Clostridial strain degeneration: new approaches to an old problem

Jonathan Humphreys, BSc

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy
September 2018

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.

Jonathan Humphreys

September 2018

Abstract

The genus *Clostridium* comprises a range of physiologically diverse species including notorious pathogens as well as industrially relevant organisms. Solventogenic *Clostridia* produce solvents via their unique acetone-butanol-ethanol (ABE) fermentation metabolism and encompass species of significance for future biofuel production. Typically these organisms undergo an initial acidogenic phase which is followed by a solventogenic phase as they enter stationary growth. The later phase also sees a morphological change in these organisms as they accumulate storage compounds and endospores. With an increased need for sustainable fuel production, solventogenic *Clostridia* offer an attractive alternative to produce biobutanol. An intrinsic problem with these organisms however is the phenomenon of spontaneous strain degeneration. This phenomenon is characterised by a reduced or complete loss in the production of solvents and the formation of spores which typically occurs when the bacteria are repeatedly subcultured in batch culture or grown in continuous fermentations.

The aim of this study was to gain a further understanding of the degeneration process by investigating the genetic, ecological and evolutionary forces behind this phenomenon. To do this, *Clostridium beijerinckii* NCIMB 8052 was selected as a model organism due to its tendency to degenerate more rapidly than other strains. Furthermore, this strain was progenitor to a hyper-butanol producing strain making it a relevant candidate for industrial use.

Through repeated subculturing of *C. beijerinckii* NCIMB 8052, four distinct colony morphologies emerged. A total of 71 degenerate isolates were obtained based on their colony appearance. These were phenotypically characterised in regards to their

solvent, spore and granulose producing capabilities. The most severe of the colony types was unable to produce solvents or endospores.

Genetic comparisons between derived isolates and their parental ancestor revealed four regions that contained more mutations than anywhere else in the genome. These regions included the master regulator gene of solvent and spore formation *spo0A* and two notable histidine kinases. The role of these genes was proven via insertional disruption which showed a huge reduction in sporulation capacity.

Social interactions were studied by mixing degenerate isolates with the wild type in various starting frequencies. At low frequencies, isolates were much fitter than the wild type however this fitness rapidly declined as the starting frequencies increased. This frequency-dependent fitness is suggestive of social cheating.

These findings led to the hypothesis that degenerate isolates have gained an increased fitness compared to the wild type and this fitness is due to the loss of Spo0A activity.

Acknowledgements

First and foremost I would like thank my primary supervisor Dr Klaus Winzer for not only his help, support and advice throughout the course of this PhD but for also initially explaining to me the wrong project which led to the work completed here. Furthermore, I'd like to thank my co-supervisors Dr Steve Diggle and Dr Alan Cockayne for their guidance during this PhD. I would also like to extend a thank you to all the CBS staff including Matt Abbott who helped a lot with sample analysis. Thanks to everyone in CRG/SBRC for their help as well. This work was made possible by the funding from the University of Nottingham's BBSRC Doctoral Training Partnership (DTP) and so a huge thanks to them.

For making the past four years so enjoyable, a big thank you goes to Pawel, Ollie, Zak and Alex for not only the pub nights, coffees and lunches but listening to all my questions and giving solid advice throughout the PhD. The Trent lads Dan, Chuckles, Strongers, Vijay and Raj all deserve a mention for all the night outs, hangover recoveries and all the great laughs.

To Emma, thank you for putting up with all the late nights in the lab, long weekends and for all your support since our chance encounter in Lithuania. I am also truly grateful to my Mum Di and Brother Sam who have helped me and supported me throughout the last 4 years and for being the best family I could ask for. A final and most important thank you goes to my grandparents Richard and Peggy for without them, none of this work would have been possible.

Table of contents

List of Figures

List of Tables

Abbreviations

µF	Micro farads
µg	Microgram
µl	Microliter
5-FC	5-Fluorocytosine
5-FOA	5-Fluoroorotic acid
ABE	Acetone butanol ethanol
ACE	Allelic coupled exchange
ADP	Adenosine diphosphate
Agr	Accessory gene regulator
ATCC	American type culture collection
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
Bp	Base pair
Cas	CRISPR associated proteins
CBM-S	Clostridial basal medium – sporulation
cDNA	Complementary DNA
CFU	Colony forming unit
CGM	Clostridial growth medium

CoA	Coenzyme A
CRISPR	Clustered regularly interspaced short palindromic repeats
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EPB	Electroporation buffer
Erm	Erythromycin
g	gravity
GC	Gas chromatography
gDNA	Genomic DNA
HPLC	High pressure liquid chromatography
HTH	Helix-turn-helix
InDel	Insertion/deletion
kJ	Kilo Joules
kV	Kilo volts
LB	Lysogeny Broth
LHA	Left homology Arm
mg	Milligram
ml	Millilitre
mm	Millimetre

mM	Millimolar
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Centre for Biotechnology information
NCIMB	National Collection of Industrial, Food and Marine Bacteria
NGS	Next Generation Sequencing
nm	Nanometres
OD	Optical density
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
RAM	Retro-transposition activated marker
RHA	Right homology arm
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RPM	Rotations per minute
rRNA	Ribosomal ribonucleic acid
SASPS	Small acid soluble proteins
SNV	Small Nucleotide Variant
TAE	Tris acetate EDTA

UV	Ultra violet
V	Volt

1. Introduction

1.1. The need for alternative fuel sources

1.1.1. Fossil fuels

Statistics from 2015 showed that fossil fuels contribute to around 80% of the total fuel source globally (International Energy Agency, 2015). These fuel sources include crude oil, coal and natural gas and occur from the decomposition of ancient organisms over millions of years (Ourisson et al., 1984). Although these fuels represent the majority of our energy production, their abundance is finite (Hughes and Rudolph, 2011; Shafiee and Topal, 2009). Predictions on our overreliance on these fuels sees oil, coal and gas exhausted in approximately 35, 107 and 37 years (Shafiee and Topal, 2009) or 50, 114 and 52 years (BP, 2016), respectively, depending on the source. Furthermore, the burning of fossil fuels contributes to the release of huge amounts of greenhouse gases into the environment. These include the gases carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) which have been shown to cause global temperatures to rise (Dentener et al., 2007; James Hansen et al., 2000).

The increase in global temperature has been associated with several detrimental effects on the planet. These include but are not limited to arctic sea ice retreat, increases in heavy rainfall and flooding, permafrost melting, loss of glaciers and snowpack, increased intensity of hurricanes and severe droughts (James Hansen et al., 2000; Karl and Trenberth, 2003; Solomon et al., 2009). Moreover, natural carbon sinks such as the rainforest and the ocean are being lost due to increased deforestation and the acidification of the ocean which has further exacerbated the harmful effects of global warming (Geider et al., 2001; Hoegh-Guldberg et al., 2007; Karl and Trenberth, 2003).

Due to the inherent problematic nature of fossil fuels and our high energy demand, alternative fuel sources are required. This has led to a concerted effort to find other fuel sources that are both sustainable and environmentally friendly to allow the eventual replacement of fossil resources (Mabee et al., 2005).

1.1.2. Biofuels

1.1.2.1. First generation biofuels

Biofuels encompass liquid or gas fuels that are predominately produced from biomass. Renewable sources of biomass include plant waste biomass, vegetable oils, sugar beets and wheat (Demirbas, 2008). Biofuels represent several advantages over fossil fuels as they are biodegradable, sustainable, carbon neutral, can help reduce our reliance on fossil fuels and have a greater potential to be environmentally friendly (Demirbas, 2008; Naik et al., 2010; Puppán, 2002). First generation biofuels include bioethanol, biodiesel and biogas and are characterised by their tendency to be blended with traditional fuels (Naik et al., 2010). Bioethanol is formed from the fermentation of sugars and starches, biodiesel from transesterification of vegetable oils, residual oils and fats and biogas, usually biomethane, is acquired via anaerobic digestion of manure. All of these fuels do require modifications to existing engines if they are to be used in traditional cars; however several flexi-fuel vehicles (which use gasoline/ethanol/methanol/gas mixtures) do exist especially in countries like Brazil where over 50% of car sales are these types of vehicles (Hira and Guilherme De Oliveira, 2009; Naik et al., 2010).

Although the use of first generation biofuels might be regarded as preferable over fossil fuels, there are several limitations of this technology. The main disadvantage is the need for extensive amounts of land to grow biofuel feedstocks that would otherwise be used to grow crops. This has led to the 'fuel vs food' debate in which many people argue that traditional biofuel production has caused food prices to rise (Mitchell, 2008; Schmidhuber, 2007; Valerie Mercer-Blackman, Hossein Samiei, 2007). Furthermore, the need to clear large areas of land for biofuel feedstocks can lead to larger carbon debt as emissions from burning land outweighs the ones saved from biofuels (Fargione et al., 2008; Searchinger et al., 2008). Moreover, the effects of damaging natural ecosystems and the extensive use of fertilisers poses a further problem when creating first generation biofuels (Mohr and Raman, 2013).

1.1.2.2. Second generation biofuels

There has now been a growing shift to the production of second generation biofuels which primarily utilise lignocellulosic material from waste plant material (Gomez et al., 2008; Naik et al., 2010). Lignocellulosic material is comprised of lignin, hemicellulose and cellulose which make up around 75% of plant material and has a huge potential for biofuel conversion (Chaturvedi and Verma, 2013; Gomez et al., 2008; Pauly and Keegstra, 2008). Potential sources of lignocellulosic material include by-products such as cereal straw, sugar cane bagasse, forest residues and waste materials such as organic components of municipal solid wastes (Sims et al., 2010).

There are two main routes for converting plant biomass into fuel, thermochemical processing and biochemical processing. Thermochemical processing utilises extreme temperatures and various levels of oxygen to define the end product. Pyrolysis occurs

when oxygen is not present and results in either solid (charcoal) or liquid (bio-oil) production which can be used for power generation. With the addition of low concentrations of oxygen, gasification will occur which sees the production of synthesis gas (syngas), a mixture of hydrogen and organic gases (CO, CO_2 and CH_4). Syngas can subsequently be converted to liquid fuel by the Fischer–Tropsch process or can be utilised by acetogenic bacteria for the production of bioethanol or fine chemicals. The disadvantage of the thermochemical process is the high costs associated with maintaining high temperatures and potential impurities found in syngas (Gomez et al., 2008; Humphreys et al., 2017; Naik et al., 2010; Ragsdale and Pierce, 2008; Zeisler et al., 2010)

The biochemical process converts the plant biomass into fermentable sugars for conversion into alcohols such as ethanol and butanol (Gomez et al., 2008). Several steps are first required until fermentable sugars are liberated. First is pre-treating the biomass for the removal of lignin as this impedes sugar hydrolysis (Chaturvedi and Verma, 2013; Naik et al., 2010). This is usually achieved via high pressure steam which breaks the lignin from the hemicellulose (Naik et al., 2010). Following lignin removal, cellulose and hemicellulose must then be hydrolysed by saccharification. This treatment step is usually achieved by acid or enzymatic break down to release the fermentable sugars (Alvira et al., 2010; Gomez et al., 2008). Although treatment is necessary, several issues may arise as these steps can produce toxic compounds and inhibitors (Jönsson et al., 2013). Once liberated, sugars are converted to alcohols via microbial fermentation similar to that found for first generation biofuels (Saini et al., 2015).

The abundance of lignocellulosic material makes the production of second generation biofuels more desirable than first generation (Mabee et al., 2011). Production of

bioethanol has been the primary focus of biofuel research and industry as many microbes easily ferment sugars to ethanol (Saini et al., 2015). There are many other fuel sources including the potentially far superior biobutanol.

1.1.3. Butanol

Butanol is a four carbon alcohol which can exist in several isomers depending whether it is in the straight-chained or branched structure (Jin et al., 2011). The most common form used as a biofuel is 1-butanol (n-butanol) which represents the straight-chained variety. Although bioethanol represents the majority of biofuel production, butanol offers several advantages over the traditional fuels (Köpke et al., 2011). Due to both a higher energy density and air-fuel ratio, butanol is able to provide greater mileage and power when used as a transportation fuel (Jin et al., 2011; Köpke et al., 2011). This further reduces consumption and the amount of undesirable by-products compared to ethanol. A lower vapour pressure means that butanol is both safer to handle and more easily ignited in cold weather compared to ethanol. Furthermore, ethanol is far more corrosive than butanol and requires transportation outside of existing pipelines (Jin et al., 2011; Köpke et al., 2011). Another significant advantage of butanol over ethanol is its ability to be used directly in vehicles without the need for modifications or blending (Szulczyk, 2010).

Although butanol does offer several advantages over ethanol, the energy yield from substrate to solvent is around half that of ethanol when produced via microbial fermentation (Pfromm et al., 2010). At present, production of butanol is more costly than ethanol and requires significant optimisation. Microbial production of butanol is however well established and has been achieved for over a 100 years. The main

sources of microbial biobutanol are from ABE (acetone-butanol-ethanol) fermentation mediated by members of the genus *Clostridium* (Jones and Woods, 1986).

1.2. The genus Clostridium

The genus *Clostridium* encompasses a wide variety of Gram-staining positive, rod shaped bacteria. Members of this genus are obligate anaerobes, able to form endospores and unable to reduce sulphate (Dürre, 2005; Lawson et al., 1993; Ludwig et al., 2009). Furthermore, members of the genus usually have a G+C genomic DNA content that is very low ranging from 22–35 mol% (Ludwig et al., 2009). Further characterisation of this genus has been achieved through the use of comparative 16S rRNA sequence analysis which revealed several phylogenetic lineages (Lawson et al., 1993). Several notable pathogenic species exists in the genus which include *Clostridium botulinum* which produces a deadly paralysis toxin (Münchau and Bhatia, 2000), *Clostridium perfringens* which causes gangrene (Rood and Cole, 1991), *Clostridium tetani* which causes tetanus (Cook et al., 2001) and *Clostridium difficile* which cause excessive diarrhoea (Poxton et al., 2001). It should be noted that comparative genomics of *C. difficile* with other members of the genus *Clostridium* show that this species should be reclassified as a *Peptoclostridium* (Yutin and Galperin, 2013).

Several important non-pathogenic species also exist in the genus which are of both medical and industrial significance. *C. sporogenes* has been proposed as an alternative for cancer treatment as spores will germinate in hypoxic tumours leading to lysis (Heap et al., 2014). Solvent producing species represent a diverse group of bacteria

that are able to ferment sugars into alcohols (Jones and Woods, 1986). Species of note include *Clostridium acetobutylicum, Clostridium beijerinckii, Clostridium saccharobutylicum* and *Clostridium saccharoperbutylacetonicum* (Poehlein et al., 2017). Many of these species ferment sugars through ABE fermentation and have been utilised industrially for many years.

1.3. ABE fermentation

1.3.1. History of ABE fermentation

Solvent production by microbial fermentation was first described by Louis Pasteur when he was able to produce butanol by microbial fermentation. This was subsequently followed by Franz Schardinger who could produce acetone in a similar fashion. Industrial scale production of solvents was pioneered by Chaim Weizmann who isolated several solvent producing strains whilst working at the University of Manchester. The most effective strain he named *Clostridium acetobutylicum* which was able to ferment starches with high solvent yields and was originally intended to create synthetic rubber (Jones and Woods, 1986). The production of solvents rapidly changed with the advent of the First World War. Previously, the British Army required cordite, a smokeless powder, for their munitions which was originally produced from acetone acquired from Germany, Austria and the USA. The onset of the war meant alternative sources were needed. Weizmann was instructed by the British government to find alternative methods for acetone production which led to the utilisation of microbial ABE fermentation. Many years later after the Second World

War, Weizmann became the first president of Israel after he refused personal honour or reward for his work during the First World War (Jones and Woods, 1986).

Butanol at the time was seen as an undesirable product until it was realised that it could be used for car lacquer for the ever growing automotive industry post the First World War (Jones and Woods, 1986). The industry grew rapidly with several fermentation plants found globally in countries such as the USA, UK, Japan, China, Russia and South Africa. Although production of solvents by microbial fermentation was originally high, by the 1950s the advent of the petrochemical industry saw their use decline (Jones and Woods, 1986). The last commercial ABE plant closed in South Africa in 1983 due to a shortage of sugar feedstock resulting from drought. The concerns of dwindling fossil fuel reserves has now led to a renewed shift towards utilising ABE fermentation once again. Since 2006, several Chinese fermentation plants have begun ABE fermentation for solvent production (Ni and Sun, 2009) but the majority are not currently operational due to low oil prices (Jiang et al., 2015).

