacterium *Acetobacterium woodii*. Environ Microbiol. 2018 Dec 24;20(12):4587–95.
101. Dönig J, Müller V. Alanine, a Novel Growth Substrate for the Acetogenic Bacterium Acetobacterium woodii. Appl Environ Microbiol. 2018 Dec 1;84(23):e02023-18.
102. Lechtenfeld M, Heine J, Sameith J, Kremp F, Müller V. Glycine betaine metabolism in the acetogenic bacterium *Acetobacterium woodii*. Environ Microbiol. 2018 Dec 5;20(12):4512–25.
103. Poehlein A, Schmidt S, Kaster AK, Goenrich M, Vollmers J, Thürmer A, et al. An ancient pathway combining carbon dioxide fixation with the generation and utilization of a sodium ion gradient for ATP synthesis. Driessen A, editor. PLoS One. 2012 Mar 29;7(3):e33439.
104. Straub M, Demler M, Weuster-Botz D, Dürre P. Selective enhancement of autotrophic acetate production with genetically modified Acetobacterium woodii. J Biotechnol. 2014 May 20;178:67–72.
105. Kremp F, Poehlein A, Daniel R, Müller V. Methanol metabolism in the acetogenic bacterium *Acetobacterium woodii*. Environ Microbiol. 2018 Dec 24;20(12):4369–84.
106. Kremp F, Poehlein A, Daniel R, Müller V. Methanol metabolism in the acetogenic bacterium Acetobacterium woodii. Environ Microbiol. 2018 Dec 24;20(12):4369–84.
107. Chang IS, Kim BH, Kim DH, Lovitt RW, Sung HC. Formulation of defined media for carbon monoxide fermentation by Eubacterium limosum KIST612

and the growth characteristics of the bacterium. J Biosci Bioeng. 1999 Jan 1;88(6):682–5.

108. Song Y, Cho B-KK. Draft Genome Sequence of Chemolithoautotrophic Acetogenic Butanol-Producing Eubacterium limosum ATCC 8486. Genome Announc. 2015 Feb 12;3(1):e01564-14.
109. Shreiner AB, Kao JY, Young VB. The gut microbiome in health and in disease. Curr Opin Gastroenterol. 2015 Jan 12;31(1):69.
110. Limbana T, Khan F, Eskander N. Gut Microbiome and Depression: How Microbes Affect the Way We Think. Cureus. 2020 Aug 23;12(8).
111. Biagi E, Nylund L, Candela M, Ostan R, Bucci L, Pini E, et al. Through Ageing, and Beyond: Gut Microbiota and Inflammatory Status in Seniors and Centenarians. PLoS One. 2010;5(5):e10667.
112. Humphreys CM, Minton NP. Advances in metabolic engineering in the microbial production of fuels and chemicals from C1 gas. Curr Opin Biotechnol. 2018;50:174–81.
113. Drake HL, Daniel SL, Küsel K, Matthies C, Kuhner C, Braus-Stromeyer S. Acetogenic bacteria: what are the in situ consequences of their diverse metabolic versatilities? BioFactors. 1997;6(1):13–24.
114. Ragsdale SW. The eastern and western branches of the Wood/Ljungdahl pathway: how the east and west were won. Biofactors. 1997;6(1):3–11.
115. Borrel G, Adam PS, Gribaldo S. Methanogenesis and the Wood-Ljungdahl Pathway: An Ancient, Versatile, and Fragile Association. Genome Biol Evol. 2016;8(6):1706–11.
116. Drake HL, Küsel K, Matthies C. Ecological consequences of the phylogenetic and physiological diversities of acetogens. Antonie van Leeuwenhoek, Int J Gen Mol Microbiol. 2002 Aug;81(1–4):203–13.
117. Mansfield R. Developments In Genetic Engineering Of Novel Acetogens. PhD Thesis. 2018;University.
118. Orita I, Sato T, Yurimoto H, Kato N, Atomi H, Imanaka T, et al. The ribulose monophosphate pathway substitutes for the missing pentose phosphate pathway in the archaeon Thermococcus kodakaraensis. J Bacteriol. 2006;188(13):4698–704.
119. Clustoserdova L V., Lidstrom ME. Genetics of the serine cycle in

Methylobacterium extorquens AM1: Identification of sgaA and mtdA and sequences of sgaA, hprA, and mtdA. J Bacteriol. 1994;176(7):1957–68.

120. Bogorad IW, Chen C-T, Theisen MK, Wu T-Y, Schlenz AR, Lam AT, et al. Building carbon–carbon bonds using a biocatalytic methanol condensation cycle. Proc Natl Acad Sci. 2014;111(45):15928–33.
121. Müller V. Energy Conservation in Acetogenic Bacteria. Vol. 69, Applied and Environmental Microbiology. American Society for Microbiology; 2003. p. 6345–53.
122. Fleischmann RD, Adams MD, White O, Clayton RA, Ewen F, Kerlavage AR, et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd .(first genome of free-living organism ever completely sequenced)(includes fold-out genome map)(Cover Story) Whole-genome random sequencing and assembly of Haemophilus influenz. Science. 1995;496(17):1–16.
123. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26(10):1135–45.
124. Najafi A, Bidkhori G, Bozorgmehr JH, Koch I, Masoudi-Nejad A. Genome scale modeling in systems biology: algorithms and resources. Curr Genomics. 2014;15(2):130–59.
125. Cuevas DA, Edirisinghe J, Henry CS, Overbeek R, O'Connell TG, Edwards RA. From DNA to FBA: How to build your own genome-scale metabolic model. Front Microbiol. 2016;7(JUN):1–12.
126. Zhang H, Ye C, Xu N, Chen C, Chen X, Yuan F, et al. Reconstruction of a Genome-scale Metabolic Network of Komagataeibacter nataicola RZS01 for Cellulose Production. Sci Rep. 2017;7(1):7911.
127. Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30(9):418–26.
128. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics. 2016;107(1):1–8.
129. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci. 1977;74(12):5463–7.
130. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012.

131. Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333–51.
132. Liao YC, Lin SH, Lin HH. Completing bacterial genome assemblies: Strategy and performance comparisons. Sci Rep. 2015;5.
133. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 2012;19(5):455–77.
134. Kamath GM, Shomorony I, Xia F, Courtade TA, Tse DN. HINGE: Long-read assembly achieves optimal repeat resolution. Genome Res. 2017;27(5):747–56.
135. Treangen TJ, Abraham AL, Touchon M, Rocha EPC. Genesis, effects and fates of repeats in prokaryotic genomes. FEMS Microbiol Rev. 2009;33(3):539–71.
136. Miller JR, Koren S, Sutton G. Assembly algorithm for next-generation sequencing data. Genomics. 2010;95(6):315–27.
137. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 2016 171. 2016 Nov 25;17(1):1–11.
138. Prakash T, Taylor TD. Functional assignment of metagenomic data: Challenges and applications. Brief Bioinform. 2012;13(6):711–27.
139. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007;17(3):377–86.
140. Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. IMG: The integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 2012;40(D1):115–22.
141. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics. 2008;9:1–15.
142. Richardson EJ, Watson M. The automatic annotation of bacterial genomes. Brief Bioinform. 2013;14(1):1–12.
143. Chen L, Liu P, Evans TC, Ettwiller LM. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. Science (80-). 2017;355(6326):752–6.
144. Sleep JA, Schreiber AW, Baumann U. Sequencing error correction without a reference genome. BMC Bioinformatics. 2013;14.

145. Apweiler R, Martin MJ, O'Donovan C, Magrane M, Alam-Faruque Y, Antunes R, et al. Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Res. 2011;39(SUPPL. 1):214–9.
146. Köpke M, Mihalcea C, Liew FM, Tizard JH, Ali MS, Conolly JJ, et al. 2,3-butanediol production by acetogenic bacteria, an alternative route to chemical synthesis, using industrial waste gas. Appl Environ Microbiol. 2011 Aug 1;77(15):5467–75.
147. Cann AF, Liao JC. Pentanol isomer synthesis in engineered microorganisms. Appl Microbiol Biotechnol. 2010 Jan 27;85(4):893–9.
148. Devarajan Y, Nagappan BK, Munuswamy DB. Performance and emissions analysis on diesel engine fuelled with cashew nut shell biodiesel and pentanol blends. Korean J Chem Eng. 2017 Apr 22;34(4):1021–6.
149. Devarajan Y, Jayabal R kumar, Ragupathy D, Venu H. Emissions analysis on second generation biodiesel. Front Environ Sci Eng. 2017 Feb 24;11(1):3.
150. Ramakrishnan P, Kasimani R, Peer MS, Rajamohan S. Assessment of n-pentanol/Calophyllum inophyllum/diesel blends on the performance, emission, and combustion characteristics of a constant-speed variable compression ratio direct injection diesel engine. Environ Sci Pollut Res. 2018 Mar 5;1–14.
151. Esvelt KM, Wang HH. Genome-scale engineering for systems and synthetic biology. Mol Syst Biol. 2013;9(641):1–17.
152. Judson N, Mekalanos JJ. Transposon-based approaches to identify essential bacterial genes. Trends Microbiol. 2000 Nov 1;8(11):521–6.
153. Altenburg E. The Limit of Radiation Frequency Effective in Producing Mutations. Am Nat. 1928 Nov 15;62(683):540–5.
154. AUERBACH C. CHEMICAL MUTAGENESIS. Biol Rev. 1949 Jul 1;24(3):355–91.
155. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921.
156. Karberg M, Guo H, Zhong J, Coon R, Perutka J, Lambowitz AM. Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. Nat Biotechnol. 2001 Dec;19(12):1162–7.
157. Frazier CL, San Filippo J, Lambowitz AM, Mills DA. Genetic manipulation of

Lactococcus lactis by using targeted group II introns: generation of stable insertions without selection. Appl Environ Microbiol. 2003 Feb 1;69(2):1121–8.

158. Köpke M, Held C, Hujer S, Liesegang H, Wiezer A, Wollherr A, et al. Clostridium ljungdahlii represents a microbial production platform based on syngas. Proc Natl Acad Sci U S A. 2010 Jul 20;107(29):13087–92.
159. Minton NP, Ehsaan M, Humphreys CM, Little GT, Baker J, Henstra AM, et al. A roadmap for gene system development in Clostridium. Anaerobe. 2016;41:104–12.
160. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP. The ClosTron: A universal gene knock-out system for the genus Clostridium. J Microbiol Methods. 2007;70(3):452–64.
161. Kuehne SA, Heap JT, Cooksley CM, Cartman ST, Minton NP. ClosTron-mediated engineering of clostridium. Methods Mol Biol. 2011;765(August):389–407.
162. Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, Scott JC, et al. The ClosTron: Mutagenesis in Clostridium refined and streamlined. J Microbiol Methods. 2010;80(1):49–55.
163. Jia K, Zhu Y, Zhang Y, Li Y. Group II Intron-Anchored Gene Deletion in Clostridium. PLoS One. 2011;6(1):16693.
164. Heap JT, Ehsaan M, Cooksley CM, Ng Y-KK, Cartman ST, Winzer K, et al. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. Nucleic Acids Res. 2012 Apr 1;40(8):e59.
165. Ehsaan M, Kuit W, Zhang Y, Cartman ST, Heap JT, Winzer K, et al. Mutant generation by allelic exchange and genome resequencing of the biobutanol organism Clostridium acetobutylicum ATCC 824. Biotechnol Biofuels. 2016 Dec 4;9(1):4.
166. Ng YK, Ehsaan M, Philip S, Collery MM, Janoir C, Collignon A, et al. Expanding the Repertoire of Gene Tools for Precise Manipulation of the Clostridium difficile Genome: Allelic Exchange Using pyrE Alleles. Popoff MR, editor. PLoS One. 2013 Feb 6;8(2):e56051.
167. Patrick HD, Eric LS, Zhang F. Development and Applications of CRISPR-Cas9 for Genome Engineering Patrick. Cell. 2014;157(6):1262–78.

168. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science (80-). 2007 Mar 23;315(5819):1709–12.
169. Marraffini LA. CRISPR-Cas immunity in prokaryotes. Vol. 526, Nature. 2015. p. 55–61.
170. Barrangou R, Marraffini LA. CRISPR-cas systems: Prokaryotes upgrade to adaptive immunity. Vol. 54, Molecular Cell. 2014. p. 234–44.
171. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013 Jul 18;154(2):442–51.
172. Rath D, Amlinger L, Rath A, Lundgren M. The CRISPR-Cas immune system: Biology, mechanisms and applications. Biochimie. 2015.
173. Nidhi S, Anand U, Oleksak P, Tripathi P, Lal JA, Thomas G, et al. Novel CRISPR–Cas Systems: An Updated Review of the Current Achievements, Applications, and Future Research Perspectives. Int J Mol Sci. 2021 Apr 1;22(7).
174. Barrangou R. Diversity of CRISPR-Cas immune systems and molecular machines. Vol. 16, Genome Biology. BioMed Central Ltd.; 2015.
175. Koonin E V., Makarova KS. Origins and evolution of CRISPR-Cas systems. Vol. 374, Philosophical Transactions of the Royal Society B: Biological Sciences. Royal Society Publishing; 2019.
176. Nishimasu H, Nureki O. Structures and mechanisms of CRISPR RNA-guided effector nucleases. Vol. 43, Current Opinion in Structural Biology. Elsevier Ltd; 2017. p. 68–78.
177. Zheng Y, Li J, Wang B, Han J, Hao Y, Wang S, et al. Endogenous Type I CRISPR-Cas: From Foreign DNA Defense to Prokaryotic Engineering. Vol. 8, Frontiers in Bioengineering and Biotechnology. Frontiers Media S.A.; 2020.
178. Morisaka H, Yoshimi K, Okuzaki Y, Gee P, Kunihiro Y, Sonpho E, et al. CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. Nat Commun. 2019 Dec 1;10(1):1–13.
179. Cameron P, Coons MM, Klompe SE, Lied AM, Smith SC, Vidal B, et al. Harnessing type I CRISPR–Cas systems for genome engineering in human cells. Nat Biotechnol. 2019 Dec 1;37(12):1471–7.

180. Hidalgo-Cantabrana C, Goh YJ, Pan M, Sanozky-Dawes R, Barrangou R. Genome editing using the endogenous type I CRISPR-Cas system in Lactobacillus crispatus. Proc Natl Acad Sci U S A. 2019 Aug 6;116(32):15774–83.
181. McMahon SA, Zhu W, Graham S, Rambo R, White MF, Gloster TM. Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. Nat Commun. 2020 Dec 1;11(1):1–11.
182. Liu TY, Liu JJ, Aditham AJ, Nogales E, Doudna JA. Target preference of Type III-A CRISPR-Cas complexes at the transcription bubble. Nat Commun. 2019 Dec 1;10(1):1–13.
183. Estrella MA, Kuo FT, Bailey S. RNA-activated DNA cleavage by the Type III-B CRISPR–Cas effector complex. Genes Dev. 2016 Feb 15;30(4):460–70.
184. Özcan A, Pausch P, Linden A, Wulf A, Schühle K, Heider J, et al. Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. Nat Microbiol. 2019 Jan 1;4(1):89–96.
185. Cañadas IC, Groothuis D, Zygouropoulou M, Rodrigues R, Minton NP. RiboCas: A Universal CRISPR-Based Editing Tool for *Clostridium*. ACS Synth Biol. 2019 Jun 21;8(6):1379–90.
186. Cong L, Ran FA, Cox D, Lin S, Barretto R, Hsu PD, et al. Multiplex Genome Engineering Using CRISPR/VCas Systems. Science (80-). 2013;339(6121):819–23.
187. Kim H, Ham S, Jo M, Lee GH, Lee YS, Shin JH, et al. CRISPR-Cas9 mediated telomere removal leads to mitochondrial stress and protein aggregation. Int J Mol Sci. 2017;18(10).
188. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol. 2013;31(9):833–8.
189. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-Guided Human Genome Engineering via Cas9 Prashant. Science (80-). 2013;339(6121):823–6.
190. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol. 2015 Nov 1;13(11):722–36.

191. Yan WX, Hunnewell P, Alfonse LE, Carte JM, Keston-Smith E, Sothiselvam S, et al. Functionally diverse type V CRISPR-Cas systems. Science (80-). 2019 Jan 4;363(6422):88–91.
192. Harrington LB, Burstein D, Chen JS, Paez-Espino D, Ma E, Witte IP, et al. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. Science (80-). 2018 Nov 16;362(6416):839–42.
193. Chan VWK, Dreolini LF, Flintoff KA, Lloyd SJ, Mattenley AA. The Effect of Increasing Plasmid Size on Transformation Efficiency in Escherichia coli. 2002.
194. Yan WX, Chong S, Zhang H, Makarova KS, Koonin E V., Cheng DR, et al. Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein. Mol Cell. 2018 Apr 19;70(2):327-339.e5.
195. East-Seletsky A, O'Connell MR, Burstein D, Knott GJ, Doudna JA. RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes. Mol Cell. 2017 May 4;66(3):373-383.e3.
196. Abudayyeh OO, Gootenberg JS, Essletzbichler P, Han S, Joung J, Belanto JJ, et al. RNA targeting with CRISPR-Cas13. Nature. 2017 Oct 12;550(7675):280–4.
197. Rahman K, Jamal M, Chen X, Zhou W, Yang B, Zou Y, et al. Reprogramming the endogenous type III-A CRISPR-Cas system for genome editing, RNA interference and CRISPRi screening in Mycobacterium tuberculosis. bioRxiv. 2020 Mar 9;2020.03.09.983494.
198. Komor AC, Badran AH, Liu DR. CRISPR-based technologies for the manipulation of eukaryotic genomes. Cell. 2018 Jan 12;168(1–2):20–36.
199. Feng Z, Zhang B, Ding W, Liu X, Yang DL, Wei P, et al. Efficient genome editing in plants using a CRISPR/Cas system. Cell Res. 2013;23(10):1229–32.
200. Lu J, Tong Y, Pan J, Yang Y, Liu Q, Tan X, et al. A redesigned CRISPR/Cas9 system for marker-free genome editing in Plasmodium falciparum. Parasites and Vectors. 2016;9(1):1–8.
201. Brissett NC, Doherty AJ. Repairing DNA double-strand breaks by the prokaryotic non-homologous end-joining pathway. Biochem Soc Trans. 2009 Jun 1;37(3):539–45.