1.3.2. ABE metabolism

A distinct feature of solventogenic clostridial ABE fermentation is the organisms' ability to utilise a biphasic metabolism. The first acidogenic phase occurs when the cells are growing exponentially and sees the production of the organic acids acetate and butyrate. The eventual accumulation of acids lowers the pH which causes a transition into the solventogenic phase which sees the production of organic solvents. The solventogenic phase sees acetate and butyrate reassimilated and converted to acetone, butanol and ethanol which for some species, such as *C. acetobutylicum*,

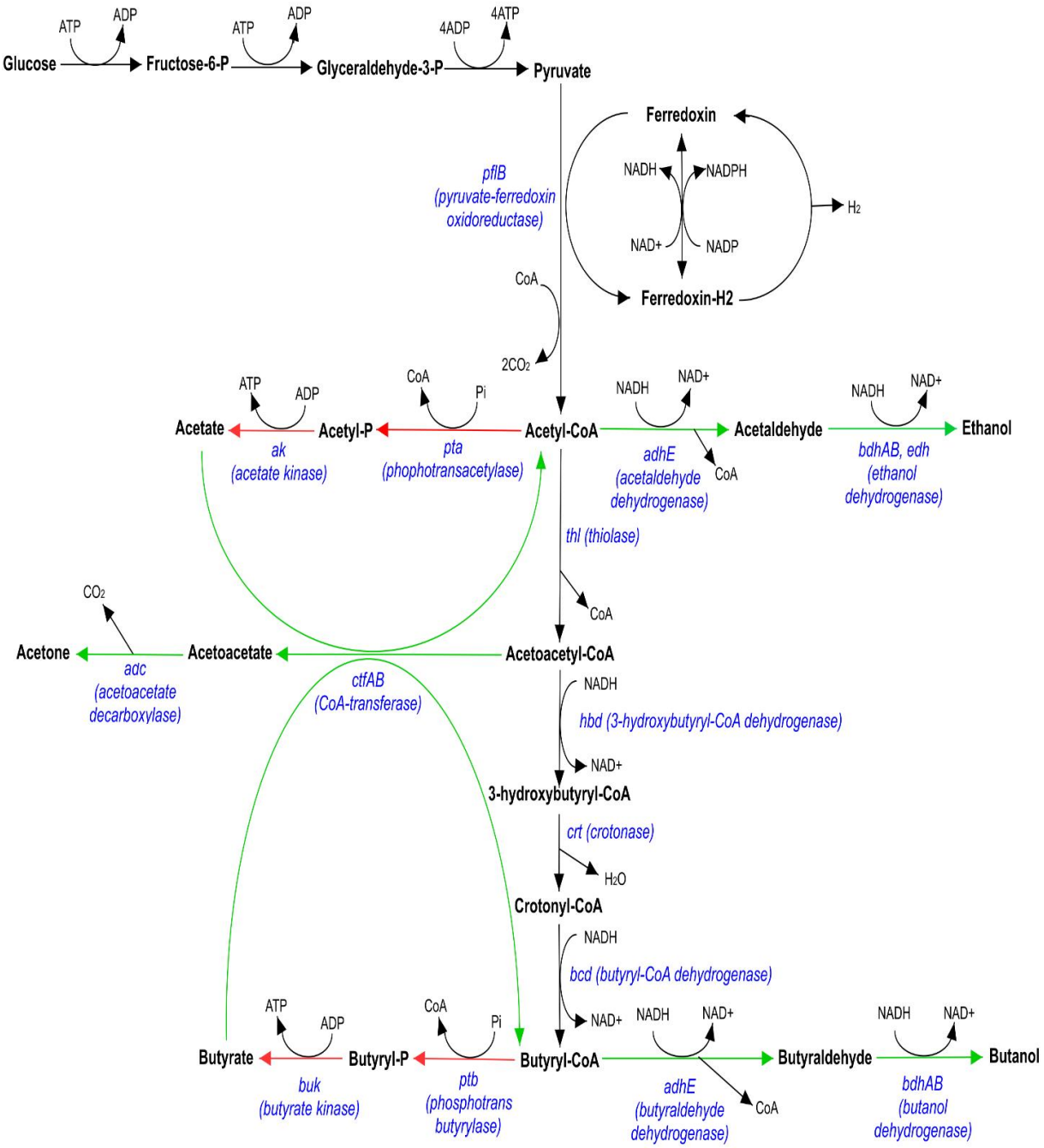


Figure 1.1 A summary of the acetone-butanol-ethanol (ABE) fermentation from *C. acetobutylicum*.

Metabolism of *C. acetobutylicum* adapted from Lee et al., 2008. Blue indicates the enzymes involved in the fermentation pathway and black indicates the products at each stage of the fermentation. Red lines indicate the acidogenic stage of the pathway and green shows the solventogenic phase of the pathway.

appears to be associated with entry into the stationary growth phase (Jones and Woods, 1986; Köpke et al., 2011; Lee et al., 2008b). This phase also sees physiological changes such as spore formation (Dürre, 2014; Jones and Woods, 1986). A summary of the ABE pathway can be seen in Figure 1.1.

Clostridial growth begins with either metabolism of hexose sugars via the Embden-Meyerhof pathway or pentose sugars via the pentose phosphate pathway. Metabolism of hexose sugars leads to 2 pyruvate from 1 molecule of hexose with a net production of 2 ATP and 2 NADH. The pentose pathway is associated with the generation of 5 ATP and 5 NADH from 3 moles of pentose (Jones and Woods, 1986; Köpke et al., 2011).

Pyruvate resulting from the glycolysis of these sugars is cleaved by pyruvate ferredoxin oxidoreductase in the presence of coenzyme A (CoA) to yield carbon dioxide, acetyl-CoA and reduced ferredoxin. Ferredoxin plays a key role in electron transfer and distribution. Many clostridial species use ferredoxin to transfer electrons to an iron containing hydrogenase to produce hydrogen gas using protons as final acceptor and thus regenerating ferredoxin (Jones and Woods, 1986; Lütke-Eversloh and Bahl, 2011). Two other key enzymes in the electron distribution were originally considered to be the NADH ferredoxin oxidoreductase and NADPH ferredoxin oxidoreductase. The former brings about either the oxidation or reduction of NAD by the equilibration of electrons between NAD and ferredoxin, although reduction of ferredoxin by NADH is thermodynamically highly unfavourable. NADPH ferredoxin oxidoreductase controls the production of NADPH for cell biosynthesis from reduced ferredoxin (Jones and Woods, 1986).

Recently however, it has been found that many Clostridia, including *C. beijerinckii*, contain an Rnf (*Rhodobacter* nitrogen fixation) complex which has NADH ferredoxin

oxidoreductase activity. This complex transfers electrons from reduced ferredoxin to NAD which is coupled to the pumping of protons out of the cell and the energy contained within the proton gradient can be used for ATP synthesis by ATPase activity from reduced ferredoxin. This system is reversible and can, in principle, also lead to energy-dependent reduction of ferredoxin by NADH (Biegel et al., 2011; Dürre, 2014; Herrmann et al., 2008; Lütke-Eversloh and Bahl, 2011).

Acetyl-CoA forms the basis for the branched fermentation of ABE formation as it leads to both acidogenic and solventogenic products (Jones and Woods, 1986; Köpke et al., 2011; Lütke-Eversloh and Bahl, 2011). The acidogenic phase sees the production of acetate, butyrate, carbon dioxide and hydrogen. Acetate is formed from acetyl-CoA via phosphotransacetylase (phosphate acetyltransferase) (Pta) and acetate kinase (Ack) and butyrate formed from butyryl-CoA via phosphotransbutyrylase (Ptb) and a butyrate kinase (Buk) (Andersch et al., 1983; Jones and Woods, 1986; Köpke et al., 2011). To form butyryl-CoA, several intermediate steps are involved which first sees acetyl-CoA converted to acetoacetyl-CoA by thiolase (Thl). This is then reduced to 3-hydroxybutyryl-CoA with NADH by a 3-hydroxybutyryl-CoA-dehydrogenase (Hbd) and further converted to crotonyl-CoA by the cronotase (Crt) with the release of H_2O. The final step sees a further reduction of crotonyl-CoA using butyryl-CoA-dehydrogenase (Bcd) to finally form butyryl-CoA in a complex reaction which involves electron bifurcation and thus the concomitant conversion of 2 NADH and 1 oxidised ferredoxin to 2 NAD and 1 reduced ferredoxin (Jones and Woods, 1986; Köpke et al., 2011; Lee et al., 2008b). The acidogenic phase allows conservation of more energy in the form of ATP than the solventogenic phase, with acetate forming twice as much ATP than butyrate (Jones and Woods, 1986). However generation of acetate leads to excess NADH as no NADH is consumed in the process and its formation therefore must be balanced by the production of butyrate. The balance between the two acids is also

maintained by the aforementioned NADH ferredoxin oxidoreductase to generate NADH and the hydrogenase for H_2 production as excess NADH will shift the pathway towards butyrate (Jones and Woods, 1986).

With the continual production of these organic acids during the acidogenic phase the pH will begin to drop. A continual drop in pH will lead to destruction of the proton gradient as undissociated acids will begin to diffuse back into the cell allowing protons to enter (Jones and Woods, 1986). The collapse of the pH gradient will lead to cessation of important metabolic functions and can cause a culture to 'acid crash' (Maddox et al., 2000). To avert this, the solventogenic pathway is expressed which sees acetate and butyrate reassimilated (partly) into the cell (Lütke-Eversloh and Bahl, 2011). The switch to this phase has been shown to be controlled by the master regulator Spo0A (Ravagnani et al., 2000). Once this switch begins cells also undergo physiological changes with the formation of spores which is also controlled by Spo0A (Dürre, 2014; Dürre and Hollergschwandner, 2004).

The reassimilation of acetate and butyrate is achieved via the acetoacetyl-CoA:acetate/butyrate-coenzyme A transferase (CftA/CtfB) which uses acetoacetyl-CoA as the CoA donor and creates acetyl-CoA and butyryl-CoA. This reaction is coupled with the conversion of acetoacetyl-CoA to acetoacetate which is subsequently converted to acetone by acetoacetate decarboxylase (Adc) (Jones and Woods, 1986; Laursen and Westheimer, 1966; Lütke-Eversloh and Bahl, 2011). The now formed acetyl-CoA and butyryl-CoA are converted to ethanol and butanol respectively. They are first converted to either acetaldehyde or butyraldehyde by acetaldehyde/butyraldehyde dehydrogenase (AdhE1/2) and then to ethanol and butanol by ethanol/butanol dehydrogenase (BdhA/B) (Grimmler et al., 2011; Lütke-Eversloh and Bahl, 2011; Sauer, 1995). Although the conversion to

solvents helps prevent acid crashing, solvents themselves are still toxic to clostridial cells (Bowles and Ellefson, 1985) and therefore the formation of spores helps preserve future survival.

1.4. Physiological changes undergone by solventogenic species

During the acidogenic phase cells appear as highly motile, thin rods which transition into swollen, cigar shaped forms with the beginning of solventogenesis (Jones et al., 1982a). This transition sees the accumulation of the starch like molecule granulose for sugar storage. It is suspected that granulose is accumulated to serve as an energy source for spore formation (Dürre, 2014; Reysenbach et al., 1986). The most significant physiological change that these bacteria also undergo is the formation of heat resistant spores. Spores are metabolically inactive and are resistant to a variety of environmental stresses including heat, UV radiation, toxic chemicals, oxygen exposure and desiccation and represent a long term survival strategy of the bacterium (Setlow, 2006). The formation of spores allows for survival when conditions are unfavourable and will lead to germination when the local environment becomes less hostile. Germination sees the spores return back to a growing cell to continue the life cycle. The formation of spores is not limited to *Clostridia* but also occurs, and is far better studied, in *Bacillus* species.

1.4.1. Morphological changes in spore formation

In the initial phases of spore formation, the cell is split in half with a larger mother cell and a smaller prespore (Piggot and Hilbert, 2004). The chromosome is segregated and packed inside the prespore which the mother cell will eventually engulf (Dürre and Hollergschwandner, 2004). Two cytoplasmic membranes thus surround the prespore, the inner and outer prespore membranes. Between these two membranes is a further two layers of peptidoglycan with different composition, one being the cortex and the other germ cell wall (Driks et al., 2000; Dürre and Hollergschwandner, 2004). The spore coat is synthesised on the outside of the prespore membrane which contains several protein layers only made by the mother cell during sporulation (Dürre and Hollergschwandner, 2004). The spore itself contains large amounts of dipicolinic acid which chelates cations and 3-phosphoglycerate which provide an energy source once the spore germinates (Driks et al., 2000). The segregated chromosomal DNA within the spore is bound by small acid-soluble spore proteins (SASPs) which changes the DNA structure to an A-like confirmation as so to protect the DNA and to provide a further energy source post germination (Driks et al., 2000; Dürre and Hollergschwandner, 2004; Mohr et al., 1991). SASP primary sequences have been found to very conserved amongst *Bacillus* and *Clostridia* (Driks et al., 2000).

A subtle difference between *Bacillus* and *Clostridia* is the appearance of appendages on clostridial spores that arise from the spore coat and the accumulation of the aforementioned granulose in the mother cell which causes cells to swell and appear cigar shaped (Dürre and Hollergschwandner, 2004; Woods and Jones, 1987).

1.4.2. Regulation and initiation of spore formation

In both *Bacillus* and *Clostridia* Spo0A plays a crucial role in spore formation and is known as the master regulator of this system. In the *B. subtilis* model, regulation of Spo0A is through the use of a phosphorelay system which sees Spo0F, the first in the relay, be phosphorylated by histidine kinases. Once phosphorylated, Spo0F transfers a phosphoryl group to Spo0B which subsequently transfers this to Spo0A which can begin gene expression of the related spore genes through the use of 0A boxes upstream of genes controlled by Spo0A. Spo0A binds to 0A boxes and, after recruiting RNA polymerase, gene transcription is initiated (Dürre, 2014; Perego, 1998; Piggot and Hilbert, 2004; Steiner et al., 2011; Strauch and Hoch, 1993). This phosphorelay system has not been found in *Clostridia* as species lack the *spo0F* and *spo0B* genes (Dürre, 2014; Dürre and Hollergschwandner, 2004). In *C. acetobutylicum* however it was found that activation of Spo0A was achieved via direct phosphorylation of Spo0A by orphan histidine kinases (Steiner et al., 2011). This may represent a truer model of Spo0A activation in *Clostridia* as potential orphan kinases have been found in many other species (Dürre, 2005; Paredes et al., 2005).

The development of a mature spore is incredibly complex and still mostly understood in *Bacillus*. In the *B. subtilis* sporulation system, σ^H is an alternative sigma factor that is always found in low amounts in vegetative cells. A feedback loop exists in which σ^H promotes the expression of *spo0A* however σ^H transcription is usually supressed by AbrB. Once Spo0A becomes phosphorylated however, it is able to indirectly cause higher expression of σ^H as Spo0A~P represses *abrB* expression (Al-Hinai et al., 2015). With higher amounts of σ^H and now Spo0A~P the *spoIIA* operon (*spoIIAA, spoIIAB* and *sigF*) can be expressed which encodes for the anti-anti sigma factor SpoIIAA, an anti-sigma factor SpoIIAB and σ^F (Dürre and Hollergschwandner, 2004; Piggot and Losick,

2002). SpoIIAB is usually bound with σ^F rendering the latter inactive. SpoIIAB also contains a binding site for either ATP or ADP. When bound with ATP, SpoIIAB is able to phosphorylate SpoIIAA which inactivates SpoIIAA (Al-Hinai et al., 2015; Dürre and Hollergschwandner, 2004). SpoIIAA~P is found mostly in the mother cell and prevents σ^F transcription through binding (Dürre, 2014). Once SpoIIAA becomes dephosphorylated by the membrane bound phosphatase SpoIIE, it is able to bind with SpoIIAB to form a ADP–SpoIIAB–SpoIIAA complex which releases σ^F which can now express σ^F controlled operons (Al-Hinai et al., 2015; Dürre and Hollergschwandner, 2004). One of these operons encodes the signal protein SpoIIR which when secreted into the space between the mother cell and prespore, activates SpoIIGA. SpoIIGA is a protease that is able to cleave pro-σ^E into σ^E within the mother cell leading to activation of σ^G in the prespore which further activates σ^K in the mother cell (Dürre, 2014; Dürre and Hollergschwandner, 2004). This is known as the 'crisscross regulation' of sporulation which is summarised in Figure 1.2 (Losick and Stragier, 1992).

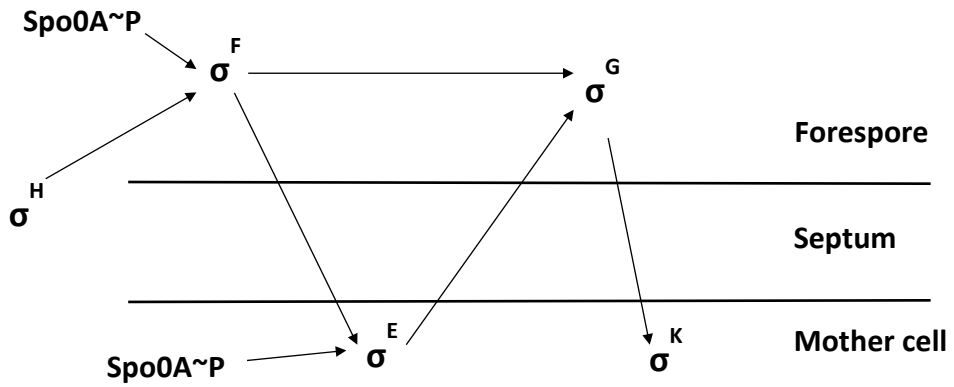


Figure 1.2 Crisscross regulation of sigma factors during endospore formation in *B. subtilis* and *C. acetobutylicum* (adapted from Dürre and Hollergschwandner (2004) and Losick and Stragier (1992)).