202. Song F, Stieger K. Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks. Mol Ther - Nucleic Acids. 2017;7(June):53–60.
203. Xie Z, Pang D, Wang K, Li M, Guo N, Yuan H, et al. Optimization of a CRISPR/Cas9-mediated Knock-in Strategy at the Porcine Rosa26 Locus in Porcine Foetal Fibroblasts. Sci Rep. 2017;7(1):1–12.
204. Shen CC, Hsu MN, Chang CW, Lin MW, Hwu JR, Tu Y, et al. Synthetic switch to minimize CRISPR off-target effects by self-restricting Cas9 transcription and translation. Nucleic Acids Res. 2019 Feb 20;47(3):e13.
205. Harms A, Maisonneuve E, Gerdes K. Mechanisms of bacterial persistence during stress and antibiotic exposure. Science (80-). 2016;354(6318).
206. Jahn LJ, Munck C, Ellabaan MMH, Sommer MOA. Adaptive laboratory evolution of antibiotic resistance using different selection regimes lead to similar phenotypes and genotypes. Front Microbiol. 2017;8(MAY):1–14.
207. Wang H, La Russa M, Qi LS. CRISPR/Cas9 in Genome Editing and Beyond. Annu Rev Biochem. 2016 Jun 2;85(1):227–64.
208. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012 Aug 17;337(6096):816–21.
209. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci. 2012 Sep 25;109(39):E2579–86.
210. Shen B, Zhang W, Zhang J, Zhou J, Wang J, Chen L, et al. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nat Methods. 2014 Apr 2;11(4):399–402.
211. Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 2014 Jan;24(1):132–41.
212. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, et al. Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. Cell. 2013 Sep 12;154(6):1380–9.
213. Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA. Programmable repression and activation of bacterial gene expression using an

engineered CRISPR-Cas system. Nucleic Acids Res. 2013 Aug;41(15):7429–37.

214. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013 Feb 28;152(5):1173–83.
215. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, et al. Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013;500(7463):472–6.
216. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods. 2013;10(10):973–6.
217. Kim S, Kim D, Cho SW, Kim J-SJ, Kim J-SJ. Highly Efficient RNA-guide genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. 2014;128:1–32.
218. Cañadas IC, Groothuis D, Zygouropoulou M, Rodrigues R, Minton NP. RiboCas: A Universal CRISPR-Based Editing Tool for Clostridium. ACS Synth Biol. 2019;
219. Selle K, Fletcher JR, Tuson H, Schmitt DS, McMillan L, Vridhambal GS, et al. In Vivo Targeting of Clostridioides difficile Using Phage-Delivered CRISPR-Cas3 Antimicrobials. MBio. 2020 Mar 1;11(2).
220. Baron U, Gossen M, Bujard H. Tetracycline-controlled transcription in eukaryotes: Novel transactivators with graded transactivation potential. Nucleic Acids Res. 1997 Jul 15;25(14):2723–9.
221. Freundlieb S, Schirra-Müller C, Bujard H. A Tetracycline Controlled Activation/Repression System with Increased Potential for Gene Transfer into Mammalian Cells. J Gene Med. 1999 Jan 1;1(1):4–12.
222. Gossen M, Freundlieb S, Bender G, Müller G, Hillen W, Bujard H. Transcriptional activation by tetracyclines in mammalian cells. Science (80-). 1995;268(5218):1766–9.
223. Li C-LF, Santhanam B, Webb AN, Zupan B, Shaulsky G. Gene discovery by chemical mutagenesis and whole-genome sequencing in Dictyostelium. Genome Res. 2016 Sep;26(9):1268–76.
224. Barquist L, Mayho M, Cummins C, Cain AK, Boinett CJ, Page AJ, et al. The

TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. Bioinformatics. 2016 Jan 21;32(7):1109–11.

225. Craig NL. Tn7: a target site-specific transposon. Mol Microbiol. 1991 Nov 1;5(11):2569–73.
226. Davies CJ, Hutchison III CA. Insertion site specificity of the transposon Tn3. Nucleic Acids Res. 1995 Feb 11;23(3):507–14.
227. Millard J, Agius A, Zhang Y, Soucaille P, Minton NP. Exploitation of a Type 1 Toxin-Antitoxin System as an Inducible Counter-Selective Marker for Genome Editing in the Acetogen Eubacterium limosum. Microorg 2023, Vol 11, Page 1256. 2023 May 10;11(5):1256.
228. Peltier J, Hamiot A, Garneau JR, Boudry P, Maikova A, Hajnsdorf E, et al. Type I toxin-antitoxin systems contribute to the maintenance of mobile genetic elements in Clostridioides difficile. Commun Biol. 2020 Dec 1;3(1):1–14.
229. Piscotta FJ, Jeffrey PD, James Link A. ParST is a widespread toxin–antitoxin module that targets nucleotide metabolism. Proc Natl Acad Sci U S A. 2019 Jan 15;116(3):826–34.
230. Guglielmini J, Van Melderen L. Bacterial toxin-antitoxin systems. Mob Genet Elements. 2011 Nov;1(4):283–306.
231. LeRoux M, Culviner PH, Liu YJ, Littlehale ML, Laub MT. Stress induces the transcription of toxin-antitoxin systems but does not activate toxin. bioRxiv. 2020 Mar 3;2020.03.02.972737.
232. Song S, Wood TK. A Primary Physiological Role of Toxin/Antitoxin Systems Is Phage Inhibition. Front Microbiol. 2020 Aug 13;11:1895.
233. Koga M, Otsuka Y, Lemire S, Yonesaki T. Escherichia coli rnlA and rnlB Compose a Novel Toxin–Antitoxin System. Genetics. 2011 Jan 1;187(1):123–30.
234. Rawlings DE. Proteic toxin-antitoxin, bacterial plasmid addiction systems and their evolution with special reference to the pas system of pTF-FC2. FEMS Microbiol Lett. 1999 Jul 1;176(2):269–77.
235. Kroll J, Klinter S, Schneider C, Voß I, Steinbüchel A. Plasmid addiction systems: perspectives and applications in biotechnology. Microb Biotechnol. 2010 Nov;3(6):634.
236. Kamruzzaman M, Iredell J. A ParDE-family toxin antitoxin system in major

resistance plasmids of Enterobacteriaceae confers antibiotic and heat tolerance. Sci Reports 2019 91. 2019 Jul 8;9(1):1–12.

237. Fernández-García L, Blasco L, Lopez M, Bou G, García-Contreras R, Wood T, et al. Toxin-Antitoxin Systems in Clinical Pathogens. Toxins (Basel). 2016 Jul 20;8(7).
238. Kedzierska B, Hayes F. Emerging Roles of Toxin-Antitoxin Modules in Bacterial Pathogenesis. Molecules. 2016 Jun 1;21(6):790.
239. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol 2019 182. 2019 Dec 19;18(2):67–83.
240. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. Genome Biol 2019 201. 2019 Jun 13;20(1):1–18.
241. Benedict MN, Mundy MB, Henry CS, Chia N, Price ND. Likelihood-Based Gene Annotations for Gap Filling and Quality Assessment in Genome-Scale Metabolic Models. PLoS Comput Biol. 2014;10(10).
242. Latendresse M. Efficiently gap-filling reaction networks. BMC Bioinformatics. 2014;15(1):1–8.
243. Thiele I, Vlassis N, Fleming RMT. FASTGAPFILL: Efficient gap filling in metabolic networks. Bioinformatics. 2014;30(17):2529–31.
244. Pan S, Reed JL. Advances in gap-filling genome-scale metabolic models and model-driven experiments lead to novel metabolic discoveries. Curr Opin Biotechnol. 2018;51:103–8.
245. Joyce AR, Palsson BØ. Predicting Gene Essentiality Using Genome-Scale in Silico Models. In Humana Press; 2008. p. 433–57.
246. Basler G, Grimbs S, Nikoloski Z. Optimizing metabolic pathways by screening for feasible synthetic reactions. BioSystems. 2012;109(2):186–91.
247. Yoo M, Bestel-Corre G, Croux C, Riviere A, Meynial-Salles I, Soucaille P. A Quantitative System-Scale Characterization of the Metabolism of Clostridium acetobutylicum. MBio. 2015;6(6):1–12.
248. Durot M, Bourguignon PY, Schachter V. Genome-scale models of bacterial metabolism: Reconstruction and applications. Vol. 33, FEMS Microbiology Reviews. 2009. p. 164–90.

249. Henry CS, Dejongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol. 2010;28(9):977–82.
250. Bioinformatics B, Xu Z, Sun X, Yu S. Genome-scale analysis to the impact of gene deletion on the metabolism of E. coli: Constraint-based simulation approach. BMC Bioinformatics. 2009;10(SUPPL. 1):1–10.
251. Huang H, Chai C, Li N, Rowe P, Minton NP, Yang S, et al. CRISPR/Cas9-Based Efficient Genome Editing in Clostridium ljungdahlii, an Autotrophic Gas-Fermenting Bacterium. ACS Synth Biol. 2016 Dec 16;5(12):1355–61.
252. Loubière P, Gros E, Paqiet V, Lindley ND. Kinetics and physiological implications of the growth behavior of Eubacterium limosum on glucose/methanol mixtures. J Gen Microbiol. 1992;138(1992):979–85.
253. Kelly WJ, Henderson G, Pacheco DM, Li D, Reilly K, Naylor GE, et al. The complete genome sequence of Eubacterium limosum SA11, a metabolically versatile rumen acetogen. Stand Genomic Sci. 2016;11(1):1–10.
254. Pacaud S, Loubiere P, Goma G. Methanol Metabolism by Eubacterium limosum B2: Effects of pH and Carbon Dioxide on Growth and Organic Acid Production. Curr Microbiol VOI. 1985;12:245–50.
255. Merico A, Sulo P, Piškur J, Compagno C. Fermentative lifestyle in yeasts belonging to the Saccharomyces complex. FEBS J. 2007;274(4):976–89.
256. Morales P, Rojas V, Quirós M, Gonzalez R. The impact of oxygen on the final alcohol content of wine fermented by a mixed starter culture. Appl Microbiol Biotechnol. 2015;99(9):3993–4003.
257. Zimmermann HF, Anderlei T, Büchs J, Binder M. Oxygen limitation is a pitfall during screening for industrial strains. Appl Microbiol Biotechnol. 2006 Oct;72(6):1157–60.
258. Rollero S, Roberts S, Bauer FF, Divol B. Agitation impacts fermentation performance as well as carbon and nitrogen metabolism in *Saccharomyces cerevisiae* under winemaking conditions. Aust J Grape Wine Res. 2018 Jul 1;24(3):360–7.
259. Tromans D. Temperature and pressure dependent solubility of oxygen in water: a thermodynamic analysis. Hydrometallurgy. 1998;48(3):327–42.
260. Cunliffe BJ. Avoiding Explosions By Means of Inerting Systems.

2001;(148):311–24.

261. Mehta R, Singhal P, Singh H, Damle D, Sharma AK. Insight into thermophiles and their wide-spectrum applications. 3 Biotech. 2016;6(1):1–9.
262. Sheng L, Kovács K, Winzer K, Zhang Y, Minton NP. Development and implementation of rapid metabolic engineering tools for chemical and fuel production in Geobacillus thermoglucosidasius NCIMB 11955. Biotechnol Biofuels. 2017;10(1):1–18.
263. Kumar M, Saini S, Gayen K. Acetone-butanol-ethanol fermentation analysis using only high performance liquid chromatography. Anal Methods. 2014 Jan 17;6(3):774–81.
264. Tsuey LS, Ariff A Bin, Mohamad R, Rahim RA. Improvements of GC and HPLC analyses in solvent (acetone-butanol-ethanol) fermentation byClostridium saccharobutylicum using a mixture of starch and glycerol as carbon source. Biotechnol Bioprocess Eng. 2006 Aug;11(4):293–8.
265. Brickwedde A, van den Broek M, Geertman J-MA, Magalhães F, Kuijpers NGA, Gibson B, et al. Evolutionary Engineering in Chemostat Cultures for Improved Maltotriose Fermentation Kinetics in Saccharomyces pastorianus Lager Brewing Yeast. Front Microbiol. 2017 Sep 8;8:1690.
266. Heap JT, Pennington OJ, Cartman ST, Minton NP. A modular system for Clostridium shuttle plasmids. J Microbiol Methods. 2009 Jul;78(1):79–85.
267. Bergkessel M, Guthrie C. Colony PCR. Methods Enzymol. 2013 Jan 1;529:299–309.
268. NEB Tm Calculator [Internet]. [cited 2020 Apr 23]. Available from: https://tmcalculator.neb.com/#!/main
269. Pardo B, Gómez-González B, Aguilera A. DNA double-strand break repair: How to fix a broken relationship. Vol. 66, Cellular and Molecular Life Sciences. 2009. p. 1039–56.
270. Jasin M, Rothstein R. Repair of strand breaks by homologous recombination. Cold Spring Harb Perspect Biol. 2013 Nov 1;5(11):a012740.
271. Shrivastav M, De Haro LP, Nickoloff JA. Regulation of DNA double-strand break repair pathway choice. Cell Res. 2008 Jan;18(1):134–47.
272. Boyaval P, Moineau S, Romero D a, Horvath P. Against Viruses in Prokaryotes. Science (80-). 2007;315(March):1709–12.

273. Horton RM. PCR-mediated recombination and mutagenesis - SOEing together tailor-made genes. Mol Biotechnol. 1995;3(2):93–9.
274. Pecheritsyna SA, Arkhipova O V., Suzina NE, Lysanskaya VY, Laurinavichyus KS, Shcherbakova VA. Intracellular polysaccharide of an anaerobic psychrophile Clostridium algoriphilum. Microbiology. 2011 Feb;80(1):37–42.
275. Price ND, Reed JL, Palsson B. Genome-scale models of microbial cells: Evaluating the consequences of constraints. Vol. 2, Nature Reviews Microbiology. Nature Publishing Group; 2004. p. 886–97.
276. Tramontano M, Andrejev S, Pruteanu M, Klünemann M, Kuhn M, Galardini M, et al. Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. Nat Microbiol. 2018 Apr 1;3(4):514–22.
277. Liew FM, Martin ME, Tappel RC, Heijstra BD, Mihalcea C, Köpke M. Gas Fermentation-A flexible platform for commercial scale production of low-carbon-fuels and chemicals from waste and renewable feedstocks. Front Microbiol. 2016;7(MAY):694.
278. Bertsch J, Müller V. Bioenergetic constraints for conversion of syngas to biofuels in acetogenic bacteria. Biotechnol Biofuels. 2015;8:210.
279. Dragosits M, Mattanovich D. Adaptive laboratory evolution - principles and applications for biotechnology. Vol. 12, Microbial Cell Factories. BioMed Central; 2013. p. 1–17.
280. Zhang Q, Jin YL, Fang Y, Zhao H. Adaptive evolution and selection of stress-resistant Saccharomyces cerevisiae for very high-gravity bioethanol fermentation. Electron J Biotechnol. 2019 Sep 1;41:88–94.
281. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Vol. 25, Nucleic Acids Research. Nucleic Acids Res; 1997. p. 3389–402.
282. Brooke JS. Stenotrophomonas maltophilia: an Emerging Global Opportunistic Pathogen. Clin Microbiol Rev. 2012 Jan;25(1):2.
283. Song. Draft Genome Sequence of Chemolithoautotrophic Acetogenic. 2015;3(1):1981–2.
284. Menouni R, Hutinet G, Petit MA, Ansaldi M. Bacterial genome remodeling

through bacteriophage recombination. FEMS Microbiol Lett. 2015 Jan 1;362(1):1–10.

285. Robert F, Pelletier J. Exploring the Impact of Single-Nucleotide Polymorphisms on Translation. Front Genet. 2018 Oct 30;9:507.
286. Kang S, Song Y, Jin S, Shin J, Bae J, Kim DR, et al. Adaptive Laboratory Evolution of Eubacterium limosum ATCC 8486 on Carbon Monoxide. Front Microbiol. 2020 Mar 11;0:402.
287. Ho SS, Urban AE, Mills RE. Structural Variation in the Sequencing Era: Comprehensive Discovery and Integration. Nat Rev Genet. 2020 Mar 1;21(3):171.
288. Chen Z, Erickson DL, Meng J. Polishing the Oxford Nanopore long-read assemblies of bacterial pathogens with Illumina short reads to improve genomic analyses. Genomics. 2021 May 1;113(3):1366–77.
289. Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M, Pevzner PA. Assembly of long error-prone reads using de Bruijn graphs. Proc Natl Acad Sci U S A. 2016 Dec 27;113(52):E8396–405.
290. Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Ech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 2018 Jul 2;46(W1):W537–44.
291. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004 Jul;14(7):1394–403.
292. Casadesús J, Low D. Epigenetic Gene Regulation in the Bacterial World. Microbiol Mol Biol Rev. 2006 Sep;70(3):830.
293. Adam M, Murali B, Glenn NO, Potter SS. Epigenetic inheritance based evolution of antibiotic resistance in bacteria. BMC Evol Biol. 2008 Feb 18;8(1):1–12.
294. Adhikari S, Curtis PD. DNA methyltransferases and epigenetic regulation in bacteria. FEMS Microbiol Rev. 2016 Sep 1;40(5):575–91.
295. Limoli DH, Jones CJ, Wozniak DJ. Bacterial Extracellular Polysaccharides in Biofilm Formation and Function. Microbiol Spectr. 2015 Jun 18;3(3).
296. Cuthbertson L, Kos V, Whitfield C. ABC Transporters Involved in Export of Cell Surface Glycoconjugates. Microbiol Mol Biol Rev. 2010 Sep;74(3):341.