σ^G and σ^K allow expression of various operons whose products are required for spore maturation (Dürre, 2014). Although this system has been most studied in *B. subtilis* the σ^F, σ^E, σ^G, and σ^K genes have been identified in *C. acetobutylicum* indicating a similar system exists (Al-Hinai et al., 2015; Nölling et al., 2001; Sauer et al., 1994).

1.5. Current limitations to ABE commercialisation

Many limitations still prevent global commercialisation of the ABE process, which include high feedstocks costs, overall running costs and low butanol yields (Green, 2011). The use of sustainable and cheaper feedstocks such as agricultural waste may make a contribution to reduce running costs. For instance, lignocellulosic material as described in 1.1.2.2. has been considered by many an ideal material. However, issues arise with its processing and pre-treatment, which aims at releasing the contained fermentable sugars in a cost-effective manner without the accumulation of toxic by-products. Unfortunately, enzymatic release of sugars from physically or chemically pre-treated biomass is still not cost-efficient (Klein-Marcuschamer et al., 2012). These issues may be overcome by the use of highly efficient enzyme complexes termed cellulosomes which are multienzyme structures attached to the cell surface via scaffoldin proteins (Artzi et al., 2017; Bayer et al., 2004). These enzyme complexes are incredibly efficient as several different enzymes, each with their own substrate specificity, can be added to the complex allowing for better substrate hydrolysis (Artzi et al., 2017). This therefore makes their use ideal when degrading different biomass sources. Many *Clostridia* are unable to produce functional cellulosomes even when the necessary genes are within the genome (López-Contreras et al., 2004). Engineering of *Clostridia* strains has yielded promising results with minicellulosome

activity, e.g. for *C. acetobutylicum* (Willson et al., 2016). Further strain improvements are still needed to make this route economically viable.

Another key area that needs tackling is the innate solvent toxicity that occurs when solvents reach above 20 g/L (Jones and Woods, 1986). High levels of butanol will increase membrane fluidity leading to membrane destabilisation (Gu et al., 2011; Vollherbst-Schneck et al., 1984). Without a stable membrane maintenance of the internal pH will be abolished, intracellular ATP levels will be reduced, glucose uptake will be inhibited, ion transport (sodium/potassium pumps) will be disrupted all leading to detrimental effects on cellular metabolism (Bowles and Ellefson, 1985; Ezeji et al., 2010). Attempts have been made to increase tolerance with some success however yields of solvents were still as before or were not a level that was economically viable (Ezeji et al., 2010; Hermann et al., 1985; Lin and Blaschek, 1983; Lütke-Eversloh and Bahl, 2011; Qureshi and Blaschek, 2001).

High extraction costs of a relatively low amount of butanol is also an issue (Green, 2011). Improved yields have been attempted by using gas stripping methods to remove the effects of solvent toxicity (Ezeji et al., 2005; Lu et al., 2013) and increase solvent recovery by using liquid–liquid extraction, adsorption, pervaporation, reverse osmosis, alginate adsorbent immobilisation and aqueous two phase separation (Ezeji and Li, 2010; Vane, 2008; Ye et al., 2018). Furthermore, increased yields can be gained by genetically engineering pathways to shift the metabolism towards solvent production. This has been achieved in several ways including reducing acetone formation using antisense RNA (Tummala et al., 2003a) and overexpressing the alcohol-aldehyde dehydrogenase whilst simultaneously downregulating the CoA transferase for acetone formation with antisense RNA (Tummala et al., 2003b). Furthermore inactivation of the butyrate kinase gene coupled with overexpression of

alcohol-aldehyde dehydrogenase led to increased butanol yields (Harris et al., 2000). This was further demonstrated by shifting the butanol pathway through disruption of the *pta* and *buk* genes whilst simultaneously overexpressing the alcohol dehydrogenase to also increase butanol yields (Jang et al., 2012). The use of genetic manipulation tools has proved key in improving strains ability to produce and tolerate butanol and technological developments in this field are still ongoing.

1.6. Genetic tools available for use in *Clostridia*

Genetic manipulation of solventogenic *Clostridia* has become one of the main areas of focus for strain improvement. Fermentation strains were originally selected based on several factors such as the fermentation substrate that was to be used, the desired end product ratios, natural phage resistance and whether the strain required additional nutrients (Jones and Woods, 1986). Nowadays with the advent of advanced genetic engineering, strains can be tailored depending of the fermentation required. Desirable areas of strain improvement include increased oxygen tolerance, higher cell densities along with increased cell viability, utilisation of cellulosic material and most importantly increased butanol yields (Papoutsakis, 2008).

1.6.1. The development of vectors

Early genetic tools included a variety of shuttle vectors derived from *B. subtilis and E. coli* that enabled the transformation of genes into clostridial hosts. Several groups developed vector systems that could bring fermentation genes into *C. acetobutylicum* to overexpress important ABE genes (Lee et al., 1992a, 1992b; Mermelstein et al.,

1992; Oultram et al., 1988). A significant problem was low transformation efficiencies due to the highly active restriction system that *C. acetobutylicum* possess which would cut unmethylated foreign DNA (Lütke-Eversloh, 2014). This problem was soon overcome by the use of methyltransferase gene from *B. subtilis* phage φ3TI which was able to methylate plasmid DNA to protect it from endonuclease activity (Mermelstein and Papoutsakis, 1993).

Another advancement in the availability of plasmids for *Clostridia* was the development of the modular pMTL8000 plasmid system (Heap et al., 2009). This *E.coli*/*Clostridium* shuttle vector system includes four basic modules that can be exchanged using various restriction enzymes. The modules comprise different replicons, resistance markers and several additional features such as multiple cloning sites or the *tra* genes for conjugative transfer and can be adapted depending on the vector the user requires (Heap et al., 2009; Lütke-Eversloh, 2014).

1.6.2. Intron based gene inactivation

In addition to a successful modular plasmid system for *Clostridia* an important development in clostridial tools was the use of mobile group II introns. Site directed gene disruption has been achieved with the use of targeting vectors (Targetrons) based on the retrohoming of mobile Group II introns. These mobile elements are able to insert themselves into specific DNA sequences and can be expressed using plasmids encoding specific ribonucleoprotein (RNP) complexes (Joseph et al., 2018; Zhong et al., 2003).

Group II introns consist of a catalytically active intron RNA (also known as a ribozyme) and an intron-encoded protein (IEP). The IEP is multifunctional as it mediates RNA

splicing, targets specific DNA and has reverse transcriptase activity (Kuehne et al., 2012; Lambowitz and Zimmerly, 2011). The ribozyme is able to catalyse self-splicing which creates an excised intron lariat RNA. This combines with the IEP to form the RNP complex which is able to recognise specific double-stranded DNA target sites for insertion (Zhong et al., 2003). The IEP recognises a small number of fixed positions in double stranded DNA which will trigger local DNA unwinding to allow for the intron RNA to pair to the DNA sequence (Zhong et al., 2003). Base pairing interactions uses three short sequence elements in the DNA target site which are recognised by the complementary sequences of the intron RNA. Once recognised the intron RNA is inserted into one of the DNA strands by reverse splicing with the IEP cleaving the opposite strand to use the 3' end as a primer for reverse transcription of the inserted intron RNA (Zhong et al., 2003). The resulting cDNA is then integrated into the genome using the host repair systems (Cousineau et al., 1998).

It was the *ltrB* gene of Lactococcus lactis (Ll.ltrB) that paved the way for group II intron use and the specificity of the insertion can be changed by modifying the intron RNA (Karberg et al., 2001). Furthermore, selection of mutants can be achieved by the development of the retro-transposition-activated selectable marker (RAM) which activates a resistance cassette during insertion. The RAM is encoded in the group II intron but is however disrupted by a group I intron (Heap et al., 2007; Zhong et al., 2003). During integration of the group II intron, the group I intron will be spliced out leading to full activation of the resistance cassette. Integrants can therefore be selected based on their resistance to various antibiotics as resistance will only occur if there has been successful integration of the group II intron (Heap et al., 2007; Zhong et al., 2003).

This Targetron system has been developed for use in *Clostridia* with the ClosTron system and is similarly based using group II introns (Heap et al., 2007). The system has been further streamlined with the use of FLP recombinase which is able to flip out the resistance cassette when FLP Recognition Target (FRT) sites flank the cassette (Heap et al., 2010). This allows for marker re-cycling so that multiple insertions can be acquired throughout the genome for successful gene disruptions. The ClosTron system has been implemented in variety of different *Clostridia* including *C. beijerinckii, C. acetobutylicum, C. autoethanogenum, C. difficile, C. sporogenes* and *C. botulinum* (Cooksley et al., 2012; Heap et al., 2010; Kuehne et al., 2012; Liew et al., 2017).

Although the use of ClosTrons represents an important tool for clostridial engineering, there are several limitations to this technology. One such limitation is the availability of sequences as smaller genes may not be sufficient in size to allow for successful insertion (Steiner et al., 2012). Furthermore, poor insertion site sequences may lead to incorrect insertions which need to be determined through Southern blotting or whole genome sequencing (Zhong et al., 2003). Gene insertion also run the risk of polar effects which may cause altered expression of surrounding genes (Kuehne et al., 2012).

1.6.3. Allelic exchange

Although ClosTrons provide rapid gene disruptions, the aforementioned limitations meant that newer technologies were required to progress improved strain engineering. The use of homologous recombination for gene integration and deletion has now been employed in several clostridial species. Complementary homology arms on a plasmid that flank target DNA sequences on the chromosome will cause

homologous recombination to occur to allow for removal of DNA sequences. This has allowed for specific in-frame deletions that do not affect downstream open reading frames and also integration of genetic material at specific sites in the genome. The use of homologous recombination has been exemplified with the use of counter/negative selection markers which in *Clostridia* include the *pyrE, codA* and *mazF* systems (Al-Hinai et al., 2012; Cartman et al., 2012; Ehsaan et al., 2016; Heap et al., 2012; Ng et al., 2013).

1.6.3.1. The *pyrE* system

The *pyrE* gene encodes orotate phosphoribosyl transferase which catalyses the *de novo* synthesis of the important pyrimidine base uracil (Ghim et al., 1994). Strains in which this gene is disrupted are no longer able to synthesise uracil but are however resistant to the chemical 5-fluoroorotic acid (5-FOA) which would normally be deadly to wild type cells (Boeke et al., 1984). Creation of *ΔpyrE* mutants has therefore be achieved through homologous recombination and selection of strains ability to grow on FOA (Heap et al., 2012). A summary of how a *ΔpyrE* mutant was created based on Heap et al. (2012) can be seen in Figure 1.3.

In the case of Heap et al. (2012), a plasmid was created that contained one homology arm with the 1200 bp sequence directly downstream of the *pyrE* gene and a second that contained the final 300 bp sequence of the *pyrE* gene itself. Homologous recombination would first occur with the longer 1200 bp sequence and subsequent integration of the plasmid will occur. This is known as the single cross over event. A second recombination event with the homologous sequences on the integrated plasmid will cause excision of the plasmid leaving a truncated *pyrE* on the

chromosome and an intact *pyrE* on the cytoplasmic plasmid which is known as the double cross over event. Mutants with a truncated *pyrE* can now be selected for by their ability to grow on FOA.

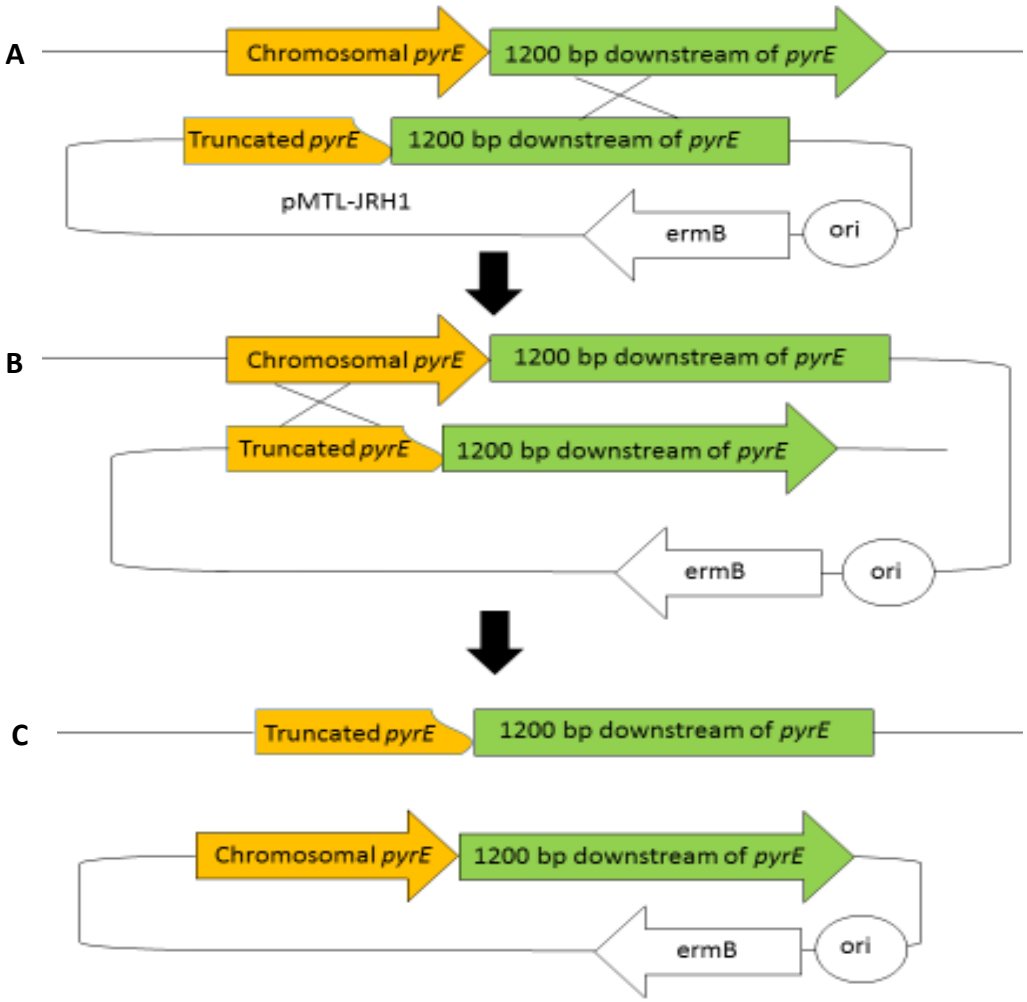


Figure 1.3 Creation of a *ΔpyrE* mutant adapted from Heap et al. (2012).

Creation of a *ΔpyrE* mutant using allele exchange. First homology event (A) with integration of the plasmid and the second homology event (B). Excision of the plasmid and truncation of the *pyrE* gene (C).

The use of the *ΔpyrE* system was first exemplified with the ability to stably integrate large portions of DNA into the *pyrE* locus (Heap et al., 2012). This process was coined

Allele-Coupled-Exchange (ACE) which uses homologous recombination to bring DNA into the chromosome and select for integration based on the state of the *pyrE* gene. Genetic material can be integrated into the chromosome by placing desired genes between the two homology arms. A summary of how this is achieved can be seen below in Figure 1.4.

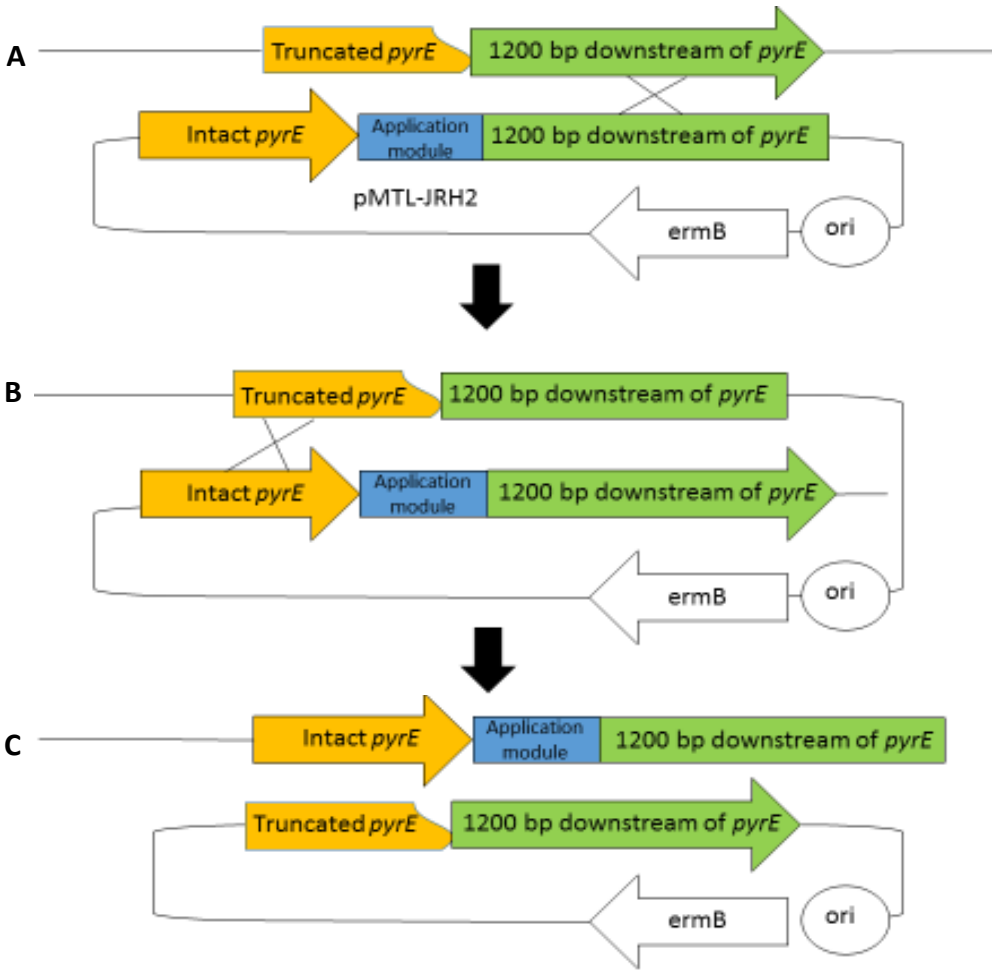


Figure 1.4 Genomic integration at the *pyrE* locus adapted from Heap et al. (2012).

First homology event (A) with integration of the plasmid and the second homology event (B). Excision of the plasmid and restoration of the *pyrE* gene (C). The application module containing any genetic material is now integrated in the chromosome.

When homologues recombination occurs, similar to the truncation of the *pyrE* gene, genetic material between the two homology arms will be brought into the chromosome and stably integrated. With restoration of the *pyrE* gene also happening, integrants can be selected for by their ability to grow on medium not supplemented with uracil. Once the integration of the desired DNA and *pyrE* restoration has occurred, the whole process can be repeated, this time by integrating and selecting for a truncated *pyrE* gene to continually bring more DNA into to the chromosome. This was successfully demonstrated by inserting over 40 kb of the lambda phage DNA into *C. acetobutylicum* (Heap et al., 2012).

The use of the *ΔpyrE* system is not limited to genomic integration and has been employed for in-frame deletion of selected genes (Ehsaan et al., 2016; Ng et al., 2013). The starting point for this method is a previously created *pyrE* mutant of the strain of interest. In this instance a plasmid is created that contains homology arms (approx. 750 bp) that flank a target region which include 2-3 codons of the start and end of the gene and a *pyrE* gene from another organism (*C. sporogenes* has been traditionally used for *C. acetobutylicum*) as it will not recombine with the native *pyrE*. Once transformed, selection of a single cross over event i.e. one of the knock out arms has recombined, can be achieved by utilising the transformants' ability to grow on uracil (Ehsaan et al., 2016; Ng et al., 2013). The double cross over event i.e. recombination of the second arm and subsequent excision and loss of the plasmid can be counter selected for by strains ability to return to growth on 5-FOA as they will now only have their truncated native *pyrE*. This process is summarised in Figure 1.5.

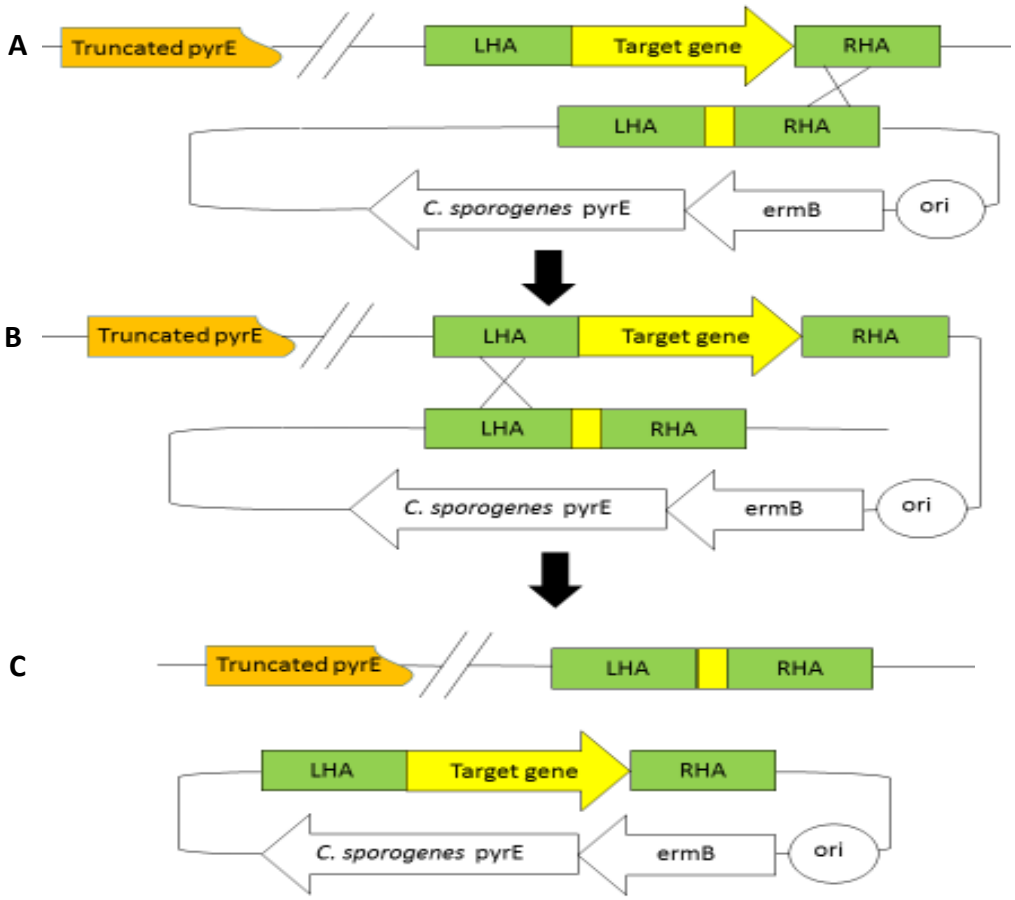


Figure 1.5 Using the *ΔpyrE* system to generate in-frame deletions.

Homologous recombination at the flanking region of the target gene (A) which can either occur at the LHA or RHA. Following integration of the plasmid the second homology event will occur (B) causing excision of the plasmid (C) in which loss of the plasmid can be counter selected for with 5-FOA.

Once confirmation of deletion has occurred, the process can be repeated to continually knock genes out. As a last step, the truncated *pyrE* gene is then repaired to allow a fair comparison to the wild type.

1.6.3.2. *codA* and *mazF* systems

Similar to the *pyrE* system for gene deletions, the use of *codA* and *mazF* as counter selection markers has also been employed for genetic manipulation of *Clostridium* species. The *codA* gene encodes a cytosine deaminase that catalyses the conversion of cytosine to uracil but due to a relaxed substrate specificity will also convert the pyrimidine analog 5-fluorocytosine (5-FC) into the highly toxic 5-fluorouracil (5FU) (Cartman et al., 2012). FU toxicity occurs as it inhibits the key enzyme thymidylate synthase in nucleotide biosynthesis and incorporation of toxic fluorinated nucleotides into DNA and RNA (Longley et al., 2003). In a similar fashion to the *pyrE* system, knockout vectors can be constructed that contain in-frame deletion homology arms and the *codA* gene. After the second cross over event, counter selection for plasmid loss is achieved by supplementing with 5-FC which would kill any strains that contained the plasmid (Cartman et al., 2012).

mazF is an ACA-specific endoribonuclease, also known as mRNA interferase, found in *E. coli* that is usually countered by the anti-toxin *mazE* (Al-Hinai et al., 2012; Yamaguchi and Inouye, 2009). The use of this system was exemplified by Al-Hinai et al., (2012) who used a lactose inducible promoter for the *mazF* on their knockout plasmid. To select for plasmid loss, *mazF* was induced with the addition of lactose and any cells that contained the plasmid would die. Further selection of deletions was achieved by the addition of an antibiotic resistance cassette between the homology arms that is flanked by FRT sites. This was to improve mutant selection and the resistance cassette would be removed once the gene deletion had occurred (Al-Hinai et al., 2012).

1.6.4. CRISPR/Cas based genome editing

The most recent addition to the genetic tool repertoire for *Clostridia* is the use of the CRISPR/Cas system for gene deletions and insertions. The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) system is a bacterial and archaeal immune system evolved to degrade foreign plasmids and DNA that most often come from bacteriophages (Sorek et al., 2013). Each CRISPR locus consists of several short repeat sequences (usually 20-50 bases) that are separated by a unique spacer. It is these unique spacers that form the CRISPR array of targets which is transcribed and processed to produce RNA fragments known as CRISPR-RNA (crRNA) (Wang et al., 2015). Along with the repeat sequence array, the CRISPR-Cas operon encodes Cas proteins which are responsible for DNA cutting (Chylinski et al., 2013). The crRNA will guide the Cas nuclease to the target DNA sequence for cleaving at the specific protospacer-adjacent motif (PAM) (Wang et al., 2015). The most commonly used system is the type II CRISPR/Cas system which uses Cas9 (Hsu et al., 2014). In this system the Cas9 forms a dual-RNA complex with crRNA and a trans-activating CRISPR-RNA (tracrRNA) which is required for the Cas9 nuclease to be functional (Chylinski et al., 2013). This system has been adapted from *Streptococcus pyogenes* and used in a variety of bacteria for genome editing such as *S. pneumonia, Lactobacillus reuteri* and *E. coli* (Jiang et al., 2013; Oh and van Pijkeren, 2014).

The development of the CRISPR/Cas system is relatively new in *Clostridia* but has however been applied to several species. Gene deletions, integrations and also the introduction of small nucleotide polymorphisms into specific genes has been achieved in *C. beijerinckii* (Wang et al., 2016) which also included knocking out the master regulator gene *spo0A* (Wang et al., 2015). Furthermore CRISPR has been employed in

several other *Clostridia* including *C. acetoubutylicum* (Li et al., 2016), *C. saccharoperbutylacetonicum* (Wang et al., 2017), *C. pasteurianum* (Pyne et al., 2016), *C. autoethanogenum* (Nagaraju et al., 2016) and *C. difficile* (Wang et al., 2018). Efficiency of gene deletions do still range from 10-100% depending on the organism however significant steps have been made and the use of CRISPR/Cas seems to be at the forefront of clostridial gene editing.

1.6.5. Transposon mutagenesis

With the use of the aforementioned genetic tools, traditional reverse genetics has been utilised to elude to a genes specific role through genetic modification to observe phenotypes. In contrast the forward genetic approaches assesses a particular phenotype without making any prior assumptions (Cartman and Minton, 2010). This approach uses random transposon insertions to create large mutant libraries which can be screened for certain phenotypes in various conditions. The mariner transposable element *Himar1* has been found to insert randomly in the genome of bacteria and forms the main system used for forward genetics (Cartman and Minton, 2010). This system has subsequently been employed in a selected group of *Clostridia* including *C. difficile* (Cartman and Minton, 2010), *C. perfringens* (Liu et al., 2013), *C. acetobutylicum* and *C. sporogenes* (Zhang et al., 2015). The advantage when using this system is the large libraries that are created which can elude to potentially important genes that may not be detected in traditional reverse genetics.

1.7. The phenomenon of clostridial strain degeneration

Although several limitations still exist when attempting to commercialise ABE fermentation (section 1.5) one major, and perhaps the least understood, issue is the phenomenon of strain degeneration. This phenomenon sees solventogenic species lose their ability to produce solvents and endospores when grown for extended periods of time (Jones and Woods, 1986). Successive serial transfers of a clostridial culture, as such seen in batch fermentations, has been described to cause a progressive decrease in the amounts of solvents and spores produced (Grimbert, 1893; Hartmanis et al., 1986; Jones and Woods, 1986; Kashket and Cao, 1995; Kutzenok and Aschner, 1952). Furthermore, the phenomenon is not limited to serial transfers as solventogenic species grown in continuous fermentations will still produce asporogenous mutants unable to produce solvents (Finn and Nowrey, 1959; Stephens et al., 1985; Woolley and Morris, 1990).

Along with the loss of solvent and spore formation, aberrant colony morphologies have also been observed (Adler and Crow, 1987; Kashket and Cao, 1995). Several different colony types have been documented which range from colonies with brown centres, colonies with outgrowths and colonies that appeared larger, more diffuse and translucent compared to the wild type. With the categorisation of morphologies it has been possible to predict each colony types solvent and spore capabilities; however this was found to not always be a true indicator (Adler and Crow, 1987; Kashket and Cao, 1995). Furthermore Fourier transform infrared spectroscopy has been employed to distinguish degenerate colonies as outgrowths gave a distinct spectra to those of the parental strain (Schuster et al., 2001). The main body of

research into degeneration was mostly undertaken at various periods between the 1950s-1990s but no definitive cause was ever established.

In *C. acetobutylicum* ATCC 824, one of the most well studied solventogenic strains, it was found that the loss of the pSOL1 megaplasmid was perhaps the main cause for this strain to degenerate (Cornillot et al., 1997). This megaplasmid contains many of the necessary solventogenic genes so its loss would lead to degenerate phenotypes. To monitor degeneration in this strain the use of the *amyP* gene on the megaplasmid which encodes an extracellular α-amylase has been employed as it can serve as a reporter gene if the megaplasmid is lost (Sabathé et al., 2002). In addition, more modern techniques have used real-time qPCR with primer sets only binding to sequences present on the megaplasmid to observe whether it was present or not to indicate if strains had degenerated (Lee et al., 2010). Although it is possible to monitor for its presence, the cause as to why the pSOL1 megaplasmid is lost is still unknown and may merely be due to inefficient segregation during growth.

Although studying megaplasmid loss is useful for *C. acetobutylicum* ATCC 824, many if not all other important solventogenic *Clostridia* do not possess a megaplasmid as all the solventogenic genes are on their chromosome (Kosaka et al., 2007). *C. beijerinckii* NCIMB 8052 has been found to degenerate more rapidly than other strains with several examples of this strain degenerating within 24 hours (Adler and Crow, 1987; Kashket and Cao, 1995; Stephens et al., 1985; Woolley and Morris, 1990). Studies using this strain have postulated that cells may degenerate due to excessive acidification of the medium depending on their growth rates (Kashket and Cao, 1995). In this case, cells that are growing rapidly will produce acetate and butyrate at an uncontrolled rate which exceeds the rate of induction of the solventogenic pathways. This will cause a decrease in pH which would lead to bactericidal levels with cells

unable to switch to solventogenesis. Therefore, if the growth rate were decreased, the rate of acid production would decrease and thus allow the solventogenic pathway to be induced fast enough for acids to be reassimilated (Kashket and Cao, 1995). This hypothesis does however not offer any explanation as to how this links to any stably inherited genetic changes. In support of this hypothesis however, when cells were maintained in a both a pH controlled and density controlled (turbidostat) fermenter, a more sustained production of solvents (and no signs of degeneration) were observed (Stephens et al., 1985). In the same study however, when the cultures were grown in various chemostats, i.e. with a constant flow rate without adjustment of turbidity, with either magnesium-, phosphate- or glucose-limitation, the cultures invariably degenerated (Stephens et al., 1985) suggesting that growth rate is important but not the only factor.

The involvement of a global regulatory gene has also been inferred for *C. beijerinckii* NCIMB 8052 as strains have been isolated that are more resistant to degeneration than the wild type (Kashket and Cao, 1993, 1995). These strains were derived from mating *C. beijerinckii* NCIMB 8052 with *Enterococcus faecalis* BM4110 harbouring the transposon Tn1545 to yield transposon mutants with increased degeneration resistance. The authors postulated that an increased longevity of solvent production is due to alteration at a regulatory locus that was effected by the insertion of the transposon (Kashket and Cao, 1993). It was subsequently found that the transposon had inserted into the 5' end of a gene encoding peptide deformylase (PDF) (Evans et al., 1998). PDFs are known to remove N-formyl from N-formylmethionine at the N-termini of polypeptides and are essential in *E. coli*. The authors hypothesised that reduced but not complete inactivation of the protein caused the organism to grow slower, therefore reducing the ability of the strain to produce excess acids that they believed is linked to degeneration (Evans et al., 1998; Kashket and Cao, 1995).

The addition of acetate and butyrate has helped prevent degeneration in several studies using *C. beijerinckii* NCIMB 8052 and *C. beijerinckii* BA101 which is derived from the former (Chen and Blaschek, 1999a, 1999b; Lee et al., 2008a). These studies found that the addition of organic acids caused an increase in CoA transferase activity (the enzymes responsible for assimilation). This increased activity was thought to be due to higher levels of Spo0A activity which may ensure expression of the CoA transferases and other enzymes involved in solventogenesis (Chen and Blaschek, 1999a, 1999b). The higher Spo0A activity was believed to be from increased acetyl phosphate concentrations which would donate the phosphate group to Spo0A in a phosphorelay system (Chen and Blaschek, 1999b). Although plausible at the time, it has been found that orphan histidine kinases regulate Spo0A activity (Steiner et al., 2011) which may respond to excess acids and begin the phosphorylation of Spo0A.