297. Schmid J, Sieber V, Rehm B. Bacterial exopolysaccharides: biosynthesis pathways and engineering strategies. Front Microbiol. 2015;6(MAY).
298. Periwal V, Scaria V. Insights into structural variations and genome rearrangements in prokaryotic genomes. Bioinformatics. 2015 Jan 1;31(1):1–9.
299. Merrikh CN, Merrikh H. Gene inversion potentiates bacterial evolvability and virulence. Nat Commun 2018 91. 2018 Nov 7;9(1):1–10.
300. Takors R, Kopf M, Mampel J, Bluemke W, Blombach B, Eikmanns B, et al. Using gas mixtures of CO, CO2 and H2 as microbial substrates: the do's and don'ts of successful technology transfer from laboratory to production scale. Microb Biotechnol. 2018 Jul 1;11(4):606.
301. Hijum SAFT van, Medema MH, Kuipers OP. Mechanisms and Evolution of Control Logic in Prokaryotic Transcriptional Regulation. Microbiol Mol Biol Rev. 2009 Sep;73(3):481.
302. Dybvig K. DNA rearrangements and phenotypic switching in prokaryotes. Mol Microbiol. 1993;10(3):465–71.
303. Weinhold B. Epigenetics: The Science of Change. Environ Health Perspect. 2006;114(3):A160.
304. Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, et al. The Epigenomic Landscape of Prokaryotes. PLOS Genet. 2016 Feb 1;12(2):e1005854.
305. Pregnon G, Minton NP, Soucaille P. Genome Sequence of Eubacterium limosum B2 and Evolution for Growth on a Mineral Medium with Methanol and CO2 as Sole Carbon Sources. Microorganisms. 2022 Sep 1;10(9):1790.
306. VanGuilder HD, Vrana KE, Freeman WM. Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques. 2008 Apr 16;44(5):619–26.
307. Kongjan P, Tohlang N, Khaonuan S, Cheirsilp B, Jariyaboon R. Characterization of the integrated gas stripping-condensation process for organic solvent removal from model acetone-butanol-ethanol aqueous solution. Biochem Eng J. 2022;182:108437.
308. Abdehagh N, Dai B, Thibault J, Tezel FH. Biobutanol Separation from ABE Model Solutions and Fermentation Broths Using a Combined Adsorption-Gas Stripping Process. J Chem Technol Biotechnol. 2016 Mar 1;92.
309. Cai D, Chen H, Chen C, Hu S, Wang Y, Chang Z, et al. Gas stripping–

pervaporation hybrid process for energy-saving product recovery from acetone–butanol–ethanol (ABE) fermentation broth. Chem Eng J. 2016;287:1–10.

310. Wood JC, Marcellin E, Plan MR, Virdis B. High methanol-to-formate ratios induce butanol production in Eubacterium limosum. Microb Biotechnol. 2021;
311. Li X, Henson MA. Dynamic metabolic modelling predicts efficient acetogen–gut bacterium cocultures for CO-to-butyrate conversion. J Appl Microbiol. 2021 Dec 1;131(6):2899–917.
312. Hazra AB, Han AW, Mehta AP, Mok KC, Osadchiy V, Begley TP, et al. Anaerobic biosynthesis of the lower ligand of vitamin B12. Proc Natl Acad Sci U S A. 2015 Aug 25;112(34):10792–7.
313. Graf D, Laistner L, Klingel V, Radde NE, Weirich S, Jeltsch A. Reversible switching and stability of the epigenetic memory system in bacteria. FEBS J. 2023 Apr;290(8):2115–26.
314. Riber L, Hansen LH. Epigenetic Memories: The Hidden Drivers of Bacterial Persistence? Trends Microbiol. 2021 Mar 1;29(3):190–4.
315. Shin J, Kang S, Song Y, Jin S, Lee JSJK, Lee JSJK, et al. Genome Engineering of Eubacterium limosum Using Expanded Genetic Tools and the CRISPR-Cas9 System. ACS Synth Biol. 2019 Sep 20;8(9):2059–68.
316. Yan M, Wang B-H, Xu X, der Meister TJ, Tabγač H-T, Hwang F-F, et al. Extrusion of Dissolved Oxygen by Exopolysaccharide From Leuconostoc mesenteroides and Its Implications in Relief of the Oxygen Stress. Front Microbiol. 2018;9:2467.
317. Singh A, Cher Soh K, Hatzimanikatis V, Gill RT. Manipulating redox and ATP balancing for improved production of succinate in E. coli. Metab Eng. 2011;13(1):76–81.
318. Kozaeva E, Nieto-Domínguez M, Hernández AD, Nikel PI. Synthetic metabolism for in vitro acetone biosynthesis driven by ATP regeneration. RSC Chem Biol. 2022 Nov;3(11):1331–41.
319. Yang H, Zhu D, Kai L, Wang L, Zhang H, Zhang J, et al. Engineering Streptomyces albulus to enhance ε-poly-L-lysine production by introducing a polyphosphate kinase-mediated ATP regeneration system. Microb Cell Fact. 2023 Mar;22(1):51.

320. Gurumurthy CB, Grati M, Ohtsuka M, Schilit SLP, Quadros RM, Liu XZ. CRISPR: a Versatile Tool for Both Forward and Reverse Genetics Research. Hum Genet. 2016 Sep 1;135(9):971.
321. Long L, Guo D-D, Gao W, Yang W-W, Hou L-P, Ma X-N, et al. Optimization of CRISPR/Cas9 genome editing in cotton by improved sgRNA expression. Plant Methods. 2018 Dec 3;14(1):85.
322. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Vol. 28, Nucleic Acids Research. Oxford University Press; 2000. p. 27–30.
323. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Vol. 28, Protein Science. Blackwell Publishing Ltd; 2019. p. 1947–51.
324. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: Integrating viruses and cellular organisms. Nucleic Acids Res. 2021 Jan 8;49(D1):D545–51.
325. Dietrich HM, Kremp F, Öppinger C, Ribaric L, Müller V. Biochemistry of methanol-dependent acetogenesis in Eubacterium callanderi KIST612. Environ Microbiol. 2021 Aug 1;23(8):4505–17.
326. Heider J. A new family of CoA-transferases. Vol. 509, FEBS Letters. No longer published by Elsevier; 2001. p. 345–9.
327. Rangarajan ES, Li Y, Ajamian E, Iannuzzi P, Kernaghan SD, Fraser ME, et al. Crystallographic trapping of the glutamyl-CoA thioester intermediate of family I CoA transferases. J Biol Chem. 2005 Dec 30;280(52):42919–28.
328. Sheppard CA, Trimmer EE, Matthews RG. Purification and Properties of NADH-Dependent 5,10-Methylenetetrahydrofolate Reductase (MetF) from Escherichia coli. J Bacteriol. 1999;181(3):718.
329. Li Q, Chen J, Minton NP, Zhang Y, Wen Z, Liu J, et al. CRISPR-based genome editing and expression control systems in Clostridium acetobutylicum and Clostridium beijerinckii. Biotechnol J. 2016 Jul 1;11(7):961–72.
330. Ingle P, Groothuis D, Rowe P, Huang H, Cockayne A, Kuehne SA, et al. Generation of a fully erythromycin-sensitive strain of Clostridioides difficile using a novel CRISPR-Cas9 genome editing system. Sci Rep. 2019 Dec 1;9(1).
331. Sreekumar S, Baer ZC, Pazhamalai A, Gunbas G, Grippo A, Blanch HW, et al. Production of an acetone-butanol-ethanol mixture from Clostridium

acetobutylicum and its conversion to high-value biofuels. Nat Protoc. 2015 Mar 2;10(3):528–37.

332. Hoffmeister S, Gerdom M, Bengelsdorf FR, Linder S, Flüchter S, Öztürk H, et al. Acetone production with metabolically engineered strains of Acetobacterium woodii. Metab Eng. 2016 Jul 1;36:37–47.
333. Li RD, Li YY, Lu LY, Ren C, Li YX, Liu L. An improved kinetic model for the acetone-butanol-ethanol pathway of Clostridium acetobutylicum and model-based perturbation analysis. BMC Syst Biol. 2011 Jun 20;5(SUPPL. 1):S12.
334. Omorotionmwan BB, Wang H, Baker JP, Gizynski K, Yoo M, Akaluka C, et al. Chromosomal engineering of inducible isopropanol- butanol-ethanol production in Clostridium acetobutylicum. Front Bioeng Biotechnol. 2023 Jun 16;11:1218099.
335. Jang YS, Malaviya A, Lee SY. Acetone-butanol-ethanol production with high productivity using Clostridium acetobutylicum BKM19. Biotechnol Bioeng. 2013 Jun;110(6):1646–53.
336. Xue C, Zhao J, Lu C, Yang ST, Bai F, Tang IC. High-titer n-butanol production by clostridium acetobutylicum JB200 in fed-batch fermentation with intermittent gas stripping. Biotechnol Bioeng. 2012 Nov 1;109(11):2746–56.
337. Peng M, Liang Z. Degeneration of industrial bacteria caused by genetic instability. World J Microbiol Biotechnol 2020 368. 2020 Jul 18;36(8):1–16.
338. Fu H, Liang Y, Zhong X, Pan Z, Huang L, Zhang H, et al. Codon optimization with deep learning to enhance protein expression. Sci Reports 2020 101. 2020 Oct 19;10(1):1–9.
339. Trachootham D, Lu W, Ogasawara MA, Valle NR Del, Huang P. Redox Regulation of Cell Survival. Antioxid Redox Signal. 2008 Aug 1;10(8):1343.
340. Lee H, Bae J, Jin S, Kang S, Cho BK. Engineering Acetogenic Bacteria for Efficient One-Carbon Utilization. Front Microbiol. 2022 May 9;13:1667.
341. Katsyv A, Müller V. Overcoming Energetic Barriers in Acetogenic C1 Conversion. Front Bioeng Biotechnol. 2020 Dec 23;8:1420.
342. Balabanova L, Averianova L, Marchenok M, Son O, Tekutyeva L. Microbial and Genetic Resources for Cobalamin (Vitamin B12) Biosynthesis: From

Ecosystems to Industrial Biotechnology. Int J Mol Sci 2021, Vol 22, Page 4522. 2021 Apr 26;22(9):4522.

343. Dulay H, Tabares M, Kashefi K, Reguera G. Cobalt Resistance via Detoxification and Mineralization in the Iron-Reducing Bacterium Geobacter sulfurreducens. Front Microbiol. 2020 Nov 26;11:600463.
344. Majtan T, Frerman FE, Kraus JP. Effect of cobalt on Escherichia coli metabolism and metalloporphyrin formation. Biometals. 2011 Apr;24(2):335.
345. Assis DA de, Matte C, Aschidamini B, Rodrigues E, Záchia Ayub MA. Biosynthesis of vitamin B12 by Propionibacterium freudenreichii subsp. shermanii ATCC 13673 using liquid acid protein residue of soybean as culture medium. Biotechnol Prog. 2020 Sep 1;36(5):e3011.
346. Xia W, Chen W, Peng W fu, Li K tai. Industrial vitamin B12 production by Pseudomonas denitrificans using maltose syrup and corn steep liquor as the cost-effective fermentation substrates. Bioprocess Biosyst Eng. 2015 Jan 6;38(6).
347. Piwowarek K, Lipińska E, Hać-Szymańczuk E, Kieliszek M, Ścibisz I. Propionibacterium spp.—source of propionic acid, vitamin B12, and other metabolites important for the industry. Appl Microbiol Biotechnol. 2018 Jan 1;102(2):515.
348. Vitamin B12 (Cobalamin) Market Share, Growth & Forecast [Internet]. [cited 2022 Aug 5]. Available from: https://www.reportsanddata.com/report-detail/vitamin-b12-cobalamin-market
349. Frederic Y-H Chen A, Jung H-W, Tsuei C-Y, Liao Correspondence JC, Y-H Chen F, Liao JC. Converting Escherichia coli to a Synthetic Methylotroph Growing Solely on Methanol. Cell. 2020;182:933-946.e14.
350. Wegat V, Fabarius JT, Sieber V. Synthetic methylotrophic yeasts for the sustainable fuel and chemical production. Biotechnol Biofuels Bioprod. 2022;15(1):113.
351. Guo Y, Zhang R, Wang J, Qin R, Feng J, Chen K, et al. Engineering Yeasts to Grow Solely on Methanol or Formic acid coupled with CO2 fixation. Research Square; 2023.
352. Espinosa MI, Gonzalez-Garcia RA, Valgepea K, Plan M, Scott C, Pretorius IS, et al. Engineering and Evolution of Methanol Assimilation in Saccharomyces

cerevisiae. bioRxiv. 2020 Jan 1;717942.

353. Fabarius JT, Wegat V, Roth A, Sieber V. Synthetic Methylotrophy in Yeasts: Towards a Circular Bioeconomy. Trends Biotechnol. 2021;39(4):348–58.

354. Du Y, Zou W, Zhang K, Ye G, Yang J. Advances and Applications of Clostridium Co-culture Systems in Biotechnology. Front Microbiol. 2020 Nov 16;0:2842.

355. Singh V, Haque S, Niwas R, Srivastava A, Pasupuleti M, Tripathi CKM. Strategies for fermentation medium optimization: An in-depth review. Front Microbiol. 2017 Jan 6;7(JAN):227613.

356. Crater JS, Lievense JC. Scale-up of industrial microbial processes. FEMS Microbiol Lett. 2018 Jul;365(13).

357. Jeswani HK, Chilvers A, Azapagic A. Environmental sustainability of biofuels: a review. Proceedings Math Phys Eng Sci. 2020 Nov;476(2243):20200351.

358. Goodall ECA, Robinson A, Johnston IG, Jabbari S, Turner KA, Cunningham AF, et al. The essential genome of Escherichia coli K-12. MBio. 2018 Jan 1;9(1).

359. Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, et al. High-throughput analysis of gene essentiality and sporulation in Clostridium difficile. MBio. 2015 Feb;6(2):e02383.

360. Seong HJ, Han S-W, Sul WJ. Prokaryotic DNA methylation and its functional roles. J Microbiol. 2021;59(3):242–8.

361. Hiraoka S, Okazaki Y, Anda M, Toyoda A, Nakano S, Iwasaki W. Metaepigenomic analysis reveals the unexplored diversity of DNA methylation in an environmental prokaryotic community. Nat Commun. 2019;10(1):159.

362. Holland I, Davies JA. Automation in the Life Science Research Laboratory. Front Bioeng Biotechnol. 2020 Nov 13;8:571777.

363. Dimauro G, Barletta VS, Catacchio CR, Colizzi L, Maglietta R, Ventura M. A systematic mapping study on machine learning techniques for the prediction of CRISPR/Cas9 sgRNA target cleavage. Comput Struct Biotechnol J. 2022 Jan 1;20:5813–23.

364. Simeonidis E, Price ND. Genome-scale modeling for metabolic engineering. J Ind Microbiol Biotechnol. 2015 Mar;42(3):327–38.

365. Matthews NE, Cizauskas CA, Layton DS, Stamford L, Shapira P.

Collaborating constructively for sustainable biotechnology. Sci Rep. 2019;9(1):19033.

366. Naylor S, Chen JY. Unraveling human complexity and disease with systems biology and personalized medicine. Per Med. 2010 May;7(3):275–89.

367. Jackson NAC, Kester KE, Casimiro D, Gurunathan S, DeRosa F. The promise of mRNA vaccines: a biotech and industrial perspective. npj Vaccines. 2020;5(1):11.

368. Kumar V, Bansal V, Madhavan A, Kumar M, Sindhu R, Awasthi MK, et al. Active pharmaceutical ingredient (API) chemicals: a critical review of current biotechnological approaches. Bioengineered. 2022 Feb;13(2):4309–27.

369. Chen S, Li Z, Zhang S, Zhou Y, Xiao X, Cui P, et al. Emerging biotechnology applications in natural product and synthetic pharmaceutical analyses. Acta Pharm Sin B. 2022;12(11):4075–97.

370. García JL, Galán B. Integrating greenhouse gas capture and C1 biotechnology: a key challenge for circular economy. Microb Biotechnol. 2022 Jan;15(1):228–39.

371. Jatain I, Dubey KK, Sharma M, Usmani Z, Sharma M, Gupta VK. Synthetic biology potential for carbon sequestration into biocommodities. J Clean Prod. 2021 Nov 10;323:129176.

372. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016 Jul;44(W1):W16-21.

Appendices

8.1 Primers

Table 26 - Table listing all primers used in this study.

Name	**Sequence**
Glg Operon-LHA F-Aatii	atatatgacgtccttttgatattctaaaagagtatcttcc
Glg Operon-LHA F-AsiSI	atatatgcgatcgccttttgatattctaaaagagtatcttcc
Glg Operon-LHA R-SOE	tccagggttaaactgttcttgcaatttcatagtcaatcagtcctttc
Glg Operon-RHA F-SOE	ttgcaagaacagtttaaccctggaccttgagctattggctcaag
Glg Operon-RHA R-AscI	atatatggcgcgcctgttcgcccattaaagagg
Glg Operon-SCR-F	aagaataaaacatcaaagctcacc
Glg Operon-SCR-R	cgagcagagacgaaagtcg
glgB-LHA F-AsiSI	atatatgcgatcgccgcaccttaattgacacg
glgB-LHA R-SOE	tccagggttaaactgttcttgcaatttcatagtcaatcagtcctttc
glgB-RHA F-SOE	ttgcaagaacagtttaaccctggaatattatgaataccgaatgtgtcg
glgB-RHA R-AscI	atatatggcgcgcccgagaagattctgaaattccg
glgB-SCR-F	ccagatgtccttttatattatcatgg
glgB-SCR-R	ccatcctcaatgattgcg
GlgA-LHA F-AsiSI	atatatgcgatcgccagcagtgatgaagatgtgg
GlgA-LHA R-SOE	tccagggttaaactgttcttgcaaatccattgttatataaccattcctttac
GlgA RHA_F_SOE	ttgcaagaacagtttaaccctggattatcataatgaattttagacacaattcc
GlgA RHA_R_Acs1	atatggcgcgccccaatccgggataatttcc
GlgA-SCR-F	gaaaggaatggtccttaagc
GlgA-SCR-R	caccagcattcgtattcg
Ines SCR F	tttctctttgatttccacatatataaatct
Ines SCR R	gcactgttatgccttttgactatcacttaat
Ines SCR R2	tgttcaaaaaaataatggcgg
Ines SCR F2	ggtaatcgctttcataaaatatattacc
Ines SCR R3	ttctttctattcagcactgttatgc
PCB102 F1	gcaaaatacattcgttgatg
PBC102 R1	ctgttatgcctttgactatc
ColE1tra F2	ccatcaagaagagcgac
GlgA-LHA F-AatII	atatatgacgtccagcagtgatgaagatgtgg

glgB-LHA F-AatII	atatatgacgtccgcaccttaattgacacg
pbt2 LHA - F (Xhol)	atatactcgagaaattgtcattaaaatggcc
Cas9scr1R	gaaacttaatcatatgcgctaagg
Cas9scr1	atggataagaaatactcaataggcttag
Cas9scr2	gctttgtcattgggtttgac
Cas9scr3	gtcgataaaggtgcttcagc
Cas9scr4	gaacatattgcaaatttagctgg
Cas9scr5	ctgacttccgaaaagatttcc
Cas9scr6	gagttagaaaacggtcgtaaacg
pyrE Screen Test Forward2	cgttttccggtttatacatgc
pyrE Screen Test Reverse	cctggggcaggatatttatt
sg act G1 Spei	atttcactagtgctcaggcagctaaaaacaggttttagagctagaaatagcaagttaaaataag gctagtccgttatcaacttgaaaaagtggcaccgagtcg
sg act G2 Spei	atttcactagtgttgaccccagaaatcaggggttttagagctagaaatagcaagttaaaataag gctagtccgttatcaacttgaaaaagtggcaccgagtcg
pbt2seq1F	gcctttctgacaatataagtcc
pbt2seq1R	ctctgtttgtatagcgaaaaca
pbt2seq2F	ctagcatgaacagacaacaaaa
pbt2seq2R	ctctgtttgtatagcgaaaaca
ygewseqF	acaacttcatttccttcctcta
ygewseqR	taccgactgtttacaccaaata
hisI_sg3_ParaE	atcttaaggaggagttttcgtcgacccgcaaagaattatggacaagttttagagctagaaatag caagtt
hisI_sg3_P_{fdx}	taggaggttagttcatatggtcgacccgcaaagaattatggacaagttttagagctagaaatag caagtt
hisI_sg4_ParaE	atcttaaggaggagttttcgtcgacgcacagtcagatgataaagcgtttagagctagaaatag caaagtt
hisI_sg4_P_{fdx}	taggaggttagttcatatggtcgacgcacagtcagatgataaagcgtttagagctagaaatag caaagtt
Ga0111251_113033 For	gattactttgtcctgagacgac
Ga0111251_113033 Rev	tccattaaactatatgcactgg