More recent studies into degeneration of *C. beijerinckii* NCIMB 8052 looked at the effect of calcium carbonate ($CaCO_3$) on a potentially degenerated strain (Jiao et al., 2016; Lv et al., 2016). The addition of $CaCO_3$ was able to partially restore solvent production which the authors attributed to stabilisation of heat shock proteins, enzymes involved in solventogenesis and DNA synthesis (Lv et al., 2016). It is however unclear how this mutant was isolated and whether it actually represents a fully degenerated individual or in fact many strains within a population. Isolation of this mutant was described to be from repeated sub-culturing of a whole culture until there was a 90% reduction in solvents which was taken forward as their degenerated strain (Lv et al., 2016). It could therefore be possible that the addition of $CaCO_3$ buffering allowed for pre-existing strains within the culture that had not fully degenerated to grow more efficiently. Although it is unclear if $CaCO_3$ plays a crucial role in restoring degenerates to wild type, comparative transcriptome studies between this degenerate strain and the wild type showed upregulation of the phosphotransferase

system, sugar metabolic pathways and chemotaxis genes along with downregulation of the genes responsible for sporulation, oxidoreduction and solventogenesis (Zhang et al., 2017). As many of these processes are controlled by sigma factors, the authors postulated a close relationship between degeneration and expression of sigma factors. The underlying causes of degeneration of this strain is however still unclear.

Whilst observing degeneration of *C. saccharoperbutylacetonicum* N1-4, the authors of the study believed that defects in NADH generation was the cause of the strain degenerating (Hayashida and Yoshino, 1989). A degenerate strain derived from repeated subculturing was found to produce constant CO_2 and H_2 during the fermentation which suggested that the electron flow did not change. Furthermore, the degenerate strain used in that study only consumed an equal amount of NADH as what would be used in the EMP (glycolysis) pathway. This indicated that reducing equivalents generated during oxidative decarboxylation of pyruvate to acetyl-CoA are only consumed for H_2 evolution and not NADH generation (Hayashida and Yoshino, 1989). The addition of benzyl viologen (BV), an election acceptor for hydrogenases, restored butanol production and the authors hypothesised that BV could act as an electron donor in NAD+ reduction to generate NADH. Elevated levels of NADH generation would allow production of butanol (Hayashida and Yoshino, 1989). What was causing the reported defects in NADH generation is still unknown.

More recent studies in *C. saccharoperbutylacetonicum* N1-4 found that degeneration of this strain was not due to the loss of the *sol* operon as this was still present and that the addition of a native quorum sensing compound was able to induce solventogenesis in a degenerated strain (Kosaka et al., 2007). The inducer that was isolated was found not to be a peptide but a low molecular compound. The authors postulated that this compound has regulatory effects on the *sol* operon either

through *spo0A* expression or by activating the *sol* genes directly as no 0A boxes were found but only putative transcriptional regulator binding sites (Kosaka et al., 2007). Moreover, the authors hypothesised that degenerated strains are unable to synthesise this molecule to induce the solvent genes and thus will only produce acids. Quorum sensing has been found to regulate several late stage phenotypes in *C. acetobutylicum* ATCC 824 such as granulose and spore formation (Kotte et al., 2017; Steiner et al., 2012) and so regulation by quorum sensing in this strain could be plausible.

Degeneration seems to vary between different strains as the rates and potentially causes have all be postulated to be different. At present there is no overarching cause as to why strains degenerate and further investigation is needed.

1.8. The recent field of sociomicrobiology

The relatively new field of sociomicrobiology aims to use concepts of evolutionary theory and apply these to bacterial communities to give a greater understanding as to why communities perform as they do. The traditional view is that an individual bacterium will behave in a unicellular fashion for its own gain (West et al., 2007a). This has now been found to not be the case as microorganisms in populations will communicate and cooperate to perform certain tasks such as biofilm formation, quorum sensing and pathogenesis (West et al., 2006). These bacterial communities are often split into two populations, the most predominant being the co-operators and the more intrusive cheaters (Cavaliere et al., 2017; West et al., 2006, 2007a). Co-operators exist as they are able to gain their fitness both directly and indirectly. Direct fitness is achieved when cells have a shared interest such as the utilisation of a waste

by-product of one individual by another. This would give fitness to both the utiliser but also to the producer as it is in their interest to make more by-product (Sachs et al., 2004; West et al., 2006, 2007b). Shared interest fitness is often seen for the early colonisers of biofilms in teeth and lungs. This was demonstrated when observing growth of *Streptococcus gordonii*, *Streptococcus oralis* and *Actinomyces naeslundii* which saw poor growth of *S. oralis* and *A. naeslundii* when grown in monocultures, unlike *S. gordonii* which grew unhampered, however once grown in co-culture, growth was unhindered for either *S. oralis* and *A. naeslundii* (Palmer et al., 2001). It is unclear why both required the other but it was believed to be some shared metabolic interaction.

The repression of cheating characteristics is also a form of direct fitness as these characteristics could be detrimental to a population. The classic example is in complex insect populations where most raise the Queens offspring rather than their own and any eggs laid by other individuals are destroyed as it would decrease productivity for the population. More relevant studies found that *Vibrio fischeri* mutants that could not luminesce were unable to colonise the light organs of the squid *Euprymna scolopes*, as they would be unproductive for the bacterial population (Frank, 2003; Visick et al., 2000; West et al., 2006, 2007b).

Indirect fitness, also known as kin selection, is more altruistic which sees individuals helping close relatives reproduce to pass on their genetic material indirectly (Hamilton, 1964; West et al., 2006, 2007a). For example, higher levels of cooperation were seen for the production of siderophores, that benefit the population in iron limiting conditions, when individuals were more related (Griffin et al., 2004). Furthermore, kin discrimination occurs when populations favour genetically related individuals over less related or non-related individuals. For example, when

populations create specialised molecules that only their linage can use such as specific siderophores made by different *Pseudomonas aeruginosa* isolates (Meyer et al., 1997) or autoinducer peptides produced by different *Staphylococcus aureus* strains that can induce virulence in the producer but repress in others (MDowell et al., 2001; West et al., 2006).

Social cheaters on the other hand are able to exist in populations without having to behave cooperatively. Cheating is often seen when the concept of 'public goods' in a population is described (West et al., 2006). Public goods are resources in a co-operating population produced by an individual that can be used by that individual and its neighbours (Cavaliere et al., 2017; West et al., 2006, 2007a). This leaves them open to exploitation as cheaters can utilise these public goods without having to produce them themselves. Several examples of public goods in a population include quorum sensing molecules (Diggle et al., 2007), the iron scavenging molecules siderophores (Ross-Gillespie et al., 2007) and various extracellular enzymes (West et al., 2007a). Making these public goods is metabolically costly leading to cheaters avoiding the cost of production and leaving it to the co-operators.

Many of these social behaviours have been studied in pathogenic species such as *P. aeruginosa* when looking at quorum sensing molecules and siderophores (Andersen et al., 2015; Darch et al., 2012; Diggle et al., 2007; Griffin et al., 2004; West et al., 2006, 2007a). With the increasing use of microbes in industrial fermentations, observing populations dynamics is key to creating stable bioprocesses (Cavaliere et al., 2017). The emergence of cheaters in a population would be detrimental when trying to optimise a specific fermentation. Strain degeneration may well represent a social behaviour that sees a small subset of a population gain a fitness advantage as they lose the ability to produce a specific product in a fermentation. Strain degeneration

has been observed in industrial microbes such as the antibiotic producing *Streptomyces rimosus* (Gravius, 1993) and mevalonic acid-producing *E. coli* (Rugbjerg et al., 2018).

Clostridial degeneration may be a form of social cheating within a community. Perhaps the emergence of degenerates occurs because they have gained an increased fitness to the wild type allowing them to propagate? The public good they may be exploiting could be the production of organic acids as these strains do not need to reassimilate the acids and may rely on the wild type to do so. The concepts of sociomicrobiology may help to explain degeneration in more detail as at present there is no universal explanation.

1.9. Aims of this study

The primary objective of this study was to elude both the genetic and evolutionary driving forces behind clostridial degeneration as this still remains unsolved. At present there is limited genetic evidence as to the genes involved in degeneration besides the inference of a global regulatory gene and potentially sigma factors. Limited genomic comparisons between wild type and any derived degenerates have currently been completed. It would seem a logical step in observing what type of genetic aberrations have occurred over time when cells degenerate. Furthermore, using the concepts of sociomicrobiology and applying them to degeneration has not been attempted as this phenomenon may represent a social trait that is driven by evolutionary forces.

Strain degeneration of solventogenic *Clostridia* still poses a problem to the commercialisation of these organisms. To prevent or limit this phenomenon would be highly advantageous and mark a milestone in making the use of these organisms

viable once again. This study used *C. beijerinckii* NCIMB 8052 as the model organism to study degeneration as it has been shown to degenerate more rapidly than other species. Moreover, this strain and derivatives of it have been used throughout the industrial use of solventogenic *Clostridia.*

The specific objectives of this study were as follows:

I. To create a robust and reliable isolation protocol which will produce degenerates of *C. beijerinckii* NCIMB 8052 quickly and effectively.

II. To apply the developed protocol for the isolation and phenotypic characterisation of mutants, including solvent, spore and granulose producing capabilities to obtain a profile for each isolated strain to more clearly define their individual properties. The stability of the culture will also be assessed to establish the progression of whole culture degeneration.

III. To undertake genomic comparisons between the isolated degenerates and the parental wild type they were derived from. This will reveal any important genetic changes that have occurred when strains have degenerated.

IV. To explore the social forces driving degeneration with the use of competition assays against the wild type and between degenerated isolates. By comparing performance against the wild type and each other, differences in relative fitness can be assessed and quantified, providing an explanation for the emergence of degenerates and changes observed at various stages of culture degeneration.

2. Materials and Methods

2.1 Bacterial strains, plasmids and oligonucleotides

Table 2.1 Bacterial strains used in this study

Strain	Properties	Source
E. coli TOP10	Cloning *E. coli* strain	Invitrogen
C. beijerinckii NCIMB 8052	Wild type *C. beijerinckii* NCIMB 8052	NCIMB
C. beijerinckii NCIMB 8052 parental wild type 1	*C. beijerinckii* 8052 parental wild type for the first set of degenerates	This Study
C. beijerinckii NCIMB 8052 parental wild type 2	*C. beijerinckii* 8052 parental wild type for the second set of degenerates	This Study
C. beijerinckii 8052 RD1	Isolated *C. beijerinckii* 8052 degenerate RD1 (first set)	This Study
C. beijerinckii 8052 RD2	Isolated *C. beijerinckii* 8052 degenerate RD2 (first set)	This Study
C. beijerinckii 8052 RD3	Isolated *C. beijerinckii* 8052 degenerate RD3 (first set)	This Study
C. beijerinckii 8052 RD4	Isolated *C. beijerinckii* 8052 degenerate RD4 (first set)	This Study
C. beijerinckii 8052 RD5	Isolated *C. beijerinckii* 8052 degenerate RD5 (first set)	This Study
C. beijerinckii 8052 RD1-1	Isolated *C. beijerinckii* 8052 degenerate RD1-1 (second set)	This Study
C. beijerinckii 8052 RD1-2	Isolated *C. beijerinckii* 8052 degenerate RD1-2 (second set)	This Study
C. beijerinckii 8052 RD2-1	Isolated *C. beijerinckii* 8052 degenerate RD2-1 (second set)	This Study
C. beijerinckii 8052 RD2-2	Isolated *C. beijerinckii* 8052 degenerate RD2-2 (second set)	This Study
C. beijerinckii 8052 RD3-1	Isolated *C. beijerinckii* 8052 degenerate RD3-1 (second set)	This Study
C. beijerinckii 8052 RD3-2	Isolated *C. beijerinckii* 8052 degenerate RD3-2 (second set)	This Study
C. beijerinckii 8052 RD4-1	Isolated *C. beijerinckii* 8052 degenerate RD4-1 (second set)	This Study
C. beijerinckii 8052 RD4-2	Isolated *C. beijerinckii* 8052 degenerate RD4-2 (second set)	This Study
C. beijerinckii 8052 RD5-1	Isolated *C. beijerinckii* 8052 degenerate RD5-1 (second set)	This Study

C. beijerinckii 8052 RD5-2	Isolated *C. beijerinckii* 8052 degenerate RD5-2 (second set)	This Study
C. beijerinckii 8052 RD6-1	Isolated *C. beijerinckii* 8052 degenerate RD6-1 (second set)	This Study
C. beijerinckii 8052 RD6-2	Isolated *C. beijerinckii* 8052 degenerate RD6-2 (second set)	This Study
C. beijerinckii 8052 CIC1	Isolated *C. beijerinckii* 8052 degenerate CIC1 (first set)	This Study
C. beijerinckii 8052 CIC2	Isolated *C. beijerinckii* 8052 degenerate CIC2 (first set)	This Study
C. beijerinckii 8052 CIC3	Isolated *C. beijerinckii* 8052 degenerate CIC3 (first set)	This Study
C. beijerinckii 8052 CIC4	Isolated *C. beijerinckii* 8052 degenerate CIC4 (first set)	This Study
C. beijerinckii 8052 CIC5	Isolated *C. beijerinckii* 8052 degenerate CIC5 (first set)	This Study
C. beijerinckii 8052 CIC6	Isolated *C. beijerinckii* 8052 degenerate CIC6 (first set)	This Study
C. beijerinckii 8052 CIC1-1	Isolated *C. beijerinckii* 8052 degenerate CIC1-1 (second set)	This Study
C. beijerinckii 8052 CIC1-2	Isolated *C. beijerinckii* 8052 degenerate CIC1-2 (second set)	This Study
C. beijerinckii 8052 CIC2-1	Isolated *C. beijerinckii* 8052 degenerate CIC2-1 (second set)	This Study
C. beijerinckii 8052 CIC2-2	Isolated *C. beijerinckii* 8052 degenerate CIC2-2 (second set)	This Study
C. beijerinckii 8052 CIC3-1	Isolated *C. beijerinckii* 8052 degenerate CIC3-1 (second set)	This Study
C. beijerinckii 8052 CIC3-2	Isolated *C. beijerinckii* 8052 degenerate CIC3-2 (second set)	This Study
C. beijerinckii 8052 CIC4-1	Isolated *C. beijerinckii* 8052 degenerate CIC4-1 (second set)	This Study
C. beijerinckii 8052 CIC4-2	Isolated *C. beijerinckii* 8052 degenerate CIC4-2 (second set)	This Study
C. beijerinckii 8052 CIC5-1	Isolated *C. beijerinckii* 8052 degenerate CIC5-1 (second set)	This Study
C. beijerinckii 8052 CIC5-2	Isolated *C. beijerinckii* 8052 degenerate CIC5-2 (second set)	This Study
C. beijerinckii 8052 CIC6-1	Isolated *C. beijerinckii* 8052 degenerate CIC6-1 (second set)	This Study
C. beijerinckii 8052 CIC6-2	Isolated *C. beijerinckii* 8052 degenerate CIC6-2 (second set)	This Study
C. beijerinckii 8052 DCOG1	Isolated *C. beijerinckii* 8052 degenerate DCOG1 (first set)	This Study
C. beijerinckii 8052 DCOG2	Isolated *C. beijerinckii* 8052 degenerate DCOG2 (first set)	This Study

C. beijerinckii 8052 DCOG3	Isolated *C. beijerinckii* 8052 degenerate DCOG3 (first set)	This Study
C. beijerinckii 8052 DCOG4	Isolated *C. beijerinckii* 8052 degenerate DCOG4 (first set)	This Study
C. beijerinckii 8052 DCOG5	Isolated *C. beijerinckii* 8052 degenerate DCOG5 (first set)	This Study
C. beijerinckii 8052 DCOG1-1	Isolated *C. beijerinckii* 8052 degenerate DCOG1-1 (second set)	This Study
C. beijerinckii 8052 DCOG1-2	Isolated *C. beijerinckii* 8052 degenerate DCOG1-2 (second set)	This Study
C. beijerinckii 8052 DCOG2-1	Isolated *C. beijerinckii* 8052 degenerate DCOG2-1 (second set)	This Study
C. beijerinckii 8052 DCOG2-2	Isolated *C. beijerinckii* 8052 degenerate DCOG2-2 (second set)	This Study
C. beijerinckii 8052 DCOG3-1	Isolated *C. beijerinckii* 8052 degenerate DCOG3-1 (second set)	This Study
C. beijerinckii 8052 DCOG3-2	Isolated *C. beijerinckii* 8052 degenerate DCOG3-2 (second set)	This Study
C. beijerinckii 8052 DCOG4-1	Isolated *C. beijerinckii* 8052 degenerate DCOG4-1 (second set)	This Study
C. beijerinckii 8052 DCOG4-2	Isolated *C. beijerinckii* 8052 degenerate DCOG4-2 (second set)	This Study
C. beijerinckii 8052 DCOG5-1	Isolated *C. beijerinckii* 8052 degenerate DCOG5-1 (second set)	This Study
C. beijerinckii 8052 DCOG5-2	Isolated *C. beijerinckii* 8052 degenerate DCOG5-2 (second set)	This Study
C. beijerinckii 8052 DCOG6-1	Isolated *C. beijerinckii* 8052 degenerate DCOG6-1 (second set)	This Study
C. beijerinckii 8052 DCOG6-2	Isolated *C. beijerinckii* 8052 degenerate DCOG6-2 (second set)	This Study
C. beijerinckii 8052 FW1	Isolated *C. beijerinckii* 8052 degenerate FW1 (first set)	This Study
C. beijerinckii 8052 FW2	Isolated *C. beijerinckii* 8052 degenerate FW2 (first set)	This Study
C. beijerinckii 8052 FW3	Isolated *C. beijerinckii* 8052 degenerate FW3 (first set)	This Study
C. beijerinckii 8052 FW4	Isolated *C. beijerinckii* 8052 degenerate FW4 (first set)	This Study
C. beijerinckii 8052 FW5	Isolated *C. beijerinckii* 8052 degenerate FW5 (first set)	This Study
C. beijerinckii 8052 FW6	Isolated *C. beijerinckii* 8052 degenerate FW6 (first set)	This Study
C. beijerinckii 8052 FW7	Isolated *C. beijerinckii* 8052 degenerate FW7 (first set)	This Study
C. beijerinckii 8052 FW1-1	Isolated *C. beijerinckii* 8052 degenerate FW1-1 (second set)	This Study