Ga0111251_113034 For	ctacattctctgggtttctttt
Ga0111251_113034 Rev	tataagcacggctcaatctata
Ga0111251_113578/9 For	gcagattcataaagtcttcca
Ga0111251_113578/9 Rev	gtctaaaaacaatgatgtcgg
pbt2 synth forward	gtcgaccaatctggaagaaa
pbt2 synth reverse	cattttaaataaggcgcgcc
RiboSwitchSwapForward	cctccttagcagggtgctg
RiboSwitchSwapReverse2	cgatatctaagcctattgagtatttcttatccatct
Cas_pRECAS_seq_F/JM115	ttttgatactgtggcggtctgtatttcc
pCB102_pRECAS_seq_r/JM116	tgtaaatttctttctattcagcactgttatgcc
BSC Internal Forward	taaagcgtctgccaggtctt
BSC Internal Reverse	ggacagagtgttgctcgtca
BSC External Forward	agtccgagtctcagttcactgc
BSC External Reverse	agaggattttccggtgccgtta
pbt1 Internal Forward	cgccgataagatcaagcaat
pbt1 Internal Reverse	atacagggttcccagcatca
pbt1 External Forward	gcttcggagcttttggatcatc
pbt1 External Reverse	gataatcggtctgcctgctgta
metVmetF Internal Forward	cttcccatgtacagcagaaggt
metVmetF Internal Reverse	cacgttcatccacttcaacacc
metVmetF External Forward	ttacccagcaaatgacacca
metVmetF External Reverse	tctccggcaatcttgatacc

cop Internal Forward	gatctggtgcagaaggcagt
cop Internal Reverse	tggaggcgtcctctgagtag
mti Internal Forward	aacccagtcaccaccaagag
mti Internal Reverse	gtcgtacacaagctgctcca
mtiI Internal Forward	tcaaggaggcggtaaacaac
mtiI Internal Reverse	gatatccgccatcattttgc
metF1 Internal Forward	cgctgcttgacacacttgat
metF1 Internal Reverse	ggaaatgtctgcccatttgt
cop Outward 1	atccttgatttcatccagctcgc
cop Outward 2	gtgcagaaggcagtggatgag
mti Outward 1	cgatgaccatgccgctcttg
mti Outward 2	gcgaaatgagccgtcctcc
mtiI Outward 1	tcgatataggtcgcaccggc
mtiI Outward 2	gcagcatcctcaatcgtttgtt
Acetone AMP Forward SalI	gattacagtcgactcaataaaaatattatagtaaaggagaaaatcat
Acetone AMP Reverse Aatii	tgtaatcctgcagggcttcttatttttatgctagccta
pthl AMP forward sbfi	gattacacctgcaggcgacgctatttgtgccgatatc
pthl AMP Reverse	cgctctccctttttcctacaagt
Overlap primers fdx terminator forward aatii	tgtaatcgacgtcataaaaataagaagcctgcatttgcaggc
Overlap primers fdx terminator reverse ascI	tgtaatcccgcgcggataaaaataagaagcctgcaaatgcaggc
Seq Primer Acetone Step 2 Primer 1	gacgagcaaggcaagaccg
Seq Primer Acetone Step 2 Primer 2	aagttgataacggactagcctta
Seq Primer Acetone Step 4 Primer 1	ttcatatggtcgacatgaactctaaaat

Seq Primer Acetone Step 4 Primer 2	aaggagcctgcataaaatgattaatg
Seq Primer Acetone Step 4 Primer 3	tatcaaaatttaggaggttagttagaa
Seq Primer Acetone Step 4 Primer 4	tccacattagacgtccttaagataatc
Seq Primer Acetone Step 6 Primer 1	tatatgattatcttaaggacgtcataa
Seq Primer Acetone Step 6 Primer 2	tttgcaatttttacaagttttcatata
Seq Ga0111251_11595 F	gtggagaacgcgatcaatcatg
Seq Ga0111251_11595 R	tgcgggagtaattacttgcact
Seq Ga0111251_113714 F	acgggaattccagtgacttctc
Seq Ga0111251_113714 R	aacggcgtactggtgagtttat
AspS_Seq_F	attgctgtttaatgctgctgtg
AspS_Seq_R	cttgataacattggcaaactcc
AspS_Amp_F	ccttgagatctgtgtcagctgt
AspS_Amp_R	tgccaaacttaaattcggcagg
alaS_Seq_F1	catttgagcagcggttctgaaatc
alaS_Seq_F2	aaaggtgttcccaggacagg
alaS_Seq_R1	gtccagaatgtccttggtgg
alaS_Seq_R2	cgatcccagtctgtcccagtttac
alaS_Amp_F	catttgagcagcggttctgaaa
alaS_Amp_R	ttgaaacctcgatcccagtctg
113175_Amp_F	aatgttgcctctcggagacatt
113175_Amp_R	cggtatccagggaaaaacagga
113175_Seq_F	aatgttgcctctcggagacattc
113175_Seq_R	aaacagttatcaatgcggcg
113714_Amp_F	cactctgcctttacctggtcat
113714_Amp_R	aacggcgtactggtgagtttat
113714_Seq_F	cactctgcctttacctggtc

113714_Seq_R	aaggataaggagacaacagaagtg
892,000_Amp_F	cgtacccgttggagatgttaca
892,000_Amp_R	tgtccgcgataatgccaaaatc
892,000_Seq_F	cgtacccgttggagatgttacag
892,000_Seq_R	cttattcctgagtgcagggttatc
711,100_Amp_F	taaggatgcaaggttgtacgct
711,100_Amp_R	tggttatgcaggcagttggtat
711,100_Seq_F	taaggatgcaaggttgtacgc
711,100_Seq_R	ggtttccacggcacagttgatg
2126000_Amp_ F	atcctgaagttggacgtctgtc
2126000_Amp_ R	tggttatgcaggcagttggtat
2126000_Seq_ F	agcctttgctcaatgatttgtc
2126000_Seq_ R	cgtagtattgtcgctcaaatc
Acetone_full_forward_1	tttaaactgtggcactccaacc
Acetone_full_reverse_1	ttagatttgcttgagactttgccg
Acetone_full_forward_2	cctcacggcaaagtctcaag
Acetone_full_reverse_2	atatctgcaagaatgtgagagc
Acetone fdx-ctfA For	ccatgcgctccatcaagaag
Acetone fdx-ctfA Rev	tttgccagtatctgggttgc
Cas9 Amp PCR F	cgaagatcgaagagaagcagga
Cas9 Amp PCR R	ctcggtgccacttttcaagtt
act internal_forward	aacccaaacaagcagcag
act internal_reverse	tccaccaccgtcataaaaatc
metF1_Int2_F	tccccatcacccattttacc
metF1_Int2_R	cgggcaataatttcagccag
hisI_Int_F	ccgcaaagaattatggacaaag
hisI_Int_R	gcacagtcagatgataaagcag
87151_forward	tttatcaggaaacagctatgaccg
87151_reverse	ctttctattcagcactgttatgcc
suicide vector seq reverse	attcaccatcttgattgattgccg

pbt1 LHAF 2	ggagcttttggatcatctccttttc
pbt1 LHAR 2	acattttggagcgaaagtaagattgtgtaatgtctttcgatgtaaaaattgaag
pbt1 RHAF 2	atcttactttcgctccaaaatgtacggcgcgtcagggtc
pbt1 RHAR 2	acaaattcactgacacatttgtgc
pbt2 LHAF 2	tagaggctgcggccca
pbt2 LHAR 2	ccaagtcgaatttggtaagctgtaaaacctcgctttcattctatgtg
pbt2 RHAF 2	gcttaccaaattcgacttggattgtcccctcgctttctaaaataacc
pbt2 RHAR 2	ttccagtaaaccgtaaagagcaga
Cas3pyrF_fwd	tggatgaactgtataaataataattttagataaccagcttgtg
Cas3pyrF_rev	cttcttatttttatactagtcatatgcgatctctcctc
16S 8F	agagtttgatcctggctcag
16S 1492R	cggttaccttgttacgactt
ctfA fake to thlCac_fwd	tgaatggcgaatggcgctaggtcgacatgaacagcaaaattattc
ctfA fake to thlCac_rev	tttacctaggcctaggctagcacttttc
adc cds_fwd	ctagcctaggcctaggtaaatttaggaaggtgacttttatg
adc cds_rev	gcaggcttcttatttttatgggcgcgccctacttaagataatc
p19_fwd	cataaaaataagaagcctgcatttgc
p19_rev	ctagcgccattcgccattc
act -HA/Rep_fwd	catcaacaacttttacagtctgtaaaatacctttttaatagcag
act -HA/Rep_rev	gactgtaaaagttgttgatg
pbt1 ha_fwd	tagagtcgacgtcacgcgtcggagcttttggatcatctc
pbt1 ha_rev	aatgtgaaagtgggtcttaaacaaattcactgacacatttg
backbone_fwd	ttaagacccactttcacatttaagttgtttttc
backbone_rev	gacgcgtgacgtcgactc
pbt2 ha_fwd	tagagtcgacgtcacgcgtctagaggctgcggcccaga
pbt2 ha_rev	aatgtgaaagtgggtcttaattccagtaaaccgtaaagagcagaaagc
pbt1_cas3_r	atatatggtctctcaacggtccatggacagaggaccgtcgatcaca
pbt1_cas3_f	atatatggtctcgaaatagggctgtgtgatcgacggtcctctgtc
pbt2_cas3_r	atatatggtctctcaacggtccatggacaggggggccgtcgatgatacagccctatt
pbt2_cas3_f	atatatggtctcgaaatagggctgtatcatcgacggccccctgtccatggacc
pyrf backbone_fwd	taattttagataaccagcttgtg
pyrf backbone_rev	catatgcgatctctcctc
gfp fdx_rev	aagctggttatctaaaattattatttatacagttcatccatgc

gfp fdx_fwd	aggaggagagatcgcatatgtatgctagtaattaaatttttaatacgg
gfp seq for	aggagcgggtagatcctaag
gfp seq rev	ttgcccacgacctgatacag
Cas3 HA Seq For	tcaacgggaatcctgctctg
Cas3 HA Seq Rev	ccagccttcttattcggccttg
pbt2 HA_multiplex fwd	tagagtcgacgtcacgcgtctagaggctgcggcccaga
pbt2 HA_multiplex rev	ggagatgatccaaaagctccttccagtaaaccgtaaagagcagaaagc
ctfA fake to thlCac_fwd 2	cccaacttaatcgccttgcagtcgacatgaacagcaaaattattc
ctfA fake to thlCac_rev 2	tttacctaggcctaggctagcacttttc
adc cds_fwd 2	ctagcctaggcctaggtaaatttaggaaggtgacttttatg
adc cds_rev 2	ctggcgaaaggggatgtgcggcgcgccctacttaagataatc
87151 lacz_fwd	cccaacttaatcgccttgcagcacatccccctttcgcc
87151 lacz_rev	tgcaaggcgattaagttggg
act suicide_fwd	attaatcgtccaaaagcgctcggataccaatgaagggaaaag
act suicide_rev	agcgcttttggacgattaattg
GFP LHA Int SCR F	agtacgaagatgtgggcgat
GFP LHA Int SCR R	cgcgggttttatagttgcca
pbt1 LHAR 3	tgtaaaacctcgctttcattctatgtgtaatgtctttcgatgtaaaaattgaag
pbt1 RHAF 3	ttacacatagaatgaaagcgaggttttacaacggcgcgtcagggtc
cop LHAF TAT	tcggcagcaaaatggatgac
cop LHAR TAT	ttcgtcaaaattttattcgctgtctgttttatatcctccgatttttgtttttattcct
cop RHAF TAT	aaaacaaaaatcggaggatataaaacagacagcgaataaaattttgacgaa
cop RHAR TAT	cactgacccggttcaaaagg
cop TAT HA HiFi F	ttttttgaacaattgacaatcggcagcaaaatggatg
cop TAT HA HiFi R	aacttgcccactggccggcccactgacccggttcaaaag
TAT V HiFi F	ggccggccagtgggcaag
TAT V HiFi R	ttgtcaattgttcaaaaaaataatggcggcgcg
mti LHAF TAT	tttatctttgacccgctgcg
mti LHAR TAT	attattaatatcaaaagttttgtatcctccacttttagttcg
mti RHAF TAT	agtggaggatacaaaactttgatattaataatctaaaaacagaacaaaaaattcacta
mti RHAR TAT	ttcggcgtatctccagtgaa

mti TAT HA HiFi F	ttttttgaacaattgacaatttatctttgacccgctg
mti TAT HA HiFi R	aacttgcccactggccggccttttgtatcctccacttttag
mtiI LHAF TAT	tgaacgtggccaaggaaaag
mtiI LHAF TAT2	catcgcggcagcggt
mtiI LHAR TAT	aatgaaatcatcgggataggatgcctcctgttcgtc
mtiI RHAF TAT	gacgaacaggaggcatcctatcccgatgatttcattctccaagt
mtiI RHAR TAT	cggttaaacggcagtgagag
mtiI RHAR TAT2	gactgggcaatgccaatgctcacagt
mtiI TAT HA HiFi F	ttttttgaacaattgacaatgaacgtggccaaggaaaag
mtiI TAT HA HiFi F2	ttttttgaacaattgacaaatacctggagtaatgaatcc
mtiI TAT HA HiFi R	aacttgcccactggccggccctctcactgccgtttaacc
mtiI TAT HA HiFi R2	aacttgcccactggccggccgactgggcaatgccaatg
act LHAF TAT	tgctgatcattgttccgctg
act LHAR TAT	tttaatatttaaaaactttctcatcaccctaaaaatccttcaaatc
act RHAF TAT	aaagtttttaaatattaaaagatatataaaggagttttttaaaatgg
act RHAR TAT	tggacttctctgcgatcctc
act RHAR TAT2	tgaaatcgtagaggtttcaagacat
act TAT HA HiFi F	ttttttgaacaattgacaatgctgatcattgttccgc
act TAT HA HiFi R	aacttgcccactggccggcctggacttctctgcgatcc
act TAT HA HiFi R2	aacttgcccactggccggcctgaaatcgtagaggtttcaag
PBT1 LHAF TAT	agagaaaggggagcagcaat
PBT1 LHAR TAT	ccctgacgcgccgttgtaaaacctcgctttcattctatgt
PBT1 RHAF TAT	aagcgaggttttacaacggcgcgtcagg
PBT1 RHAR TAT	accgatcatgttgtccctgt
PBT1 TAT HA HiFi F	ttttttgaacaattgacaaagagaaaggggagcagcaataaaaatc
PBT1 TAT HA HiFi R	aacttgcccactggccggccaccgatcatgttgtccctg
PBT2 LHAF TAT	cgattaaagaaggcctccgc
PBT2 LHAF TAT2	taggagatgtttcagcaggaggt
PBT2 LHAR TAT	tttagaaagcgaggggacaatggttcagcttaaaccattcgataaaaag
PBT2 RHAF TAT	gaatggtttaagctgaaccattgtcccctcgctttctaaaataac
PBT2 RHAR TAT	tcaccaggggcttctttcaa
PBT2 TAT HA HiFi F	ttttttgaacaattgacaacgattaaagaaggcctcc
PBT2 TAT HA HiFi F2	ttttttgaacaattgacaataggagatgtttcagcagg

PBT2 TAT HA HiFi R	aacttgcccactggccggcctcaccaggggcttctttc
TAT SEQ_forward	aaaccctggcgttacccaac
TAT SEQ_reverse	ttcagattaaagccatcgaagg
87151_fwd 2	gcggccgctgtatccatatg
87151_rev 2	ggtcatagctgtttcctgataaaaaaattg
ctfA ctfB_fwd 2	tttttatcaggaaacagctatgaccatgaacagcaaaattattcgc
ctfA ctfB_rev 2	caacttctttcatctaaacagccatgggtctaag
ctfA ctfB_rev no THL	cttcctaaatttactaaacagccatgggtctaag
thlCac_fwd 2	atggctgtttagatgaaagaagttgtaatagctag
thlCac_rev 2	cttcctaaatttactagcacttttctagcaatattg
adc cds_fwd 2	gaaaagtgctagtaaatttaggaaggtgacttttatg
adc cds_rev 2	atggtcatatggatacagcggccgccttaagataatcatatataacttcagc
adc cds_fwd no THL	atggctgtttagtaaatttaggaaggtgacttttatg
83151 FAST_fwd	atggaacacgtagcatttggaag
83151 FAST_rev	gcatgtctgcaggcctcg
tcdr_fwd	ctcgaggcctgcagacatgcgttcaattttattaaaagtatcagaaaaaaat
tcdr_rev	tcaagtatactacgttttcagaaaaatacttaattaac
P_{fdx}_fwd	tgaaaacgtagtatacttgattggaggatttaaaac
P_{fdx}_rev	aattcattaatgatcatatttttattcttcggac
ptcb_fwd	aaatatgatcattaatgaatttaaagaaatatttacaatagaaatc
ptcb_rev	ccaaatgctacgtgttccatatgattttctcctttactataatattttattg
ptcb_fwd leak	ctcgaggcctgcagacatgcttaatgaatttaaagaaatatttacaatagaaatc
ptcb_rev leak	ccaaatgctacgtgttccatatgattttctcctttactataatattttattg
cop_loss_FOR	tactgtggcggtctgtatttcc
cop_loss_REV	tcttcatcatggtggtggtgag
cop PL SCR_1_for	tactgtggcggtctgtatttcc
cop PL SCR_5_rev	tgagcagtgaggacatgcctac
pMTL83151_TetR_Cas3_FDX_fwd	taattttagataaccagcttgtg
pMTL83151_TetR_Cas3_FDX_rev	catatgaactaacctcctaaattttg
FAST_fwd	ttaggaggttagttcatatgatggaacacgtagcatttg
FAST_rev	aagctggttatctaaaattatcataccctcttaacgaaaac
83151_No_Fast_fwd	gcgtaatagcgaagaggcc

83151_No_Fast_rev	gcatgtctgcaggcctcg
Tcdr_region_fwd	ctcgaggcctgcagacatgctttatatacttggtttatttacttgattatttc
Tcdr_region_rev	gggcctcttcgctattacgcttacaagttaaaataattttcatagtctttttttatttttttaaatg
pMTL83151tcdRCHp1_fwd	acaattcatttcttatttttattaagtg
pMTL83151tcdRCHp1_rev	caattgttcaaaaaaataatggc
pTcdB_fwd	cattatttttttgaacaattgttaatgaatttaaagaaatatttacaatagaaatc
pTcdB_rev	cgtgttccatatgattttctcctttactataatattttattg
FAST_fwd2	agaaaatcatatggaacacgtagcatttg
FAST_rev2	aaaaaataagaaatgaattgttcataccctcttaacgaaaac
metF1 LHAF TAT	ccagttcgtgatttccccagta
metF1 LHAR TAT	tctttcaatgctatcggaaaatccctccatttttttgatttctactga
metF1 RHAF TAT	aaaaaatggagggattttccgatagcattgaaagatcggac
metF1 RHAR TAT	ctcggtctttttcgttcatcgg
metF1 RHAR TAT2	ggctattgtgagccctttgc
metF1 TAT HA HiFi F	ttttttgaacaattgacaaccagttcgtgatttcccc
metF1 TAT HA HiFi R	aacttgcccactggccggccctcggtctttttcgttcatc
metF1 TAT HA HiFi R2	aacttgcccactggccggccggctattgtgagccctttg
metVmetF LHAF TAT	cagtaaaagtgctgttcctggc
metVmetF LHAR TAT	cagtaattctcattagttcactactcttttacttcctccgtctatatttaatccg
metVmetF RHAF TAT	agacggaggaagtaaaagagtagtgaactaatgagaattactgtaataggt
metVmetF RHAR TAT	agcgacgaagccttttacctaa
metVmetF RHAR TAT2	gccctcacagggacagctt
metVmetF TAT HA HiFi F	ttttttgaacaattgacaacagtaaaagtgctgttcc

metVmetF TAT HA HiFi F2	ttttttgaacaattgacaacagtaaaagtgctgttcctggc
metVmetF TAT HA HiFi R	aacttgcccactggccggccagcgacgaagccttttac
metVmetF TAT HA HiFi R2	aacttgcccactggccggccgccctcacagggacagct
act TAT SCR F	gctttattttgtgctgctggga
act TAT SCR R	gccccatccagataataggctc
cop TAT SCR F	agactggggaatactttgcagg
cop TAT SCR R	atgacaggctgtgtgacaatct
mti TAT SCR F	catggaaaactgcgggtacac
mti TAT SCR R	taagagattgatacgctccgcc
mtiI TAT SCR F	aggaaatcaatgacgaagccct
mtiI TAT SCR R	acaggcttgtcataacacggat
pbt1 TAT SCR F	tccttttggttatcgtggcgta
pbt1 TAT SCR R	atcccttccagaccgtagtaca
pbt2 TAT SCR F	gtttcagggacgggtctattgt
pbt2 TAT SCR R	atatctcacaggggcaaaagca
metF1 TAT SCR F	cggcattcaatttcagggttgt
metF1 TAT SCR R	cctgaccatctatccggttgag
metVmetF TAT SCR F	aacgttgctctgtccgaagtat
metVmetF TAT SCR R	tcacgataacgccatgagttca
Euk-A	aacctggttgatcctgccagt
Euk-B	gatccttctgcaggttcacctac
Euk-570F	gtaattccagctccaatagc
pMTL87171_CA_acetone_fwd	atgaacagcaaaattattcg
pMTL87171_CA_acetone_rev	ggtcatagctgtttcctg
TetR_fwd	atcaggaaacagctatgaccttaagacccactttcacatttaag
TetR_rev	cgaataattttgctgttcatcgctacgatcaagcttattttaattatac
P_{fdx}_Acetone_fwd	atcaggaaacagctatgaccactagtataaaaataagaagcctg
P_{fdx}_Acetone_rev	cgaataattttgctgttcatcatatgaactaacctcctaaattttg

mti TAT HA HiFi R 2	aacttgcccactggccggccgtcattgtctggacaacaaac
tcdr_fwd2	ctcgaggcctgcagacatgcttacaagttaaaataattttcatagtctttttttatttttttaaatg
tcdr_rev2	tcaagtatacatgcaaaagtctttttatgaattaattg
P_{fdx}_fwd tcdr P_{fdx} 2	acttttgcatgtatacttgattggaggatttaaaac
ptcb AMP F	tacatcaccgacgagcaag
ptcb AMP R	aatcagtacctggagccac
tcdr region AMP F	gacagaagagcagaaggtgtca
tcdr region AMP R	actacaactaacactgctgcca
pbt2 Cas3 G2 F	atatatggtctcgatcattcatctttgagagcactgccgcgtcat
pbt2 Cas3 G2 R	atatatggtctctcaactgatggctatgacgcggcagtgctctcaa
pbt2 Cas3 G3 F	atatatggtctcgatcaggcaggaaattttcggcatctcaca
pbt2 Cas3 G3 R	atatatggtctctcaacatgccctgggctgtgagatgccgaaaatt
pbt2 Cas3 G4 F	atatatggtctcgatcaaaggcggtaacggcatttaagagtttgccggt
pbt2 Cas3 G4 R	atatatggtctctcaacgcggaccggcaaactcttaaatgccgt
tcdr region AMP R2	ctgtggatttttctccgtgagc
tcdr region AMP F2	atcctgacagggcaaaagatgt
tcdr region AMP R3	tgccgctgcattttcttttcat
tcdr region AMP F	gacagaagagcagaaggtgtca
tcdr region AMP R2	ctgtggatttttctccgtgagc
tcdr region AMP F2	atcctgacagggcaaaagatgt
tcdr region AMP R2	ctgtggatttttctccgtgagc
tcdr region AMP F	gacagaagagcagaaggtgtca
tcdr region AMP R3	tgccgctgcattttcttttcat
TAT SCR F	aaaccctggcgttacccaac
TAT SCR R	aaacggcaaatgtgaaatccgtc
TAT PyrF KI SCR F	tctggacattacctgtgagg
TAT PyrF KI SCR R	acagtcacttcatcggccac
Cas3 Guide SCR F	cttcgattccgacctcattaagc
Cas3 Guide SCR R	gagttctctaacggcttgatgtg
tyr HA HiFI F	ttttttgaacaattgacaacagacgggtggagggcct
tyr HA HiFI R	aacttgcccactggccggcccctcaaaatactctgatataacagccttgttggc
tyr LHAR	tgaaccagttgagcgcataggagagtgaaccattcatgataaaagtacg
tyr LHAF	cagacgggtggaggg
tyr RHAF	tcatgaatggttcactctcctatgcgctcaactggttcatcag
tyr RHAR	cctcaaaatactctgatataacagccttgtt

ilvE HA HiFI F	ttttttgaacaattgacaaaagaaaatgcagagcttg
ilvE HA HiFI R	aacttgcccactggccggccttcccatattcattgacc
ilvE LHAF	aagaaaatgcagagcttgaagagggc
ilvE LHAR	tggtttttttatgttttttatttactgcagaagttcaatcacactcttacgg
ilvE RHAF	agtgtgattgaacttctgcagtaaataaaaaaacataaaaaaaaccagtaagaatcacgaaa
ilvE RHAR	ttcccatattcattgaccgggtcat
glyA HA HiFI F	ttttttgaacaattgacaaccgagatatgagccgatg
glyA HA HiFI R	aacttgcccactggccggccaatctcaaaggcatgcac
glyA LHAF	ccgagatatgagccgatgatcc
glyA LHAR	gcatcttcataaattctcttttaggacacgatccgaaagccc
glyA RHAF	tttcggatcgtgtcctaaaagagaatttatgaagatgctgtccat
glyA RHAR	aatctcaaaggcatgcacgttc
glyA SCR F	cgatgacaacgaggatgattgc
glyA SCR R	gtacaggcggttcttcagtgta
IlvE SCR F	ttcaggatcaggggtatccagt
IlvE SCR R	tccgtctctgtatttgccgaaa
tyr SCR F	ttgatttcccagtggatcaggg
tyr SCR R	tttagcacagaatggcccagat
Cas3 VG Amp Rev	atttcgagaccatttaaatacatctcatgttaatcttcaaccgctacgatca
Cas3 VG Amp For	gttgagagaccgttgaagattaacatgagatgtatttaaattgggaaaaccct
mti TAT SCR2 F	tatatcctgtttgagccgtccg
mti TAT SCR2 R	ccaaagtccgctgtgacaaaat
act TAT SCR2 F	gctgcctcagttggaaaaatga
act TAT SCR2 R	tataagcgtcaaccgggaaagt
cop TAT SCR2 F	agactggggaatactttgcagg
cop TAT SCR2 R	agttgacgtcgatataggtcgc
mti TAT SCR2 F	agactggggaatactttgcagg
mti TAT SCR2 R	attcaattgatgagcagccagc
mtiI TAT SCR2 F	aggaaatcaatgacgaagccct
mtiI TAT SCR2 R	taagagattgatacgctccgcc
metF1 TAT SCR2 F	cggcattcaatttcagggttgt
metF1 TAT SCR2 R	cactgctggatgcgtataaagc
metVmetF TAT SCR2 F	aacgttgctctgtccgaagtat

metVmetF TAT SCR2 R	tcggaatacagccacggtttaa
pbt1 TAT SCR2 F	gacgcaaaattgacgatggaca
pbt1 TAT SCR2 R	tatgtctccacgccggttaaaa
pbt2 TAT SCR2 F	gtttcagggacgggtctattgt
pbt2 TAT SCR2 R	ggaccactagcatgaacagaca
IlvE TAT SCR2 F	ttcaggatcaggggtatccagt
IlvE TAT SCR2 R	gcaccctaacggttttcaagac
glyA TAT SCR2 F	cgatgacaacgaggatgattgc
glyA TAT SCR2 R	aagctcgcccttattctcaaca
tyr TAT SCR2 F	cggtagtggacacggttaaaga
tyr TAT SCR2 R	tttagcacagaatggcccagat
act LHAF TAT2	ggtgtgattacaattaatcgtccaaaa
act RHAR TAT3	atggctaccttaaagccct
cop LHAF TAT2	caaggactgcgcctatgaagg
cop RHAR TAT2	tcattttgccattggtgggg
mti LHAF TAT2	tgcttttgctaaagtcaaaagcc
mti RHAR TAT2	cctctgagtagccgtcc
mtiI LHAF TAT2	ttagtacacactgcgcgc
mtiI RHAR TAT2	aactctccctccagctgttc
metVmetF LHAF TAT2	aacacaggctgattttgtagcaaaat
metVmetF RHAR TAT3	gctttaacaccatcgtcagaaacttc
pbt1 LHAF TAT2	ctccatgatcatgacccccat
pbt1 RHAR TAT2	ccggattaacggcgatcc
pbt2 LHAF TAT2	ggtcggcatagcggtacag
pbt2 RHAR TAT2	ccctccagcgccttcag
act T HiFi F	ttttttgaacaattgacaaggtgtgattacaattaatcgtc
act T HiFi R	aacttgcccactggccggccatggctaccttaaagccc
cop T HiFi F	ttttttgaacaattgacaacaaggactgcgcctatgaag
cop T HiFi R	aacttgcccactggccggcctcattttgccattggtggg
mti T HiFi F	ttttttgaacaattgacaatgcttttgctaaagtcaaaagcc
mti T HiFi R	aacttgcccactggccggcccctctgagtagccgtccg
mtiI T HiFi F	ttttttgaacaattgacaattagtacacactgcgcgc

mtiI T HiFi R	aacttgcccactggccggccaactctccctccagctgttc
metVmetF T HiFi F	ttttttgaacaattgacaaaacacaggctgattttgtag
metVmetF T HiFi R	aacttgcccactggccggccgctttaacaccatcgtcag
pbt1 T HiFi F	ttttttgaacaattgacaactccatgatcatgacccc
pbt1 T HiFi R	aacttgcccactggccggccccggattaacggcgatcc
pbt2 T HiFi F	ttttttgaacaattgacaaggtcggcatagcggtacag
pbt2 T HiFi R	aacttgcccactggccggccccctccagcgccttcaga
GFP SCR F2	gaaggaaccatgaaaaaggggc
GFP SCR R2	tatgatcgcgtttttcgttcgg
FAST SCR F2	aagatttcagcgaagacgacct
FAST SCR R2	tgaagtatgtttccgtcaccgt
PyrF KO SCR F	acgaagcaaacgcaaatgatga
PyrF KO SCR R	gcttctacttggtgggtgatga
lac tcdr HiFi F	ggcgcgccgccattattt
lac tcdr HiFi R	gctagctatatactcgagtatataccgcg
lac tcdr ptcdb_fwd	tactcgagtatatagctagcttaatgaatttaaagaaatatttacaatagaaatc
lac tcdr ptcdb_rev	cgtgttccatatgattttctcctttactataatatttttattg
lac tcdr Fast_fwd	agaaaatcatatggaacacgtagcatttg
lac tcdr Fast_rev	aaaaataatggcggcgcgcctcataccctcttaacgaaaac
83151 leak_fwd	cataaaaataagaagcctgcatttgcagg
83151 leak_rev	atggatacagcggccgcg
leak ptcdb_fwd	accgcggccgctgtatccatttaatgaatttaaagaaatatttacaatagaaatc
leak ptcdb_rev	cgtgttccatatgattttctcctttactataatatttttattg
leak Fast_fwd	agaaaatcatatggaacacgtagcatttg
leak Fast_rev	gcaggcttcttatttttatgtcataccctcttaacgaaaac
83151 ptcdb fast_fwd	ttaatgaatttaaagaaatatttacaatagaaatcaaattttagaattaactttattgtaaaatcaata acttaatctaagaatatcttaatttttatattttatatag
83151 ptcdb fast_rev	atggatacagcggccgcg
tcdr_fwd	accgcggccgctgtatccatttacaagttaaaataattttcatagtctttttttatttttttaaatg
tcdr_rev	tcaagtatacatgcaaaagtctttttatgaattaattg
P_{fdx}_fwd	acttttgcatgtatacttgattggaggatttaaaac
P_{fdx}_rev	aatatttctttaaattcattaatgatcatatttttattcttcggac
pbt1 cas3 g2 F	atatatggtctcgatcaaaaatgaatgatcgcggccagatcaagggctgt
pbt1 cas3 g2 R	atatatggtctctcaacacagcccttgatctggccgcgatcattcatttttga
pbt1 cas3 g3 F	atatatggtctcgatcacagccaacctccttggcatattgcttgatctta

pbt1 cas3 g3 R	atatatggtctctcaactaagatcaagcaatatgccaaggaggttggctgtga
pbt1 cas3 g4 F	atatatggtctcgatcatctcgattttatccttcaggctcggctccacga
pbt1 cas3 g4 R	atatatggtctctcaactcgtggagccgagcctgaaggataaaatcga
ilvE Short HA F	ttttttgaacaattgacaaaataatttggttttcgggac
ilvE Short HA R	aacttgcccactggccggccatattaattgtattaacatagtctatttgatc
tyr Short HA F	ttttttgaacaattgacaaggtattaagcgctcccagctccc
tyr Short HA R	aacttgcccactggccggcccacggccctacctgcgcc
Jon Ace back_fwd	atgaactctaaaataattagatttgaaaatttaaggtcattctttaaagatgggatgacaa
Jon Ace back_rev	gcggccgccattgc
Jon Ace P_{fdx} F	gcggccgcaatggcggccgcactagtataaaaataagaagcctg
Jon Ace P_{fdx} R	ctaattattttagagttcatcatatgaactaacctcctaaattttg
glya short_fwd	ttttttgaacaattgacaaccaggagcccgttattatg
glya short_rev	aacttgcccactggccggcccatgataccggccggatag
TAT KI backbone_fwd	ggccggccagtgggcaag
TAT KI backbone_rev	ttgtcaattgttcaaaaaaataatggcggcgcg
Fast hisI LHA fwd	cattatttttttgaacaattgacaatgagaaaaaaatgccgtctcaggtttgaaag
Fast hisI LHA rev	aattactagcatagcgcttgtgcgccgaaca
Fast hisI RHA fwd	taagagggtatgatgtggtttgcaatgcccataac
Fast hisI RHA rev	ttttcaacttgcccactggccggccgcgagacctccggcaact
fast tat ki fwd	gcgcacaagcgctatgctagtaattaaatttttaatacgg
fast tat ki rev	ttgcaaaccacatcataccctcttaacgaaaac
hisI TARS fast backbone_fwd	ggccggccagtgggcaag
hisI TARS fast backbone_rev	gactctagaggatccccgggtacc
LHA and fdx_fwd	cggtacccggggatcctctagagtcagtccttcaggactattac
LHA and fdx_rev	ttttagagttcatcatatgaactaacctcctaaattttg
Acetone operon_fwd	ttagttcatatgatgaactctaaaataattagatttgaaaatttaag
Acetone operon_rev	ttttaataaggccttaagataatcatatataacttcagc
RHA_fwd	tgattatcttaaggccttattaaaaaaagaacctg
RHA_rev	ttttcaacttgcccactggccggccgaaaatatctttttggataaaatttgtg
pbt2 int F	gctcaacggacatttcggag
pbt2 int R	caaacaggcccagaccaaaa
TH SCR F	cgccattcagagtttaggacg