C. beijerinckii 8052 FW1-2	Isolated *C. beijerinckii* 8052 degenerate FW1-2 (second set)	This Study
C. beijerinckii 8052 FW2-1	Isolated *C. beijerinckii* 8052 degenerate FW2-1 (second set)	This Study
C. beijerinckii 8052 FW2-2	Isolated *C. beijerinckii* 8052 degenerate FW2-2 (second set)	This Study
C. beijerinckii 8052 FW3-1	Isolated *C. beijerinckii* 8052 degenerate FW3-1 (second set)	This Study
C. beijerinckii 8052 FW3-2	Isolated *C. beijerinckii* 8052 degenerate FW3-2 (second set)	This Study
C. beijerinckii 8052 FW4-1	Isolated *C. beijerinckii* 8052 degenerate FW4-1 (second set)	This Study
C. beijerinckii 8052 FW4-2	Isolated *C. beijerinckii* 8052 degenerate FW4-2 (second set)	This Study
C. beijerinckii 8052 FW5-1	Isolated *C. beijerinckii* 8052 degenerate FW5-1 (second set)	This Study
C. beijerinckii 8052 FW5-2	Isolated *C. beijerinckii* 8052 degenerate FW5-2 (second set)	This Study
C. beijerinckii 8052 FW6-1	Isolated *C. beijerinckii* 8052 degenerate FW6-1 (second set)	This Study
C. beijerinckii 8052 FW6-2	Isolated *C. beijerinckii* 8052 degenerate FW6-2 (second set)	This Study
C. beijerinckii 8052 Δ*pyrE*	*C. beijerinckii* 8052 mutant with truncated *pyrE* gene	This Study
C. beijerinckii 8052 Δ*pyrE spo0A*::CT*ermB*	*C. beijerinckii* 8052 *spo0A* ClosTron mutant with truncated *pyrE* gene	This Study
C. beijerinckii 8052 Δ*pyrE* Cbei_0017::CT*ermB*	*C. beijerinckii* 8052 C.bei_0017 ClosTron mutant with truncated *pyrE* gene	This Study
C. beijerinckii 8052 Δ*pyrE* Cbei_3078::CT*ermB*	*C. beijerinckii* 8052 C.bei_3078 ClosTron mutant with truncated *pyrE* gene	This Study
C. beijerinckii 8052 *spo0A*::CT*ermB*	*C. beijerinckii* 8052 *spo0A* ClosTron mutant with repaired *pyrE* gene	This Study
C. beijerinckii 8052 Cbei_0017::CT*ermB*	*C. beijerinckii* 8052 C.bei_0017 ClosTron mutant with repaired *pyrE* gene	This Study
C. beijerinckii 8052 Cbei_3078::CT*ermB*	*C. beijerinckii* 8052 C.bei_3078 ClosTron mutant with repaired *pyrE* gene	This Study
C. beijerinckii 8052 *spo0A*::CT*ermB*_comp	*C. beijerinckii* 8052 *spo0A* ClosTron mutant with complemented gene at the *pyrE* locus	This Study
C. beijerinckii 8052 Cbei_0017::CT*ermB*_comp	*C. beijerinckii* 8052 C.bei_0017 ClosTron mutant with complemented gene at the *pyrE* locus	This Study
C. beijerinckii 8052 Cbei_3078::CT*ermB*_comp	*C. beijerinckii* 8052 C.bei_3078 ClosTron mutant with complemented gene at the *pyrE* locus	This Study

Table 2.2 List of plasmids used in this study.

Plasmid	Properties	Source
pMTL83251	*Clostridium* modular plasmid containing a pCB102 Gram-positive replicon, a *ermB* resistance marker, ColE1 + TraJ Gram-negative replicon and a *lacZ* multiple cloning site	Heap et al 2009
pMTL-JRH1	*pyrE* truncation vector containing a pCB102 Gram-positive replicon, a *ermB* resistance marker, ColE1 + TraJ Gram-negative replicon and a *lacZ* multiple cloning site	This study
pMTL84351	Clostridium modular plasmid containing a pCD6 Gram-positive replicon, a *aad9* resistance marker, a ColE1 + TraJ Gram-negative replicon and a *lacZ* multiple cloning site	Heap et al 2009
pMTL-JRH4	*pyrE* repair vector containing a pCD6 Gram-positive replicon, a aad9 resistance marker, a ColE1 + TraJ Gram-negative replicon and a *lacZ* multiple cloning site	This study
pMTL-JRH4_Cbei_0017	pMTL-JRH4 containing the Cbei_0017 gene and approximately 200 bp upstream for the promoter region	This study
pMTL-JRH4_*spo0A*	pMTL-JRH4 containing the *spo0A* gene and approximately 200 bp upstream for the promoter region	This study
pMTL-JRH4_Cbei_3078	pMTL-JRH4 containing the Cbei_3078 gene and approximately 200 bp upstream for the promoter region	This study
pMTL007S-E2::cbei_0017-694|695a	ClosTron plasmid targeting the Cbei_0017 gene	ATUM
pMTL007S-E2::cbei_3078-969|970s	ClosTron plasmid targeting the Cbei_3078 gene	ATUM
pMTL007S-E2::cbei_3077-138|139a	ClosTron plasmid targeting the Cbei_3077 gene	ATUM
pMTL007S-E2::cbei_4884-159|160s	ClosTron plasmid targeting the Cbei_4884 gene	ATUM
pMTL007S-E2::cbei_4885-108|109s	ClosTron plasmid targeting the Cbei_4885 gene	ATUM

pMTL007S-E2::Cbe-spo0A-407a	ClosTron plasmid targeting the *spo0A* gene	Heap et al 2010

Table 2.3 List of oligonucleotides used in this study.

Oligonucleotide	Use in this study	Sequence
8052 LAC_F	Confirmation that degenerates were *C. beijerinckii* NCIMB 8052 were not contaminations	AGAGTGCTGTGAACTAATGGA
8052 LAC_R	Confirmation that degenerates were *C. beijerinckii* NCIMB 8052 were not contaminations	TACCTCCGTGTTCTTTCGTT
pyrE KO LHA_F	Generation of the 300bp LHA for *pyrE* KO using HiFi assembly	CAGGAAACAGCTATGACCGCATGAAGGCATATAAGAAAGAATTTATC
pyrE KO LHA_R	Generation of the 300bp LHA for *pyrE* KO using HiFi assembly	CTTAATACTTTTATCTATTTGAACAATATCTTACATC
pyrE KO RHA_F	Generation of the 1200bp RHA for *pyrE* KO using HiFi assembly	CAAATAGATAAAAGTATTAAGAACAACTCAACGTG
pyrE KO RHA_R	Generation of the 1200bp RHA for *pyrE* KO using HiFi assembly	CATGGTCATATGGATACAGCGGAATGTTAAGATTTAATACCAC
pyrE KO confirmation_F	Confirmation that the *pyrE* of *C. beijerinckii* NCIMB 8052 had been truncated	CAAAGTAGATATTGGAGGAATT
pyrE KO confirmation_R	Confirmation that the *pyrE* of *C. beijerinckii* NCIMB 8052 had been truncated	AACACTTGCAGTCTTATGAAC
pyrE repair LHA_F	Generation of the intact *pyrE* gene for repair. Contains an application module for future complementation	GGTTCCTGCAGGATGAAGGCATATAAGAAAGAAT
pyrE repair LHA_R	Generation of the intact *pyrE* gene for repair. Contains an application module for future complementation	CGCGGCGGCCGCTTATTTTGCACCATATTGTTT
pyrE repair RHA_F	Generation of the intact *pyrE* gene for repair. Contains an application module for future complementation	GGTTGCTAGCAAGTATTAAGAACAACTCAACG
pyrE repair RHA_R	Generation of the intact *pyrE* gene for repair. Contains an application module for future complementation	CCCCGGCGCGCCGGAATGTTAAGATTTAATACCAC
spo0A gene and promoter_F	*spo0A* gene and its native promoter for complementation	CGCGCATATGAACAAATGATGGGTAGCGTATT

spo0A gene and promoter_R	*spo0A* gene and its native promoter for complementation	GGCCGGATCGCA ATTAGCTAACTTT ATTTTTAAGTCTTA A
Cbei_0017 gene and promoter_F	Cbei_0017 gene and its native promoter for complementation	GGCCGCGGCCGC TAATATAAGCATA TAT
Cbei_0017 gene and promoter_R	Cbei_0017 gene and its native promoter for complementation	CGCGCATATGCTT TTACAAATATATA TCA
Cbei_3077 gene and promoter_F	Cbei_3077 gene and its native promoter for complementation	AATTGCGGCCGA GGCTTATAAGGCA GCACTAATA
Cbei_3077 gene and promoter_R	Cbei_3077 gene and its native promoter for complementation	GGCCGGATCCAA CTTACATTAATTC AAAAAATATATCC
Cbei_3078 gene and promoter_F	Cbei_3078 gene and its native promoter for complementation	AATTGCGGCCGC GAATGAAAGAGT GTGTTATAAAAAT AAT
Cbei_3078 gene and promoter_R	Cbei_3078 gene and its native promoter for complementation	GCGCGGATCCCCT CTATGTTCTTAAG TTGAATTCTATA
Cbei_4884 gene and promoter_F	Cbei_4884 gene and its native promoter for complementation	AATTGCGGCCGCA AAAATTCCTCCTA ATTCCAT
Cbei_4884 gene and promoter_R	Cbei_4884 gene and its native promoter for complementation	GGCCGGATCCAAT TTATTCTTCTTCAT ACTCTTTAACT
Cbei_4885 gene and promoter_F	Cbei_4885 gene and its native promoter for complementation	AATTGCGGCCGCT ATCATGCACCTAT TCTTTCTCAAT
Cbei_4885 gene and promoter_R	Cbei_4885 gene and its native promoter for complementation	GGCCGGATCCATA CTATCTGTTTGTTT TAATTTCGTCT
spo0A check_F	Confirmation primers of ClosTron insertion and for SNVs detection	AGAATGGACAAA AGGAGAGA
spo0A check_R	Confirmation primers of ClosTron insertion and for SNVs detection	CCTTCAACGTAGC TTTTCAT
Cbei_0017 check_F	Confirmation primers of ClosTron insertion and for SNVs detection	CTAAATTTGTAAT ATTCAATTATTAA AC
Cbei_0017 check_R	Confirmation primers of ClosTron insertion and for SNVs detection	ATATTCAATTTTCA CAACAAGTT
Cbei_3078 check_F	Confirmation primers of ClosTron insertion and for SNVs detection	ATTGTATTTTTCCC CATTC
Cbei_3078 check_R	Confirmation primers of ClosTron insertion and for SNVs detection	ATAATATAAGCGA GGCCTT
Cbei_4884 check_F	Confirmation primers of ClosTron insertion and for SNVs detection	CCTAGCACTTCCA GAAAAC

Cbei_4884 check_R	Confirmation primers of ClosTron insertion and for SNVs detection	TAAAGTAATATAA GGCATATGCC
Cbei_4885 check_F	Confirmation primers of ClosTron insertion and for SNVs detection	CTCTAACCATTGG TACATCTAG
Cbei_48845 check_R	Confirmation primers of ClosTron insertion and for SNVs detection	ACTTGAAGAGATA AAGGAAGAG
8X3XX-LF	Sequencing primers for the spectinomycin resistance marker	GTGGAATCATCCT CCCAAAC
8X3XX-LR-RR	Sequencing primers for the spectinomycin resistance marker	GATTCAGCCACTG CATTTCC
8XX5X-LF	Sequencing primer for the ColE1+TraJ Gram negative replicon	GTTCCACTGAGCG TCAGACC
8XXX1-LR	Sequencing primer for the *LacZ* multiple cloning site	GCGGTCATAGCTG TTTCCTG
8XXX1-RR	Sequencing primer for the *LacZ* multiple cloning site	TAGCGCCATTCGC CATTCAG

2.2 Microbiological methods

2.2.1 Preparation of Media

Escherichia coli was grown in LB (lysogeny broth) medium which consisted of 10 g/l tryptone, 5 g/l yeast extract, 5 g/l sodium chloride at pH 7.2 with the addition of 15 g/l agar (Bacto-agar, BD) if for solid agar medium. If required, 50 µg/ml erythromycin and 250 µg/ml spectinomycin were added for selection of plasmid bearing strains.

Clostridium beijerinckii NCIMB 8052 was grown in CBM (Clostridial Basal Medium) broth for all phenotypic analyses and CBM agar for colony counting and strain revival (R. W. O'Brien & J. G. Morris, 1971). CBM consisted of 20 ml of 20 g/l $MgSO_4$ x 7 H_2O, 379 µl of 20 g/l $MnSO_4$ x H_2O, 500 µl of 20 g/l $FeSO_4$ x 7 H_2O, 1 ml of 1 g/l p-aminobenzoic acid, 1 ml of 1 g/l thiamine-HCl, 20 µl of 0.1 g/l biotin and 4 g casamino acids (Bacto-casamino acids, acid hydrolysed casein, BD). The mixture was made to a

volume of 760 ml using deionised water and autoclaved. After autoclaving the medium was made to a final volume of 1 litre using the following solutions sterilised by autoclaving: 10 ml 50 g/l K_2HPO_4 anhydrous, 10 ml 50 g/l KH_2PO_4 anhydrous, 20 ml 250 g/l $CaCO_3$ suspension and 200 ml of 30% (w/v) glucose solution (unless stated otherwise). For solid CBM agar 13 g/l of technical agar 1 (Oxoid, UK) was added before autoclaving.

For transformation of *C. beijerinckii* NCIMB 8052 cultures were grown in liquid 2xYTG which consisted of 16 g/l tryptone, 10 g/l yeast extract and 5 g/l sodium chloride. The mixture was made up to a final volume of 800-980 ml depending on the concentration of glucose required and autoclaved. Glucose was added after autoclaving from a 30% (w/v) glucose solution.

Enumeration of heat resistant colonies for spore assays was carried out using CGM (Clostridial Growth Medium) agar which consisted of 50 g glucose, 2 g tryptone, 2 g $(NH_4)_2SO_4$, 1 g yeast extract, 1 g K_2HPO_4, 0.5 g KH_2PO_4, 0.1 g $MgSO_4$ x 7 H_2O, 750 µl 20 g/l $FeSO_4$ x 7 H_2O, 500 µl 20 g/l $CaCl_2$, 500 µl 20 g/l $MgSO_4$ x H_2O, 100 µl 20 g/l $CoCl_2$, 100 µl 20 g/l $ZnSO_4$, 10.2 g agar (Bacto-agar, BD) and deionised water to make up to 1 litre.

For anaerobic work, liquid medium was left to reduce overnight and solid agar plates 4 hours inside an anaerobic workstation (MG1000 anaerobic work station, Don Whitley Scientific). Concentrations of erythromycin (erm) and 5-fluorouracil (5-FOA) to use were determined using MIC plate testing on CBM solid agar. Concentrations of erythromycin varied from 10 µg/ml - 60 µg/ml and 5-FOA from 200 µg/ml - 700 µg/ml. 5-FOA testing agar was also supplemented with 2 µg/ml uracil in the solid agar mix. 50 µl of a culture of *C. beijerinckii* NCIMB 8052 at an optical density (at 600 nm; OD_{600}) of 0.4-0.7 was added to each concentration plate and left for 24 hours for colony

growth. Colonies were enumerated and the MIC determined. Selections used for *Clostridium beijerinckii* NCIMB 8052 were at the following concentrations and added no earlier than 12 hours before use: erythromycin 10 µg/ml, uracil 20 µg/ml or 2 µg/ml in the presence of 5-FOA at 600 µg/ml and spectinomycin 750 µg/ml.

2.2.2 Growth

E. coli was grown aerobically in LB medium at 37 °C and liquid cultures shaking at 200 revolutions per minute (RPM). *E. coli* was taken from bead stocks and put directly into 5 ml liquid media for growth to use for either pre-cultures or for plasmid preparations.