TH SCR R	tgaaatccgtcacatactgcg
ch20_fwd	atttttgataggattagaattgattttc
ch20_rev	taaatctcctcctatttcttac
RepH_fwd	aagaaataggaggagatttaatgaaaagaagggcgtatatg
RepH_rev	attctaatcctatcaaaaatttatttttttcttttttactcttctttacatg
ch20 reph_fwd	ttctacctcctaacctataaaattag
ch20 reph_rev	tttaaaatacagctttaaaacaaaaaaatttc
Tet system_fwd	gttttaaagctgtattttaaattaagacccactttcacatttaag
Tet system_rev	ttataggttaggaggtagaaatgtatttcctcctcttcaatatatttaag
Fast KI SCR F	ggacaggtgctgtcaatctg
Fast KI SCR R	cttagccgcatctgtcttcc
TAT Backbone_fwd	ggccggccagtgggcaag
TAT Backbone_rev	ttgtcaattgttcaaaaaaataatggcggcgcg
glya LHA 1000_fwd	cattatttttttgaacaattgacaattagaaaaactgtctaaccggc
glya LHA 1000_rev	aaattctcttttaggacacgatccgaaagcc
glya RHA 1000_fwd	cggatcgtgtcctaaaagagaatttatgaagatgc
glya RHA 1000_rev	ttttcaacttgcccactggccggccaggcttccataaaatacc
cop TAT SCR2 F	agactggggaatactttgcagg
cop TAT SCR2 R	agttgacgtcgatataggtcgc
mtiI TAT SCR2 F	aggaaatcaatgacgaagccct
mtiI TAT SCR2 R	taagagattgatacgctccgcc
IlvE TAT SCR2 F	ttcaggatcaggggtatccagt
IlvE TAT SCR2 R	gcaccctaacggttttcaagac
glyA TAT SCR2 F	cgatgacaacgaggatgattgc
glyA TAT SCR2 R	aagctcgcccttattctcaaca
IlvE LHA 1000_fwd	cattatttttttgaacaattgacaatttacaagctttagaagctatc
IlvE LHA 1000_rev	gttttttttatttactgcagaagttcaatcac
IlvE RHA 1000_fwd	gaacttctgcagtaaataaaaaaacataaaaaaaaccagtaagaatcacgaaaaaaatggttgaca
IlvE RHA 1000_rev	ttttcaacttgcccactggccggccctccggccatgtggcgga
cop LHA 500_fwd	cattatttttttgaacaattgacaacttatgatgcgcgacctg
cop LHA 500_rev	attcgctgtctgttttatatcctccgatttttgtttttattc
cop RHA 500_fwd	ggaggatataaaacagacagcgaataaaattttgacg
cop RHA 500_rev	ttttcaacttgcccactggccggccgccttcggccttaagcttg
cop LHA 1000_fwd	cattatttttttgaacaattgacaaatatccgaaaagccattttc

cop LHA 1000_rev	tattcgctgtctgttttatatcctccgatttttgtttttattc
cop RHA 1000_fwd	gaggatataaaacagacagcgaataaaattttg
cop RHA 1000_rev	ttttcaacttgcccactggccggccttaagatgcttactgttaaatgttaaaaatttttttg
mtiI LHA 1000_fwd	cattatttttttgaacaattgacaagccgaaaaggttaaggctg
mtiI LHA 1000_rev	tgaaatcatcgggataggatgcctcctgttc
mtiI RHA 1000_fwd	gaggcatcctatcccgatgatttcattctcc
mtiI RHA 1000_rev	ttttcaacttgcccactggccggcccttgagcacagtatcgcc
mtiI LHA 500_fwd	cattatttttttgaacaattgacaatgaacgtggccaaggaaaag
mtiI LHA 500_rev	gaaatcatcgggataggatgcctcctgttcg
mtiI RHA 500_fwd	ggaggcatcctatcccgatgatttcattctcc
mtiI RHA 500_rev	ttttcaacttgcccactggccggccgactgggcaatgccaatg
Adapted Backbone F	ggccggccagtgggcaag
Adapted Backbone R	ttgtcaattgttcaaaaaaataatggcggcgcg
AT LHA 500_fwd	ttttttttgaacaattgacaataaatcatatttaaacttatcttccc
AT LHA 500_rev	ggaaagaaaaaaagccaaatacagaggttc
AT RHA 500_fwd	atttggctttttttctttccttttatattttatattaaaactcatctaaatctaacatctaaaaaaatcagagccc
AT RHA 500_rev	aacttgcccactggccggccgctgcacctcaccgacgc
AT LHA 750_fwd	ttttttttgaacaattgacaaaacaaggtttttcttccatataaag
AT LHA 750_rev	ggaaagaaaaaaagccaaatacagaggttc
AT RHA 750_fwd	atttggctttttttctttccttttatattttatattaaaactcatctaaatc
AT RHA 750_rev	aacttgcccactggccggccatacccgctgctgctatc
M23 LHA 500_fwd	ttttttttgaacaattgacaagtccctctggcaattggttc
M23 LHA 500_rev	ggaggaaaactaagacaacaataaaaaacaaggtactgg
M23 RHA 500_fwd	tgttgtcttagttttcctcctggctttc
M23 RHA 500_rev	aacttgcccactggccggccctacgatgaactggaatg
M23 LHA 750_fwd	ttttttttgaacaattgacaaagctaaaaaaattcctgatcc
M23 LHA 750_rev	ggaggaaaactaagacaacaataaaaaacaagg
M23 RHA 750_fwd	tgttgtcttagttttcctcctggctttc
M23 RHA 750_rev	aacttgcccactggccggccaagatggaatctggcagg
2IST LHA 750_fwd	ttttttttgaacaattgacaacacgcctttacatcccgttttattc
2IST LHA 750_rev	cgcgatgatcctcgccctcctattcctg
2IST RHA 750_fwd	ggagggcgaggatcatcgcggtgactaac
2IST RHA 750_rev	aacttgcccactggccggcctctccgtcatttttttaaatttcc
2IST LHA 500_fwd	ttttttttgaacaattgacaattgtgaagattgaagtcattcatg

2IST LHA 500_rev	cgcgatgatcctcgccctcctattcctg
2IST RHA 500_fwd	ggagggcgaggatcatcgcggtgactaac
2IST RHA 500_rev	aacttgcccactggccggccgcacacttcggatattttc
pbt1 int F	aaggtggtttcagttgcagc
pbt1 int R	catggaattgaccttcgcgt
pbt2 int F	catggcgacgctgttgag
pbt2 int R	caaacaggcccagaccaaaa
act int F	ttttgttgggagctgctgtc
act int R	caactcgtcccggaaaatcc
2IST TAKOS_fwd	gatcatcgcggtgactaac
2IST TAKOS_rev	ctcgccctcctattcctg
2IST_KI fwd	tgcaggaataggagggcgagatgaaagaacgaatcgac
2IST_KI rev	ggttagtcaccgcgatgatctcaaacatagatatactcgc
AT TAKOS_fwd	ttttctttccttttatattttatattaaaactc
AT TAKOS_rev	aaagccaaatacagaggttc
AT KI_fwd	gaacctctgtatttggctttttaaaacaatgctttaatcatgttg
AT KI_rev	aaaatataaaaggaaagaaaaatggataagcccgaaaag
M23 TAKOS_fwd	gttttcctcctggctttc
M23 TAKOS_rev	taagacaacaataaaaaacaagg
M23 KI_fwd	tgtttttattgttgtcttattaccagtgcatatagtttaatg
M23 KI_rev	ttgaaagccaggaggaaaacatgaaaaagaagtcaattttctc
2IST SCR F	acatcccgttttattctgggct
2IST SCR R	ccaccgcagttcctcaaaaatt
AT SCR F	ctttccgccttttttcgtaagca
AT SCR R	tctgaatgacgaccgtacgttt
M23 SCR F	aggggcacagctttgatagaaa
M23 SCR R	ctcaaaagcccaaaggaaccag
InIB LHA 750_fwd	ttttttgaacaattgacaagtacagcatccaattcgag
InIB LHA 750_rev	tgataaaaagcatctaaatttcctttctaaattttg
InIB RHA 750_fwd	aatttagatgctttttatcattggaaagacacag
InIB RHA 750_rev	aacttgcccactggccggccgacagcaatgccaaacaac
InIB LHA 500_fwd	ttttttgaacaattgacaaattgtcacaagccagagtattg
InIB LHA 500_rev	tgataaaaagcatctaaatttcctttctaaattttgaac
InIB RHA 500_fwd	aatttagatgctttttatcattggaaagacacagg
InIB RHA 500_rev	aacttgcccactggccggcccacttcaatccccgctcaatataaaaag

InIB SCR F	aagactggctacacctttacgg
InIB SCR R	cagcaatgccaaacaacaggta
cop Insert 500_KI_fwd	aaaaatcggaggatataaaaatggcaattttagaagatattcaaaac
cop Insert 500_KI_rev	aaaattttattcgctgtctgctacattgccatcatgcg
cop Backbone_KI_fwd	cagacagcgaataaaattttg
cop Backbone_KI_rev	ttttatatcctccgatttttgtttttattc
mti Insert 500_KI_fwd	aaaaagtggaggatacaaaaatggcaaagaaatttgataaac
mti Insert 500_KI_rev	ttagattattaatatcaaaagttaagcctttaagccgtag
mti Backbone_KI_fwd	ttttgtatcctccacttttttag
mti Backbone_KI_rev	cttttgatattaataatctaaaaacagaac
mtiI Insert 500_KI_fwd	acgaacaggaggcatcctatatgctgacaattgttggtgaactc
mtiI Insert 500_KI_rev	tggagaatgaaatcatcgggttacgcgtcaggcgatgc
mtiI Backbone_KI_fwd	cccgatgatttcattctc
mtiI Backbone_KI_rev	ataggatgcctcctgttc
AT adapt SCR F	ccgttggatgattaaatgcgc
AT adapt SCR R	gtgcacctcaaagccatcc
AT WT SCR F	ccaaactcgtactcctgagcat
AT WT SCR R	tctgaatgacgaccgtacgttt
hisI integrant F	gatctgcttttagggctggc
hisI integrant R	ttagccgcatctgtcttcca
M23 adapt2 SCR F	tgcttcttccaatatctctgcg
M23 adapt2 SCR R	tctgttggtcttatctggcact
M23 WT SCR F	tacatgctcttccactcgcttt
M23 WT SCR R	gcagatgccctttgaggaaaaa
metF 500 LHAF	ttttttgaacaattgacaatctggcaaccaaggttttaac
metF 500 LHAR	ataaaccagccttattcaccatccttttcctg
metF 500 RHAF	ggtgaataaggctggtttataggagtag

metF 500 RHAR	aacttgcccactggccggcctgatgacctttttccgtc
MetV 500 LHAF	ttttttgaacaattgacaaggcaaagatgttgtcattatc
MetV 500 LHAR	attcaccatcaatcatttttacttcctccg
MetV 500 RHAF	aaaaatgattgatggtgaataagatgagc
MetV 500 RHAR	aacttgcccactggccggccaaacaggggtaacacaag
TAKOS Vector_fwd	ggccggccagtgggcaag
TAKOS Vector_rev	ttgtcaattgttcaaaaaaataatggcggcgcg
MetV/F SCR F	cttcccatgtacagcagaaggt
MetV/F SCR R	tcggaatacagccacggtttaa
TraDIS 11 LHAF	ttttttgaacaattgacaaagcacctcgatgataatg
TraDIS 11 LHAR	cgtcttatttttttcatttgagtttcctcc
TraDIS 11 RHAF	tcaaatgaaaaaataagacgataaggccttccaaaaacc
TraDIS 11 RHAR	aacttgcccactggccggccctctgggcaatggcccgt
TraDIS 11 SCR F	cacgtttcaaacacatcgtcca
TraDIS 11 SCR R	agggctttctggatagtctcct
TraDIS 12 LHAF	ttttttgaacaattgacaaagctttttggcggctgag
TraDIS 12 LHAR	gctctcattctctcatttccgcacatttttcc
TraDIS 12 RHAF	ggaaatgagagaatgagagcagataaaatg
TraDIS 12 RHAR	aacttgcccactggccggccattaatctttttctccataatgg
TraDIS 12 SCR F	attgacctgccaaaatccaacg
TraDIS 12 SCR R	gggaagaacacatccacatcct
TraDIS 13 LHAF	ttttttgaacaattgacaagacggtgatccgggcaca
TraDIS 13 LHAR	ttaggtgaatatttaatggaaagtggtgaacatagatgagtttac
TraDIS 13 RHAF	tccattaaatattcacctaaatttcctcc
TraDIS 13 RHAR	aacttgcccactggccggccttctcggtaaagaaggtaaag
TraDIS 13 SCR F	cgtttgtcggaagtttctgcat
TraDIS 13 SCR R	aaaggcaccaagggtatctctg
TraDIS 14 LHAF	ttttttgaacaattgacaatatcaatccgaacgatctg
TraDIS 14 LHAR	ttgcttagcttttcatttttcctcctgc
TraDIS 14 RHAF	aaaaatgaaaagctaagcaaaaagatatacgc
TraDIS 14 RHAR	aacttgcccactggccggccaggcgtatcaaactcggtatag
TraDIS 14 SCR F	agatgaactgaaggtcctgctg
TraDIS 14 SCR R	aagccctccggattacaatgtt
TraDIS 15 LHAF	ttttttgaacaattgacaagcgtccgcctcaccattg
TraDIS 15 LHAR	tcggttaactgttcatgggtttcctcctgttttc

TraDIS 15 RHAF	acccatgaacagttaaccgaaaggagttatgatgaaaatattagaaatgc
TraDIS 15 RHAR	aacttgcccactggccggcctccagattgccggtgggc
TraDIS 15 SCR F	agcagtccaaggtcgatttcat
TraDIS 15 SCR R	cccaggatcgcgtacatttttc
TraDIS 16 LHAF	ttttttgaacaattgacaatgggagcaaacgcgcaac
TraDIS 16 LHAR	tggattattccatttcctccttttatatacgaaaattgattaaaagcattatatc
TraDIS 16 RHAF	ggaggaaatggaataatccatagaaaaaaacgg
TraDIS 16 RHAR	aacttgcccactggccggccgaataatgatcttacccatcg
TraDIS 16 SCR F	aaaaacgctccgggtaaaaagg
TraDIS 16 SCR R	ttctttgagattcgtcccaggg
TraDIS 17 LHAF	ttttttgaacaattgacaacccctacccccagggctt
TraDIS 17 LHAR	aaaaatggaagaatagagagtctgaatgacgaccgtacg
TraDIS 17 RHAF	ctctctattcttccatttttatgctccttcagc
TraDIS 17 RHAR	aacttgcccactggccggcccttcatcggctcactggc
TraDIS 17 SCR F	aaacgtacggtcgtcattcaga
TraDIS 17 SCR R	atgtggaaaaatggatcgggga
TraDIS 18 LHAF	ttttttgaacaattgacaaattgaggcgtctttttgac
TraDIS 18 LHAR	ttctctaaagactcatttttctgcctc
TraDIS 18 RHAF	aaaaatgagtctttagagaaagcgaacactagtgag
TraDIS 18 RHAR	aacttgcccactggccggccccccaagtgcaggcctattc
TraDIS 18 SCR F	tctgtgggaaaatctaggtgcc
TraDIS 18 SCR R	acgccagcccctcttaaaatta
TraDIS 19 LHAF	ttttttgaacaattgacaaattatcccatatacgggc
TraDIS 19 LHAR	gatgcttaatcttcaaaaataaacaccgc
TraDIS 19 RHAF	atttttgaagattaagcatcagaggaggaattatgaaagtaaaaaaag
TraDIS 19 RHAR	aacttgcccactggccggccgcacgccgatcacagagg
TraDIS 19 SCR F	agaaacagcggaatcagtggat
TraDIS 19 SCR R	ctggcacattgtcgaaatggtt
TraDIS 20 LHAF	ttttttgaacaattgacaatggagcaggtggagaatg
TraDIS 20 LHAR	ctgctcattcgttcacatttcctcctcatc
TraDIS 20 RHAF	aaatgtgaacgaatgagcaggataaccag
TraDIS 20 RHAR	aacttgcccactggccggcccttccatgaaatgccggc
TraDIS 20 SCR F	aaggtgcttaaagagagggagc
TraDIS 20 SCR R	gatccgcagaacatgaccattg
BCACT LHAF	ttttttgaacaattgacaaatcactgtcatatcatgc

BCACT LHAR	aaaaatgagtaactaaaacattcagggac
BCACT RHAF	tgttttagttactcatttttccctcccatttg
BCACT RHAR	aacttgcccactggccggcctgtactacaattttctgggaatc
BCACT SCR F	atagtggcctggtttcctgaag
BCACT SCR R	ctgtgagcatggaggaaagagt
3,173,000 SCR F	tttagctgtccctctggcaa
3,173,000 SCR R	ctttcgccggtttatgctga
act LHA SCR F	aggaaacgtcgaaagacg
act RHA SCR R	agctgataaagtataaggtcttcc
act LHA SCR F2	gccaacaagtaatttgcataca
act RHA SCR R2	aaaccatcaggagttttggggt
hisI LHA scr F	tatgagggcaccatcgatc
hisI RHA scr R	ggacaggtgctgtcaatctg
BCact LHA scr F	aggaaacgtcgaaagacg
BCact RHA scr R	agctgataaagtataaggtcttcc

8.2 Genomic alignments between adapted fermentation strains and their parent strain

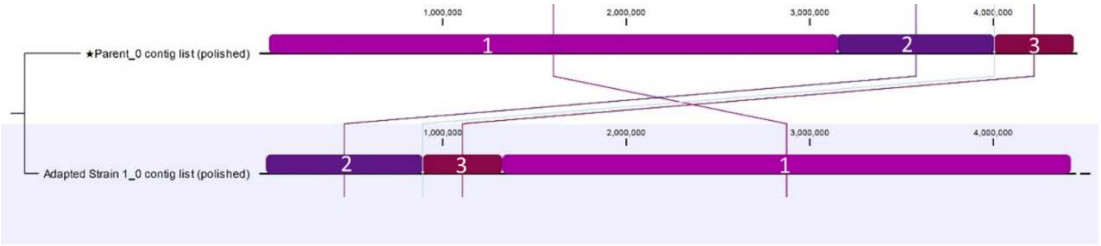

Figure 103 - Output of whole genome alignment between ASF01 ATCC 8486 and Parent ATCC 8486

The ATCC 8486 ASF01 (on the bottom) aligned against the parent ATCC 8486 (on top). The blocks of colour represent different regions and serve to distinguish regions which may have moved. The apparent repositioning of blocks 2 and 3 of the parent ATCC 8486 to the left of the ATCC 8486 ASF01 block 1 is mostly likely due to a difference in the start of the de novo sequence and would not account for a rearrangement once the circular nature of the chromosome is accounted for.

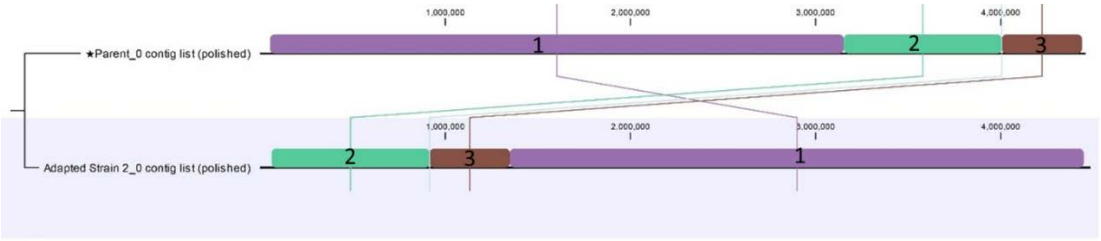

Figure 104 - Output of whole genome alignment between ASF02 ATCC 8486 and Parent ATCC 8486

The ATCC 8486 ASF02 (on the bottom) aligned against the parent ATCC 8486 (on top). The blocks of colour represent different regions and serve to distinguish regions which may have moved. The apparent repositing of blocks 2 and 3 of the parent ATCC 8486 to the left of the ATCC 8486 ASF02 block 1 is mostly likely due to a difference in the start of the de novo sequence and would not account for a rearrangement once the circular nature of the chromosome is accounted for.

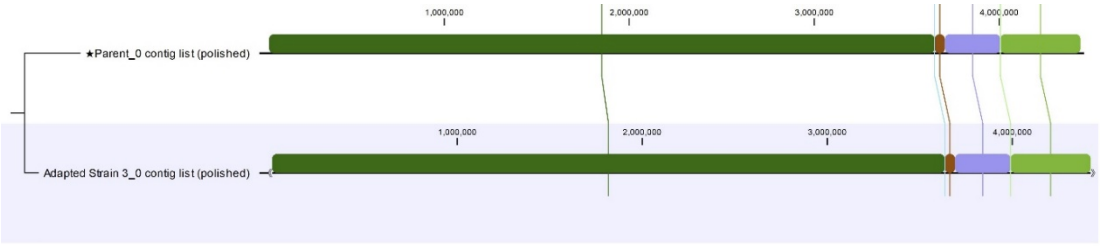

Figure 105 - Output of whole genome alignment between ASF03 ATCC 8486 and Parent ATCC 8486

The ATCC 8486 ASF03 (on the bottom) aligned against the parent ATCC 8486 (on top). The blocks of colour represent different regions and serve to distinguish regions which may have moved. No rearrangements occurred.

Table 27 - Genes in the 12 kb region which moved from the parent position 1873248 – 1885522

Genes (KAIST locus tags)	**Putative gene function**
B2M23_RS21185*	Helix-turn-helix transcriptional regulator
B2M23_RS11135	Recombinase family protein
B2M23_RS11140	DUF4041 domain-containing protein
B2M23_RS11145	S24 family peptidase
B2M23_RS11150	Helix-turn-helix transcriptional regulator
B2M23_RS11155	Rha family transcriptional regulator
B2M23_RS21190	Hypothetical protein
B2M23_RS11160	Helix-turn-helix transcriptional regulator
B2M23_RS21195	Hypothetical protein
B2M23_RS21200	Hypothetical protein
B2M23_RS21205	Hypothetical protein
B2M23_RS11165	Hypothetical protein
B2M23_RS11170	Siphovirus Gp157 family protein
B2M23_RS21445	Hypothetical protein
B2M23_RS11180	Single-stranded DNA-binding protein
B2M23_RS11185	Hypothetical protein
B2M23_RS11190	Replicative helicase loader/inhibitor
B2M23_RS11195	Hypothetical protein
B2M23_RS11200	Hypothetical protein
B2M23_RS11205	Hypothetical protein
B2M23_RS11210	Hypothetical protein
B2M23_RS11215	Hypothetical protein
B2M23_RS11220	Hypothetical protein
B2M23_RS11225*	Hypothetical protein

*Locus tags provided are that of the KAIST strain (283) though comparison was performed against the ATCC 8486 parent progenitor strain. Gene function is assigned as a combination of BLASTp data, PHASTER (372) and KAIST annotations. All genes are within this region except for those marked with * which are flanking sequences.*

8.3 Further butyrate production mutant confirmation

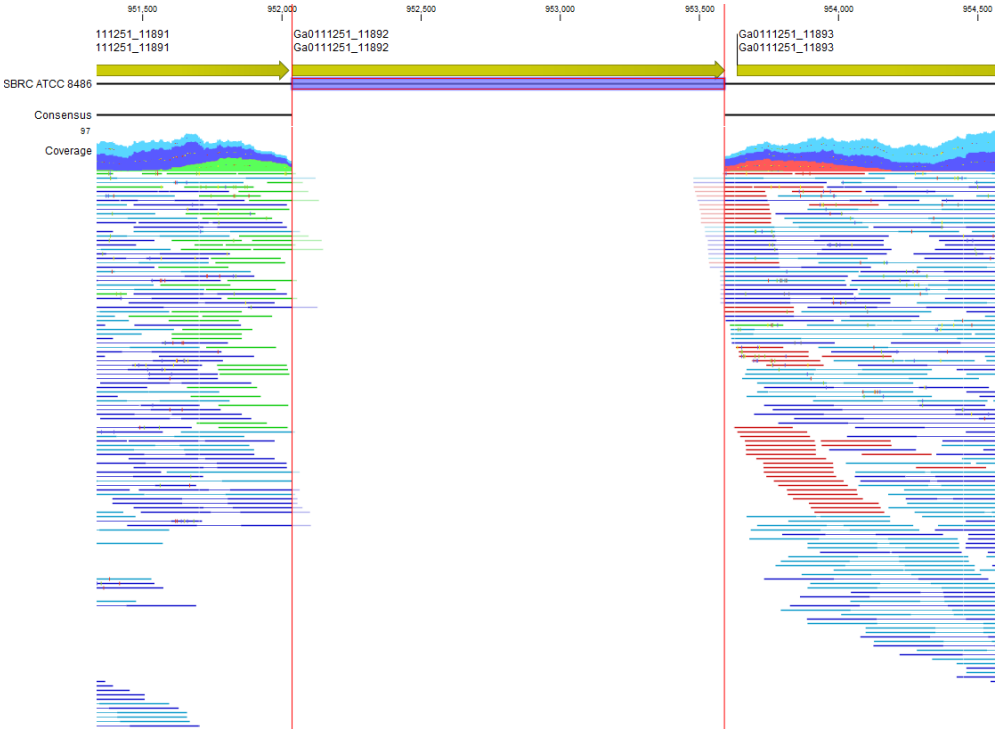

Figure 106 – Navigation area of CLC genomics workbench 21 displaying act gene region in purported Δact mutant
No aligned reads exist between the act gene against purported Δact mutant, either at the site of knockout or elsewhere in the chromosome.

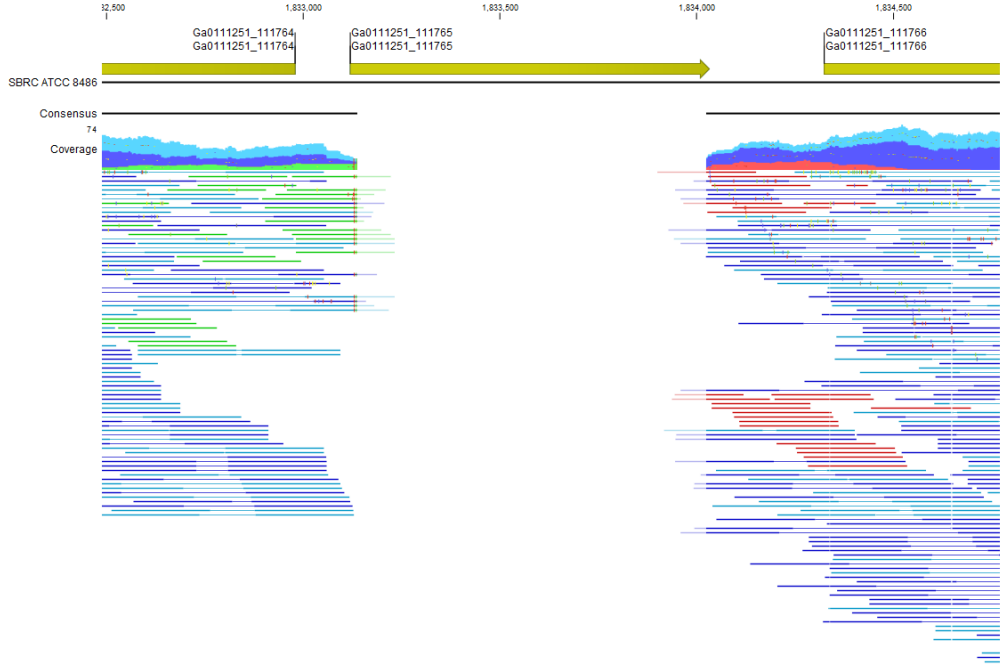

Figure 107 – Navigation area of CLC genomics workbench 21 displaying pbt1 gene region in purported Δpbt1 mutant.
No aligned reads exist between the pbt1 gene against purported Δpbt1 mutant, either at the site of knockout or elsewhere in the chromosome.

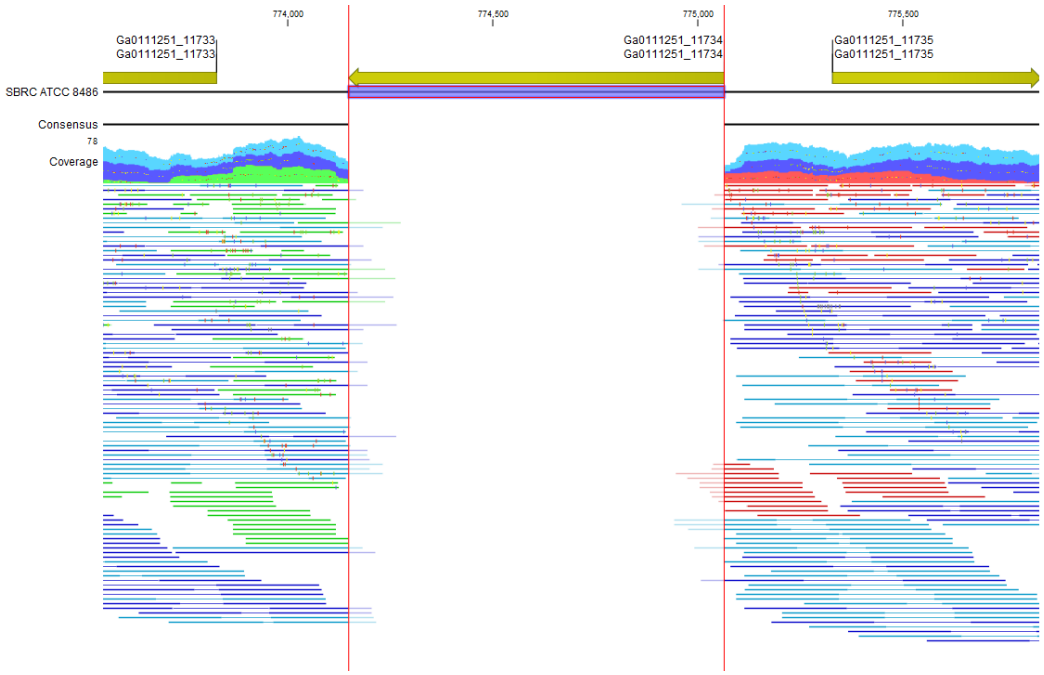

Figure 108 – Navigation area of CLC genomics workbench 21 displaying pbt2 gene region in purported Δpbt2 mutant.
No aligned reads exist between the pbt2 gene against purported Δpbt2 mutant, either at the site of knockout or elsewhere in the chromosome.

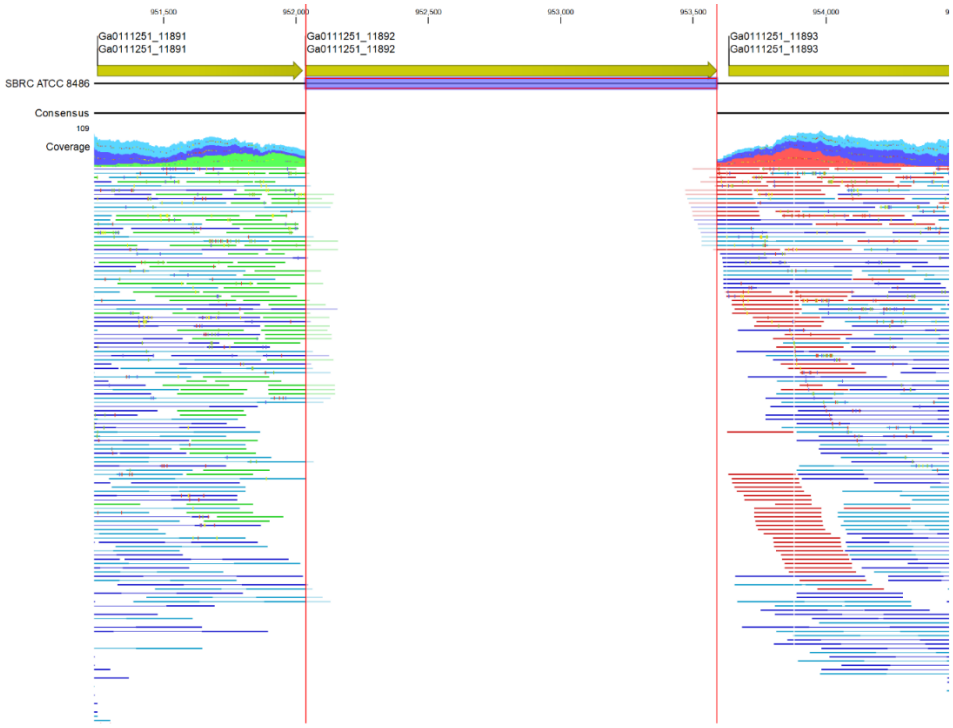

Figure 109 – Navigation area of CLC genomics workbench 21 displaying act gene region in purported ΔactΔpbt1 mutant.
No aligned reads exist between the act gene against purported Δact mutant, either at the site of knockout or elsewhere in the chromosome.

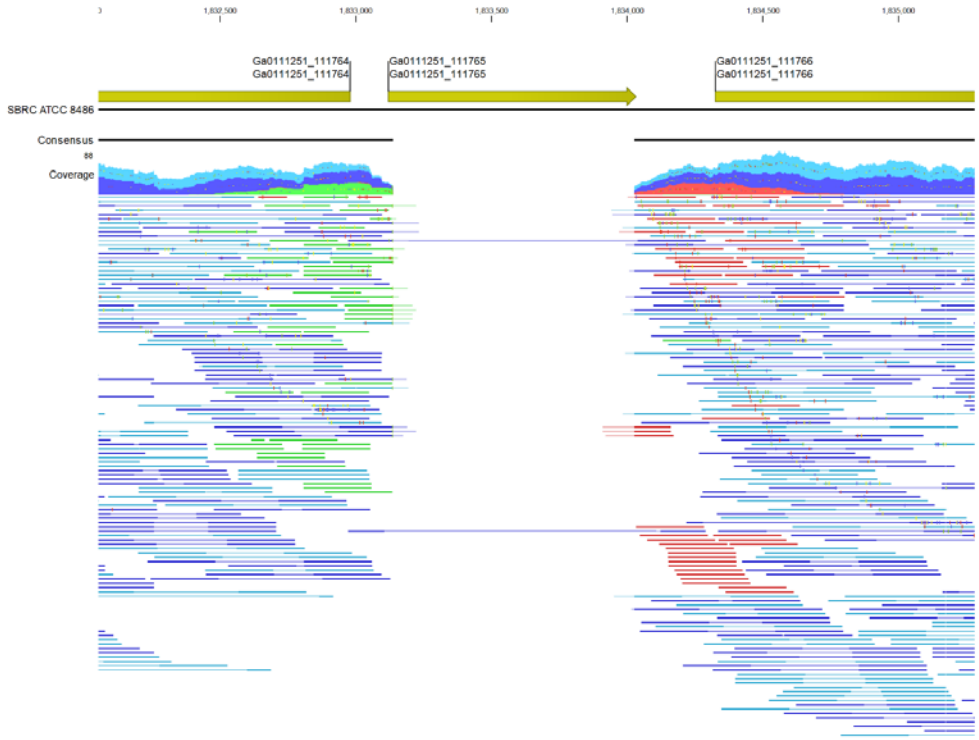

Figure 110 – Navigation area of CLC genomics workbench 21 displaying pbt1 gene region in purported ΔactΔpbt1 mutant.
Two apparent alignments exist across this region, these were investigated further, with no alignment observed. This indicated no alignments exist for pbt1 in the ΔactΔpbt1 mutant.