C. beijerinckii NCIMB 8052 was grown at 37 °C anaerobically in an atmosphere of 80% nitrogen, 10% hydrogen and 10% carbon dioxide using an anaerobic workstation (Don Whitley Scientific). Cultures were grown on CBM, CGM and 2xYTG. *C. beijerinckii* NCIMB 8052 was first grown from plating either a streak of a glycerol stock or 50 µl of a heat shocked spore suspension (80 °C for 10 minutes in a heating block) and incubated overnight. Either single colonies or several colonies (depending on the type of experiment) were inoculated into 10 ml of liquid medium (either CBM or 2xYTG) and serially 1:10 diluted down to 10^{-4} and left for overnight growth. Cultures with growth from 0.2-0.8 OD_{600} were used for subsequent experimentation to inoculate larger medium volumes to any desired OD.

2.2.3. Storage

E. coli stocks were made by adding 600 µl of cell culture to a storage bead vial (Microbank, Pro-lab), inverting 5-10 times for attachment followed by the removal of

the top liquid layer. Bead vials were then stored at -80 °C. *C. beijerinckii* was stored either as a 10% (v/v) glycerol stock for mutants and degenerates at -80 °C or as a spore suspension either at 4 °C for continued use or -20 °C for prolonged storage. Spore stocks were made from cultures grown in liquid medium containing 6% glucose for 5 days. Cultures were centrifuged at max speed for 5 minutes and the pellets resuspended in 1 ml phosphate buffered saline (PBS, 0.001 M phosphate buffer, 0.0027 M KCl, 0.137 M NaCl, pH 7.4).

2.2.4 Preparation and transformation of electrocompetent *E. coli*

Desired *E. coli* strains were grown overnight in 10 ml LB liquid medium in 30 ml universal tubes. 2 ml of the overnight medium was inoculated into 200 ml LB liquid medium in 500 ml shaking flasks and grown aerobically at 37 °C shaking at 200 RPM. Hourly OD_{600} measurements were taken until the OD_{600} reached between 0.6-0.8. The culture was then cooled on ice for 15 minutes and subsequently centrifuged using a chilled rotor at 4000xg for 15 minutes. The supernatant was removed and the pellet was suspended in 30 ml ice cold distilled water. Cells were centrifuged again at 4000xg for 15 minutes and the pellet resuspended in 1 ml ice cold distilled water containing 10% (v/v) glycerol. The cells were stored as 50 µl aliquots at -80 °C until needed.

For transformations, aliquots were thawed on ice and 1-5 µl of desired plasmid was added. The mixture was added to precooled 2mm electroporation cuvettes (BioRad, USA) and electroporation was carried out at 2.5 kV, 25 µF and 200 Ω using a Bio-Rad Micro Pulse. Afterwards 950 µl LB medium was added to the cell mixture which was transferred to a sterile 1.5 ml reaction tube. The cell mixture was then incubated at 37 °C shaking at 200 RPM for 1 hour. After recovery, 100 µl of the mixture was plated

on LB agar with antibiotic selection and incubated overnight at 37 °C. Colonies found were screened for the presence of the plasmid.

2.2.5 Preparation and transformation of electrocompetent *Clostridium beijerinckii* NCIMB 8052

Wild type *C. beijerinckii* spore suspension was heated in 50 µl volumes at 80 °C for 10 minutes in a heating block and plated on CGM agar and left to grow overnight. Colonies were inoculated and overnight 1:10 serial dilutions (down to 10^{-4}) in 2xYTG liquid medium were carried out the previous day. The overnight culture at an OD_{600} between 0.2-0.7 was added to 100 ml 2xYTG liquid medium and grown anaerobically at 37 °C until the culture reached an OD_{600} of 0.6. The culture was then cooled on ice for 10 minutes and subsequently centrifuged at 2000xg for 10 minutes at 4 °C. The supernatant was removed and cells washed in 10 ml ice cold anaerobic electroporation buffer (EPB) which consisted of 270 mM sucrose, 1 mM $MgCl_2$ and 7 mM $NaHPO_4$ buffer(pH 7.4). The mixture was then centrifuged again at 2000xg for 10 minutes at 4 °C with the supernatant being removed and the cell pellet resuspended in ice cold 5ml EPB. If the prepared cells were not to be used straight away, 10% (v/v) DMSO was added and aliquots of 800 µl were frozen at -80 °C.

Aliquots were incubated on ice for 10 minutes before use or thawed if frozen. Plasmid DNA (0.5 µg) was added to the prepared cells and the mixture transferred to a precooled electroporation cuvette (electrode distance 0.4 cm, BioRad). The mixture was then incubated for 8 minutes on ice. Electroporation was carried out at 1.25 kV, 25 µF and 200 Ω and cells were returned immediately to ice for 10 minutes. The cell suspension was added to 10 ml 2xYTG liquid medium and incubated anaerobically at 37 °C for 4 hours. Cells were concentrated by centrifuging at 2000xg for 10 minutes at

room temperature and the pellet resuspended in 100 µl 2xYTG. The resuspension was plated on CGM plates with appropriate antibiotic selection and incubated at 37 °C.

2.3 Molecular biology methods

2.3.1 Isolation of DNA

2.3.1.1 Isolation of plasmid DNA from *E. coli*

Plasmid DNA was isolated from 10 ml overnight *E. coli* culture grown in LB medium using Monarch Plasmid Miniprep Kit (New England BioLabs) following the manufacturer's instructions except DNA elution was always in 50 µl deionised water. Isolated plasmids was stored at -20 °C.

2.3.1.2 Isolation of *C. beijerinckii* genomic DNA

Genomic DNA was isolated from 5-10 ml overnight *C. beijerinckii* culture using the Sigma-Aldrich GenElute Bacterial Genomic DNA kit following the manufacturer's instructions except DNA elution was with 50 µl deionised water. Isolated gDNA was stored at -20 °C.

2.3.2 Analysis and manipulation of isolated DNA

2.3.2.1 DNA quantification and purity analysis

Isolated DNA was analysed using a NanoDrop Lite (Thermo Scientific) measuring the absorption at 260 nm and, to estimate purity, the 260/280 nm absorption ratio. Good quality DNA had an A260/A280 ratio of 1.8-2.0, with an absorption of 1 representing 50 μg of dsDNA/ml. Concentrations and purity of genomic DNA that was to be sent for whole genome sequencing was analysed using Qubit Fluorometric quantification (Thermo Scientific) following the manufacturer's instructions.

2.3.2.2 Polymerase chain reaction (PCR) for amplification of DNA

Screening PCRs for colony, plasmid or mutant detection were carried out using DreamTaq (Thermo Scientific) in 20 μl reactions. Each reaction contained 10 μl DreamTaq master mix, 1 μl 10 mM forward primer, 1 μl 10 mM reverse primer, 7 μl deionised water and 1 μl of DNA template at a concentration of between 1-100 ng/μl. The thermal cycling regime for DreamTaq PCRs was 95 °C for 2 minutes (extended to 5 minutes for colony PCR) followed by 30 cycles of 95 °C for 10 seconds, 50 °C for 30 seconds and 72 °C for 1 minute per kb of DNA amplified and a final extension of 72 °C for 10 minutes. DNA that was to be sent for sequencing or was needed for cloning was amplified using Phusion polymerase (New England BioLabs) in 25 μl reactions. Each reaction contained 12.5 μl Phusion master mix, 1.25 μl 10 mM forward primer, 1.25 μl 10 mM reverse primer, 9 μl deionised water and 1 μl DNA at a concentration of between 1-50 ng/μl. The thermal cycling regime for Phusion-based PCRs was 98 °C

for 2 minutes followed by 30 cycles of 98 °C for 10 seconds, 51 °C (55 °C for amplifications of fragments needed for New England BioLabs HiFi assembly) for 30 seconds, 72 °C at 30 seconds per kb and a final extension at 72 °C for 10 minutes. PCR amplifications was carried using a TProfessional Trio cycler (Biometra).

2.3.2.3 Restriction digest of DNA

Cleavage of DNA was carried out using various CutSmart restriction enzymes (New England BioLabs) following the manufacturer's instructions except the incubation time was increased from 30 minutes to 1 hour.

2.3.2.4 Ligation of DNA

DNA ligations were carried out in 10 µl reactions using 1 µl T4 DNA ligase (New England Biolabs), 1 µl 10x ligation buffer, 7 µl of purified DNA fragments and 1 µl cut vector. Reactions were incubated for 1 hour at room temperature.

2.3.2.5 Other DNA assembly methods

Plasmid DNA was also assembled using NEBuilder HiFi DNA assembly following the manufacturer's instructions. This method permits the combination of multiple (up to six) DNA fragments irrespective of end compatibility. 20 µl reactions consisted of 10 µl HiFi master mix, 1 µl DNA fragments, 1 µl vector and 7 µl deionised water. Fragments for the assembly were created from NEBuilder HiFi DNA assembly tool (http://nebuilder.neb.com).

2.3.2.6 Agarose gel electrophoresis

DNA was separated using agarose gel electrophoresis at 100-120 V for 1–1.5 hours in 1x TAE running buffer (40 mM Tris, 1 mM EDTA and 0.1% (v/v) glacial acetic acid) using a compact gel tank (Biometra) and a PowerPac 300 (Bio-Rad). Agarose gel was routinely made to 1% agarose (w/v) with 1x TAE buffer and 10 µl per 100 ml SYBR Safe (Thermo Scientific). SYBR Safe-stained DNA was visualised either with a Syngene UltraBright-LED blue light transilluminator (Syngene) or a Gel Dock XR+ System (BioRad). Band sizes were determined using a 2-log DNA ladder (New England BioLabs).

2.3.2.7 DNA purification

DNA from agarose gel was purified using Monarch DNA Gel Extraction Kit (New England BioLabs) following the manufacturer's instructions. DNA from PCR reactions was purified using Monarch PCR & DNA Cleanup Kit (New England BioLabs) following the manufacturer's instructions.

2.3.2.8 DNA sequencing

Sanger sequencing to confirm correct DNA assembly or confirming targeted and untargeted chromosomal mutations was carried out by Source BioScience Gene Service (Nottingham) with samples prepared as per the service's instructions. Whole genome sequencing was carried out by microbesNG (University of Birmingham) with samples prepared as per the service's instructions.

2.3.3 Plasmid construction

2.3.3.1 pMTL-JRH1 *pyrE* knock vector

pMTL-JRH1 was constructed for the truncation of the *pyrE* gene in *C. beijerinckii* NCIMB 8052. The plasmid was derived from pMTL83251 (Heap et al., 2009). pMTL83251 was digested with NotI, run on an agarose gel and purified. The left homology arm, amplified with primer pair *pyrE* KO LHA_F and *pyrE* KO LHA_R, consisted of the first 300 bp of the 5' part of the gene whereas the right homology arm, amplified with primer pair *pyrE* KO RHA_F and *pyrE* KO RHA_R, encoded the 1200 bp downstream of the *pyrE* gene. PCR amplification of the left and right homology arms was carried out from the *C. beijerinckii* NCIMB 8052 chromosome, run on an agarose gel and purified. The cut plasmid and homology arm fragments were assembled using HiFi assembly (New England BioLabs) and transformed into *E.coli* TOP10. Several clones were sequenced and the correct plasmid was stored in an *E.coli* TOP10 host at -80 °C. A vector map is supplied in the appendix.

2.3.3.2 pMTL-JRH4 pyrE repair and insertion vector

pMTL-JRH4 was constructed for *pyrE* repair and insertion of desired genes at the *pyrE* locus using a *lacZ* multiple cloning site (MCS) between the homology arms. The *lacZ* MCS existed to allow for cloning of DNA into the pMTL-JRH4 plasmid for integration when the *pyrE* gene was repaired. PCR amplification of the left and right homology arms (primers *pyrE* repair LHA_F/LHA_R and *pyrE* repair RHA_F/RHA_R) was carried out from the *C. beijerinckii* NCIMB 8052 chromosome, run on an agarose gel and purified. The left arm contained the intact *pyrE* gene with the restriction sites 5' -SbfI

and 3’ -NotI and the right arm contained the 1200 bp region immediately downstream of the *pyrE* gene with the restriction sites 5’ -NheI and 3’ -AscI for insertion into the vector. Cloning of the two fragments was carried out in two steps. First, pMTL84351 and the left homology arm fragment were digested with SbfI and NotI, run on an agarose gel and purified. The resulting digested plasmid and fragment were subsequently ligated, transformed into *E. coli* TOP10 and clones were sequenced and the correct plasmid stored at -20 °C. The resulting plasmid, pMTL-JRH4.5, and the right homology arm fragment were then digested with NheI and AscI run on an agarose gel and purified. The digested plasmid and fragment were subsequently ligated, transformed into *E. coli* TOP10 and clones were sequenced. The final correct plasmid, pMTL-JRH4, thus contained a multiple cloning site (also encoding the *lacZ* alpha peptide) separating the left and right homology arms. It was stored in an *E.coli* TOP10 host at -80 °C. A vector map is supplied in the appendix.

2.3.3.3 ClosTron plasmid synthesis

ClosTron mutagenesis was carried out as described by Heap et al., 2010. ClosTron variants were designed using the design tool provided on the ClosTron.com website (http://www.clostron.com/clostron2.php), with target sites chosen based on the highest score obtained by the provided Perutka algorithm. The respective ClosTron variant encoding plasmids were synthesised and purchased from ATUM (California, USA, formerly DNA2.0). Plasmids were stored in an *E.coli* TOP10 host at -80 °C.

2.4 Genetic manipulation of *C. beijerinckii*

2.4.1 Generation and isolation of *C. beijerinckii* Δ*pyrE*

Truncation of the *pyrE* gene was carried out similar to Heap et al., 2012. To truncate the *pyrE* gene of *C. beijerinckii* NCIMB 8052, cells were transformed with pMTL-JRH1 and initially plated on CGM + 10 μg/ml erm as described previously. Faster growing colonies were picked and streaked on CBM + 600 μg/ml FOA + 2 μg/ml uracil and incubated overnight. Colonies were restreaked onto CBM + 600 μg/ml FOA + 2 μg/ml uracil and incubated overnight. Colonies were inoculated into 10 ml CBM liquid medium containing 600 μg/ml FOA + 2 μg/ml uracil and incubated overnight. 1 ml stocks were taken to be stored at –80 °C and the remaining culture had genomic DNA extracted as previously described. Genomic DNA was analysed to confirm *pyrE* truncation and to confirm plasmid loss. Clones were then Sanger sequenced to further confirm correct truncation.

2.4.2 Δ*pyrE* repair and/or insertion of desired genes

To correct the *pyrE* gene of Δ*pyrE* mutants back to the wild type sequence, cells were transformed with the plasmid pMTL-JRH4 (with or without genes to insert at the *pyrE* locus) and initially plated on CGM + 750 μg/ml spectinomycin (spec) + 20 μg/ml uracil as described previously. Colonies were picked and streaked onto CBM with no selection and incubated overnight. Emerging colonies were streaked again on CBM with no selection and incubated overnight. Colonies were inoculated into CBM liquid medium and incubated overnight. 1 ml stocks were taken to be stored at –80 °C and

the remaining culture had genomic DNA extracted as previously described. Genomic DNA was analysed using PCR amplification to confirm *pyrE* repair (primer pair *pyrE* KO confirmation_F/R) and to confirm plasmid loss (primer pair 8X3XX-LF and 8X3XX-LR-RR). PCR products were then Sanger sequenced to further confirm correction.

2.4.3 Isolation of ClosTron insertion mutants in *C. beijerinckii*

Isolation of ClosTron mutants was carried out similar to Heap et al., 2010. *C. beijerinckii* NCIMB 8052 Δ*pyrE* was transformed with the desired ClosTron plasmid and recovered as previously described. Transformants were selected for on CGM agar containing 750 μg/ml spec + 20 μg/ml uracil and left to grow overnight. Colonies were streaked onto CGM agar containing 10 μg/ml erm and 20 μg/ml uracil and grown overnight. Colonies were inoculated into 2xYTG containing 10 μg/ml erm and 20 μg/ml uracil and left to grow overnight. 1 ml of culture was stored at – 80 °C and for the remaining cells the genomic DNA was extracted as previously described.

Genomic DNA was screened by PCR using primers that flanked the target gene to confirm for ClosTron insertion (see oligonucleotide table) and the resulting PCR fragments visualised using gel electrophoresis.

2.5 Generation of degenerate isolates and experimentation in culture dynamics

To generate degenerate isolates of *C. beijerinckii*, a spore suspension was heated in 50 μl volumes at 80 °C for 10 minutes in a heating block, serially diluted 1:10 down to

10^{-3} in PBS, plated on CBM agar and left to germinate overnight at 37 °C in the anaerobic work station (Don Whitley Scientific). Single colonies were then inoculated into 10 ml CBM liquid medium. Every 24 hrs 1 ml of culture was inoculated into fresh 9 ml CBM liquid medium in a 1:10 dilution. This was repeated for 5 subcultures. Afterwards, 100 µl of culture was serially diluted 1:10 down to 10^{-5} with anaerobic PBS, plated on CBM agar and incubated for 2 days to give single colonies. Single colonies with different morphologies to the wild type and ones that appeared as the wild type were inoculated into 10 ml CBM liquid medium and incubated overnight. 1 ml of overnight culture was made into a 20% glycerol stock and the remaining culture had the genomic DNA extracted (see 2.3.1.2).