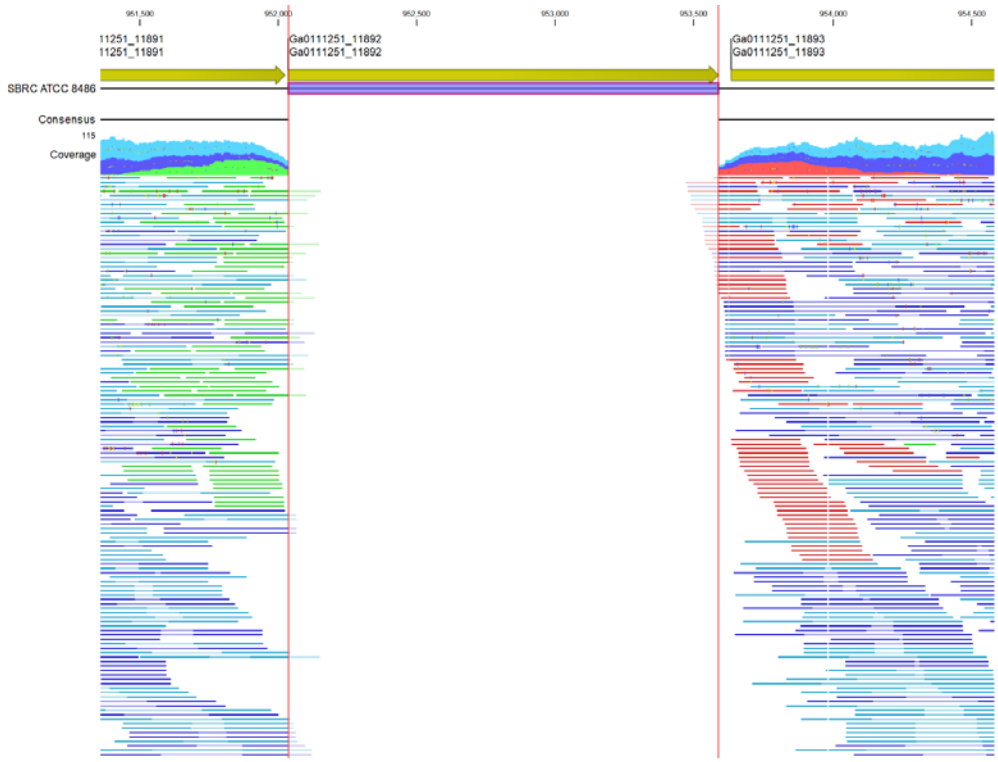

Figure 111 – Navigation area of CLC genomics workbench 21 displaying act gene region in purported ΔactΔpbt2 mutant.
No aligned reads exist between the act gene against purported ΔactΔpbt2 mutant, either at the site of knockout or elsewhere in the chromosome.

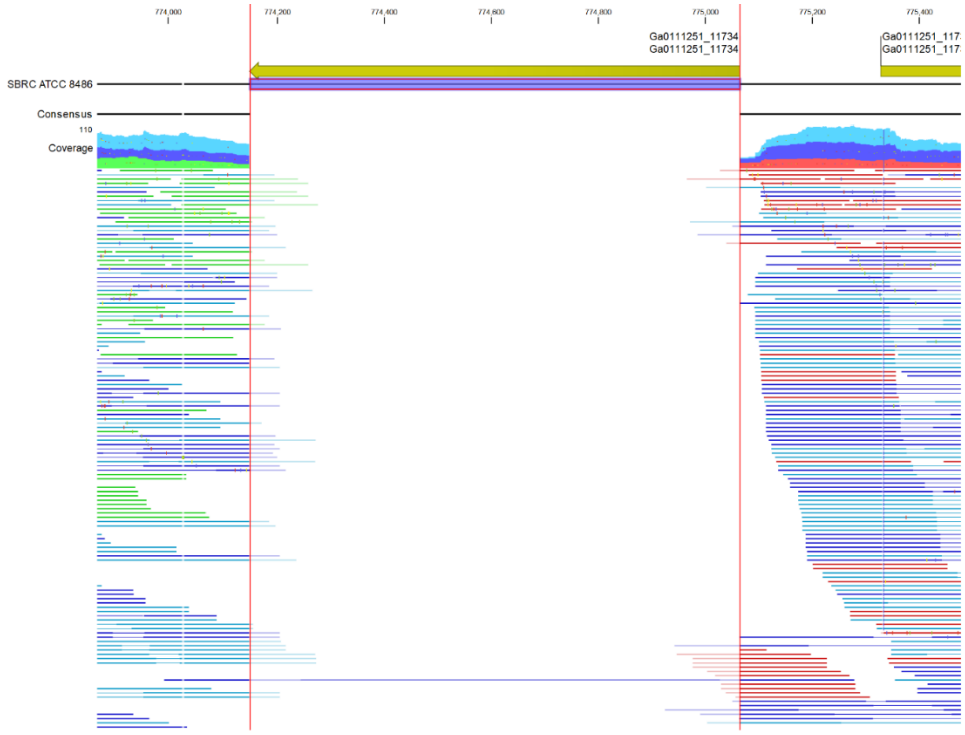

Figure 112 – Navigation area of CLC genomics workbench 21 displaying pbt2 gene region in purported ΔactΔpbt2 mutant.
No aligned reads exist between the pbt2 gene against purported ΔactΔpbt2 mutant, either at the site of knockout or elsewhere in the chromosome.

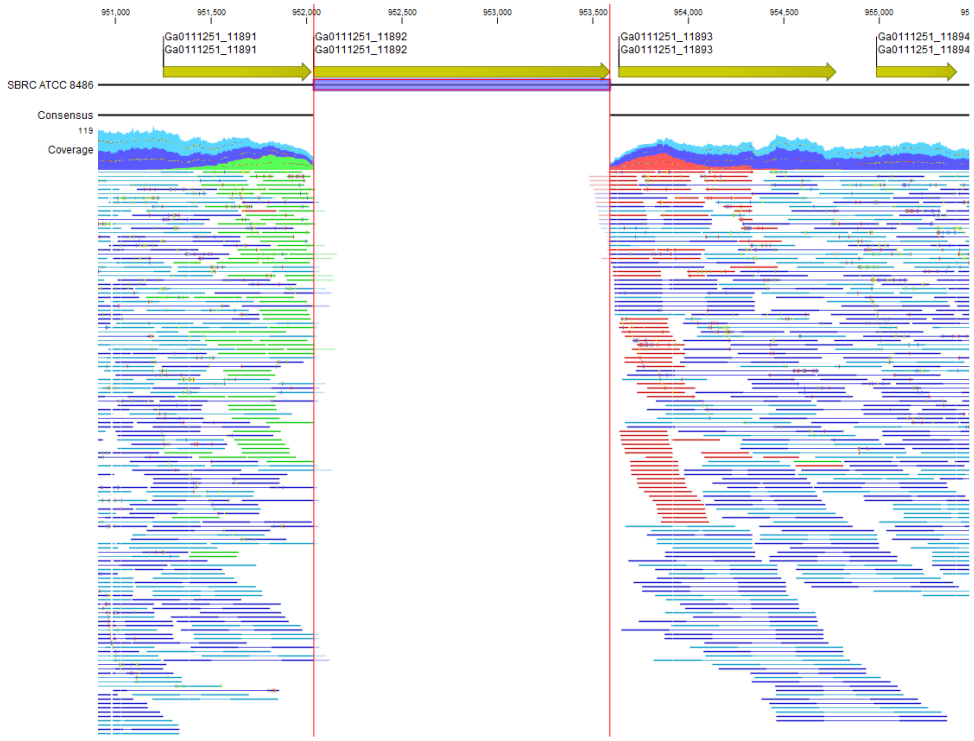

Figure 113 – Navigation area of CLC genomics workbench 21 displaying act gene region in purported ΔactΔpbt1Δpbt2 mutant.
No aligned reads exist between the act gene against purported ΔactΔpbt1Δpbt2 mutant, either at the site of knockout or elsewhere in the chromosome.

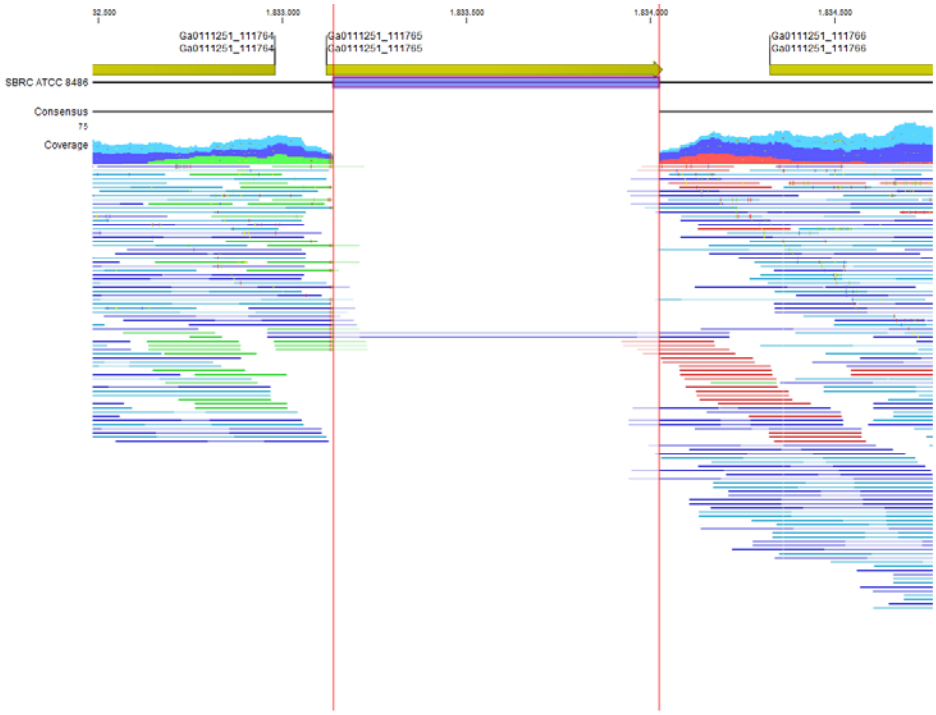

Figure 114 – Navigation area of CLC genomics workbench 21 displaying pbt1 gene region in purported ΔactΔpbt1Δpbt2 mutants.
Two lines of across the region, presumably two reads, were observed. Given the lack of a consensus across these regions a manual analysis was performed to investigate purported alignments (Figure 115).

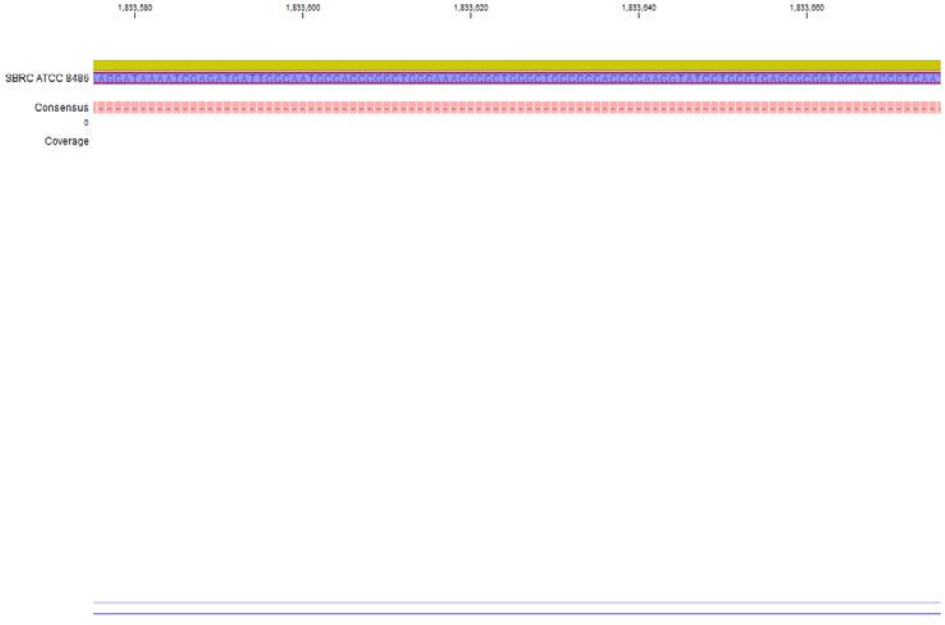

Figure 115 – Close up image of navigation area of CLC genomics workbench 21 displaying pbt1 gene region in purported ΔactΔpbt1Δpbt2 mutant.
The lack of any bases being displayed with these two blue lines, as occurs at this scale at anywhere there are reads suggests there is no actual alignment in the ΔactΔpbt1Δpbt2 mutant background, which suggests pbt1 is in fact knocked out.

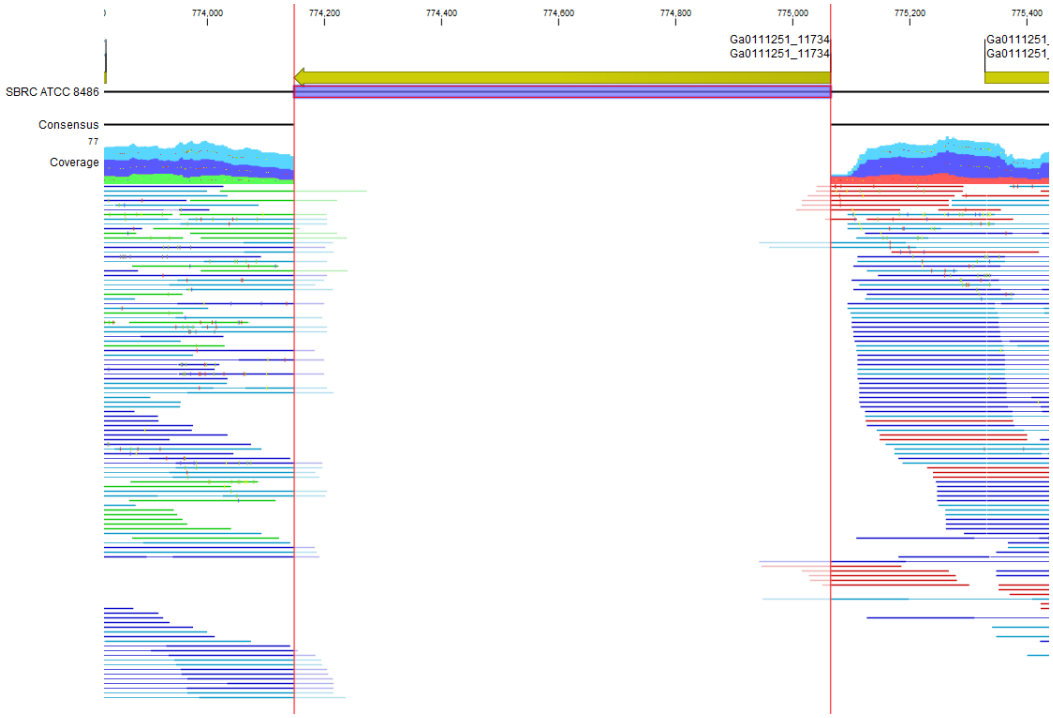

Figure 116 – Navigation area of CLC genomics workbench 21 displaying pbt2 gene region in purported ΔactΔpbt1Δpbt2 mutant.
No aligned reads exist between the pbt2 gene against purported ΔactΔpbt1Δpbt2 mutant, either at the site of knockout or elsewhere in the chromosome.

8.4 Δ*bcact*

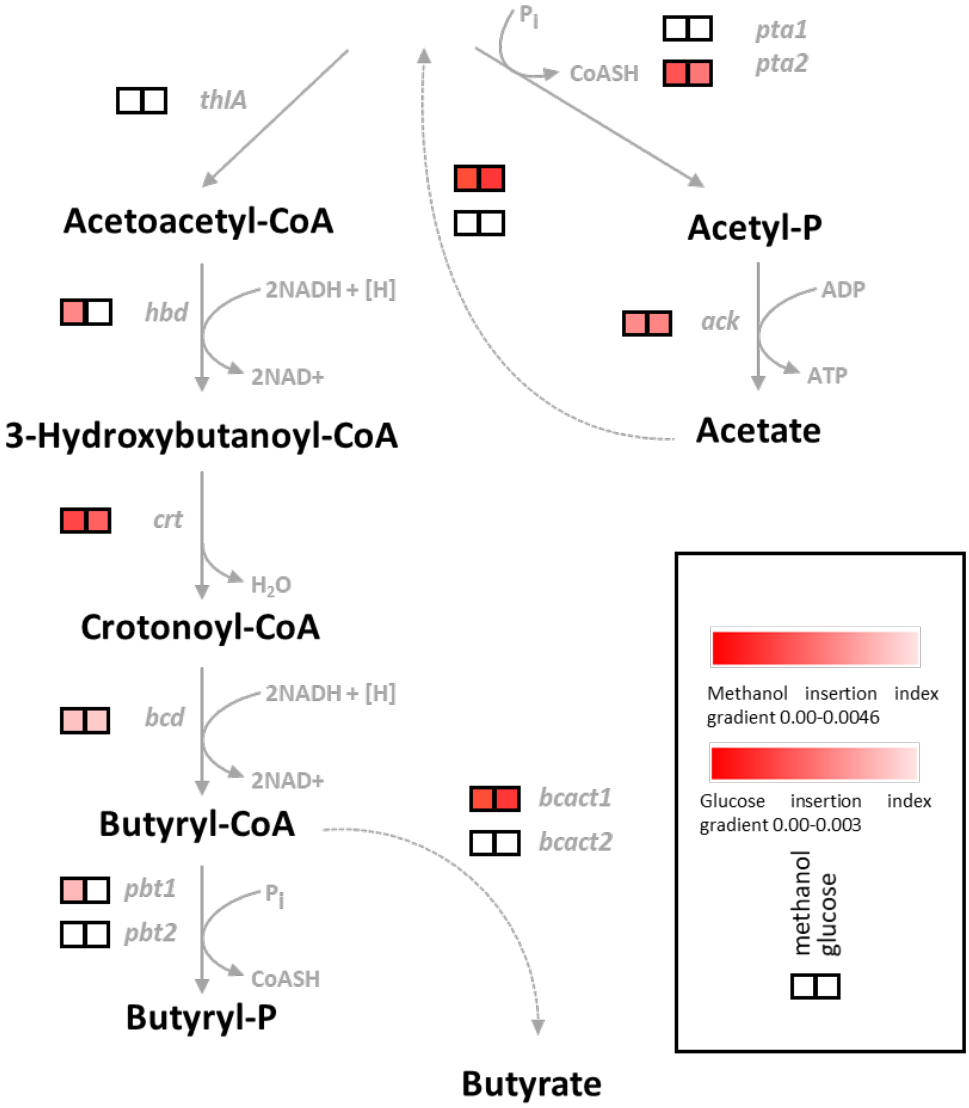

Figure 117 - Butyrate pathway displaying gene essentiality (Dr. Elizabeth Redfern, University of Nottingham, personal communication)
Figure displays the butyrate production pathway in E. limosum. Any genes with a white box are essential on the respective carbon source. bcact1 is act (4.3.2.1), bcact2 is the bcact that was mentioned in 4.3.4.8. This data suggests that bcact is essential on both methanol and glucose and thus it should be impossible to knockout. This data also suggests that Δpbt1 and 2 is not viable on glucose, which is counter to the evidence of 4.3.4.2.5 and 4.3.4.3.