To study culture dynamics with respect to culture productivity (organic solvents, acids and spores) and colony morphologies, the same initial procedure for isolating degenerates (see above) was carried out except the number of subcultures was changed from 5-15. 1 ml fermentation product samples were taken at the point of inoculation and 2 days after inoculation at each successive subculture. These were centrifuged at full speed for 1 minute and the supernatant removed and frozen at -20 °C for future analysis. 1 ml spore samples were taken after 5 days of growth and stored at 4 °C until used in a spore assay (2.6.6). After 1 day of growth, 200 µl samples were taken, serially diluted down to 10^{-5} and plated onto CBM agar. Colonies were left to grow for 2 days and the CFU of each colony type was enumerated.

2.6 Phenotypic analysis of *C. beijerinckii* and derived mutants

2.6.1 Culture preparation for analysis

Wild type and mutants were streaked on CBM agar and left to grow overnight. Colonies of *C. beijerinckii* wild type or derived mutants were inoculated into 10 ml CBM 6% glucose liquid media and prepared as overnight pre-cultures as previously described (2.2.5). The optical densities of overnight pre-cultures were measured to determine those that has reached an OD_{600} between 0.2-0.8. These were then used to inoculate the main culture to an OD_{600} of 0.01. Culture volumes were 60 ml CBM 6% glucose liquid medium in 100 ml conical flasks for growth curves and 25 ml CBM 6% glucose liquid medium in 50 ml falcon tubes for degenerate analysis. Strains were grown in technical triplicates. Solvent and pH samples for growth curves were taken every hour until the culture reached stationary phase and then taken for 5 subsequent days every 24 hours after the point of inoculation. Degenerate fermentation solvent samples were taken at the point of inoculation, 2 days and 5 days post inoculation. Spore samples were taken 5 days after inoculation for both analyses.

2.6.2 Optical density measurements

Optical density at 600 nm was used as a measure of cell growth and determined using a 1 cm gap plastic cuvette (Sigma-Aldrich) and a Thermo Scientific Biomate 3 spectrophotometer using sterile liquid medium as a blank. If samples had OD_{600} above 0.8, samples were diluted 1:10 and re-measured.

2.6.3 Gas chromatography (GC) analysis

Gas chromatography was used to quantify concentrations of acetone, acetate, butyrate, butanol and ethanol produced by *C. beijerinckii* NCIMB 8052 and derived mutants. 1 ml samples for analysis were taken, centrifuged in a table top centrifuge (Eppendorf 5424 R) at full speed for 1 minute and the supernatant removed and frozen until analysis. 500 µl of sample was used for extraction and initially acidified via the addition of 5 µl 10 M sulphuric acid. 500 µl of propyl propionate containing 50 mM of valeric acid as an internal standard was then added. The mixture was vortexed for 10 seconds and centrifuged at full speed for 1 minute. 300 µl of the resulting organic (upper) phase was removed and transferred to a 2 ml glass vial containing a 350 µl glass insert and the vial sealed with a snap cap. Each run had 2 sets of standards prepared from a 1 M stock containing acetone, acetate, butyrate, butanol and ethanol. Standards were prepared to the following concentrations using ELGA water 0 mM, 1 mM, 3 mM, 5 mM, 10 mM, 30 mM, 50 mM, 100 mM and 150 mM. Extraction for the standards was carried out as described for the samples.

Analysis was carried out using a Thermo Focus gas chromatograph equipped with a 30 meter TR-FFAP column (0.25 mm internal diameter) and a flame ionisation detector. The carrier gas was hydrogen supplied at 0.8 ml/min and the flame was maintained by compressed air (350 ml/min), hydrogen (35 ml/min) and nitrogen (30 ml/min). Sample injection volume was 1 µl which was performed by an auto sampler. The injector temperature was 240 °C and the detector temperature was 270 °C. The oven temperature profile was 40 °C held for 2 minutes then ramped at 80 °C/min to

150 °C and a 45 °C/min ramp to 210 °C at which point the temperature remained for 1 minute.

2.6.4 High performance liquid chromatography (HPLC) analysis

HPLC was used to quantify the concentrations of glucose and lactic acid from *C. beijerinckii* NCIMB 8052 and derived mutants. 1 ml samples for analysis were taken, centrifuged in a table top centrifuge (Eppendorf 5424 R) at max speed for 1 minute and the supernatant removed and frozen until analysis. 300 µl of sample supernatant was mixed in equal volume with the internal standard solution which contained 80 mM valeric acid and 0.005 M sulphuric acid. The mixture was filtered through a 0.22 µm filter (Kinesis) into a 300 µl insert inside a HPLC vial. Glucose and lactic acid standards were prepared to the following concentrations using ELGA water, for glucose: 0 mM, 10 mM, 30 mM, 50 mM, 100 mM, 150 mM, 250 mM, 300 mM and 350 mM. For lactic acid: 0 mM, 1 mM, 3 mM, 5 mM, 10 mM, 15 mM, 20 mM, 30 mM and 50 mM. Injection volumes were 20 µl using a Dionex Ultimate 3000 HPLC system (Thermo Fisher Scientific, UK), equipped with a BioRad Aminex HPX-87H column (BioRad, UK), a refractive index and diode array detector at UV 210 nm with an isocratic flow rate of 0.5 ml/min of 0.005 M sulphuric acid as mobile phase and a column temperature of 35 °C. Concentrations of glucose and lactic acid were determined using Chromeleon 7.2 Chromatography Data system (Thermo Fisher Scientific, UK).

2.6.5 Sporulation assay for determination of heat resistant CFU/ml

Viable spore numbers were quantified by enumerating heat resistant colony forming units. 100 µl of 5-day culture was transferred to either a 1.5 ml sterile reaction tube or a 96 well plate for multiple assays and heat-shocked at 80 °C for 10 minutes either in a heating block (reaction tube) or a PCR block (96 well plate). Serial dilutions at 1:10 in PBS down to 10^{-5} were carried out with thorough mixing. Dilutions were transferred to the anaerobic work station (Don Whitley Scientific) and 3x 10 µl drops per dilution of the heat treated cell suspension was plated onto a CGM agar. Colonies were enumerated after 24 hrs of incubation at 37 °C.

2.6.6 Granulose staining

Wild type and mutants were streaked on CBM agar and left to grow overnight. Resulting colonies were inoculated into 10 ml CBM 6% glucose liquid media and prepared as overnight pre-cultures as previously described (2.2.5). The optical density of overnight pre-cultures was measured to see which were growing at OD_{600} between 0.2-0.8. For each respective culture, three 10 µl aliquots were taken and placed into three separate wells of a 48-well microplate (Corning Incorporated) containing 1.25 ml CBM agar per well. Cultures were left to grow for 48 hours. Gram's iodine solution was dropped onto the cultures to visualise the presence of granulose.

2.7 Mono- and co- culture experiments

2.7.1 Monoculture experiments

Wild type spores were heat shocked as previously described (2.2.5) and plated on CBM agar whereas for non- or poorly sporulating mutants, glycerol stocks were used. Developing colonies were inoculated into 10 ml CBM 6% glucose liquid media with and without calcium carbonate buffering (as required) and prepared as overnight pre-cultures as previously described (2.2.5). Overnight pre-cultures were measured to see which were growing at OD_{600} between 0.2-0.8. Respective cultures were used to inoculate the main culture to an OD_{600} of 0.01. Each fermentation was carried out in 50 ml CBM liquid medium, either with or without calcium carbonate, as required. 1 ml samples were taken every hour until the cultures reached stationary phase and then samples taken every 24 hours at the point of inoculation for 3 days. CFU/ml were enumerated at 12, 24, 48 and 72 hours.

2.7.2 Co- culture experiments

2.7.2.1 Exponentially growing starting culture

Wild type spores were heat shocked as previously described (2.2.5) and plated on CBM agar whereas for non- or poorly sporulating mutants, glycerol stocks were used. Developing colonies were inoculated into 10 ml CBM 6% glucose liquid media and prepared as overnight pre-cultures as previously described (2.2.5). Overnight pre-cultures were measured to see which were growing at OD_{600} between 0.2-0.8. Respective cultures of wild type and a given mutant were mixed at various ratios so

that the final OD was 0.02. Each fermentation was performed in 60 ml CBM liquid medium in 100 ml conical flasks. 1 ml samples were taken every hour up to 12 hours and then samples taken every 24 hours at the point of inoculation for 3 days. CFU/ml were enumerated at 0, 12, 24, 48 and 72 hours. Monoculture controls for each strain were treated in the same way except that that the second inoculum was omitted and that they were inoculated to an OD of 0.02.

2.7.2.2 Stationary growing starting culture

Wild type spores were heat shocked as previously described (2.2.5) and plated on CBM agar and a desired degenerated strain was streaked on CBM agar. Colonies were inoculated into 10 ml CBM 6% glucose liquid and left to grow for 24 hours. 24 hour cultures of WT and the desired degenerate were added in varying volumetric ratios to make a total of 1 ml which was used to inoculate to 9 ml CBM 6% glucose liquid so that the final volume was 10 ml. For example, 100 µl degenerate isolate + 900 µl wild type added to 9 ml CBM-S broth 6% glucose. This would give a starting ratio of 1:10 (degenerate:WT) in the 1 ml. CFU/ml were enumerated at 0, 12, 24, 48 and 72 hours using serial dilutions of the culture. Fitness of individuals was calculated from $w = [x_2(1 - x_1)]/[x_1(1 - x_2)]$ from Diggle et al. (2007) and Ross-Gillespie et al. (2007) where x_1 is the initial proportion of mutant in the population and x_2 is the final proportion at a given time point.

2.7.3 pH-controlled bioreactor fermentations

2.7.3.1 Monocultures with exponentially growing starting cultures

Wild type spores were heat shocked as previously described (2.25) and plated on CBM agar and degenerated strains were steaked on CBM agar. Colonies were inoculated into 70 ml CBM 6% glucose liquid medium per strain, serially diluted to 10^{-4} in the same volume and left to grow overnight at 37 °C in the anaerobic cabinet. 20 ml of culture from a flask growing exponentially (OD_{600} between 0.2-0.8) was added to autoclaved 100-ml serum flasks and subsequently transferred to an autoclaved Multifors bioreactor vessel (Infors, Switzerland). The experiment was carried out in triplicate, resulting in the parallel running of 6 bioreactors (3 for wild type and 3 for the mutant) containing 330 ml anoxic CBM 6% glucose liquid medium, thus bringing the total volume to 350 ml. Cultures were grown at 37 °C with the pH maintained at 6. Growth curve samples for OD were taken every hour until the culture reached stationary phase and then subsequently every day after the point of inoculation. 1.5 ml samples were taken for CFU/ml, GC and HPLC analysis at 0, 6, 12, 24, 48, 76, 99 and 120 hours and processed as previously described.

2.7.3.2 Mixed culture stationary starting culture

Wild type spores were heat shocked as previously described (2.2.5) and plated on CBM agar and a degenerate mutant strain was steaked on CBM agar. Colonies were inoculated into three tubes per strain, each containing 35 ml CBM 6% glucose. The

resulting cultures were left to grow for 24 hours at 37 °C in the anaerobic cabinet. The resulting stationary phase cultures were used to generate mixed cultures containing wild type and mutant in different ratios in a culture volume of 30ml, contained within a conical flask. A total of six different mixed cultures generated in this way were transferred into autoclaved serum flask bottles and used to inoculate six Multifors bioreactor vessels (Infor, Switzerland) already containing 270 ml anoxic CBM 6% glucose liquid medium. Cultures were grown as previously described with 1.5 ml samples taken at 0, 12, 24, 48, 76, 99 and 120 hours for CFU/ml, GC and HPLC analysis.

2.8 *In silico* work and tools

2.8.1 Analysis and storage of raw numerical data

Raw numerical data was analysed and stored using Microsoft Excel 2013 (Microsoft). Graphical data and statistical analysis was presented and calculated using GraphPad Prism 7 (GraphPad software). Significant differences were determined via T-test and ANOVA analysis.

2.8.2 Plasmid, nucleotide and protein sequences

Benchling (www.benchling.com) and SnapGene viewer (www.snapgene.com) were used for plasmid construction and visualisations. Nucleotide and protein sequence were acquired from NCBI and sequences compared the Basic Local Alignment Search Tool (BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi). ClosTron plasmids were constructed using the intron assembly tool (http://clostron.com/clostron1.php).

Illustrator for Biological Sequences (Liu et al., 2015a) was used for protein domain visualisation. BAGEL3 (http://bagel.molgenrug.nl/) was used to determine potential bacteriocin genes.

2.8.3 Genome sequencing analysis

2.8.3.1 Genome sequence quality

Genome sequences were acquired from MicrobesNG (Birmingham, UK) and were subjected to trimming and the quality assessed using their in house software. Subsequent draft genome sequences are prepared but are not closed.

2.8.3.2 Read mapping

Paired-end reads were mapped against the published *C. beijerinckii* NCIMB 8052 reference genome (Wang et al., 2012) using CLC Genomics Workbench 9 (Qiagen) with the following parameters set by the read mapping analysis of CLC Genomics Workbench 9 (Qiagen):

Masking mode = no masking
Mismatch cost = 2
Cost of insertions and deletions = linear gap cost
Insertion cost = 3
Deletion cost = 3
Insertion open cost = 6
Insertion extend cost = 1
Deletion open cost = 6
Deletion extend cost = 6
Length fraction = 0.5
Similarity fraction = 0.8
Global alignment = No
Auto-detect paired distances = yes
Non-specific match handling = map randomly

2.8.3.3 Small nucleotide variant calling

Small Nucleotide Variants (SNVs) were called from the mapped reads using CLC Genomics Workbench 9 (Qiagen) with the following parameters created by Dr Grosse for the Mapping and Variants workflow:

Ploidy = 1
Ignore positions with coverage above = 100,000
Restrict calling to target regions = not set
Ignore broken pairs = yes
Ignore non-specific matches = reads
Minimum coverage = 10
Minimum count = 2
Minimum frequency (%) = 40.0
Base quality filter = yes
Neighbourhood radius = 5
Minimum central quality = 20
Minimum neighbourhood quality = 15
Read direction filter = no
Relative read direction filter = yes
Significance (%) = 1.0
Read position filter = no
Remove pyro-error variants = yes
In homopolymer regions with minimum length = 5
With frequency below = 0.7

3. Isolation and phenotypic characterisation of degenerates derived from *Clostridium beijerinckii* NCIMB 8052

3.1. Introduction

Microorganisms have been utilised for a variety of industrial bioprocesses whether it be the production of food for humans and animals, fuel for our vehicles or speciality chemicals (Jones and Woods, 1986; Tamang et al., 2016; Wackett, 2015; Zeikus, 1980). Industrial organisms are often derived from specifically selected strains, through either genetic or evolutionary selection (Debabov, 2015; Lütke-Eversloh and Bahl, 2011), to be used either to produce a certain product or aid a particular bioprocess (Stanbury et al., 2017). As described in Chapter 1, solventogenic *Clostridia* were used throughout the early to mid-1900s to make the solvents acetone, butanol and ethanol until the advent of the petrochemical industry saw their use decline (Jones and Woods, 1986). With a renewed shift towards utilising these organisms once again, challenges must be overcome to make their use economically viable. Maintaining culture stability is one such challenge as it is integral to allow for maximum efficiency and production of the desired product or process (Hunter-Cevera and Belt, 1996). Culture viability can theoretically be maintained through repeated subculturing into fresh medium as this allows for constant re-growth of culture. The constant exponential growth of a culture however runs the risk of strain degeneration as with continued cell division there is a chance that random mutations establish themselves in the genome, resulting in the propagation of undesired mutants (Stanbury et al., 2017).

Strain degeneration of solventogenic *Clostridia* is one such example of this. This phenomenon is characterised by a reduction or complete loss in solvent and endospore production making it a significant problem for the commercialisation of these organisms (Jones and Woods, 1986; Kashket and Cao, 1995; Kutzenok and Aschner, 1952; Woolley and Morris, 1990). Strain degeneration is well documented

to occur through serial transfer of a clostridial culture, a technique that is employed in industrial batch culture fermentations (Grimbert, 1893; Hartmanis et al., 1986; Jones and Woods, 1986; Kashket and Cao, 1995). To alleviate the problem of degeneration in batch fermentations, it became standard practice to use freshly germinated spores at each successive inoculation (Stephens et al., 1985). Batch fermentations are associated with several problems such as significant down time between fermentations due to cleaning for contamination removal, low reactor productivity and product inhibition (Jones and Woods, 1986; Qureshi and Maddox, 1995; Yang and Tsao, 1995). Whilst many microbial fermentations, including ABE fermentation, were operated in batch, a continuous process offers several advantages over batch fermentations and industry is shifting towards them. These include increased product yield and uniformity, reduction in processing time, reduction in equipment size (and thus capital expenditure) and finer instrument control to allow the process to be run more economically (Maxon, 1955; Ni and Sun, 2009) . Although continuous fermentations has its advantages over batch, it should be noted that strain degeneration of solventogenic *Clostridia* also occurs in continuous fermentations exemplifying the problematic nature of this phenomenon (Finn and Nowrey, 1959; Stephens et al., 1985; Woolley and Morris, 1990).

Besides the loss of solvent and spore production, changes in cell physiology and colony morphology were found to be distinct characteristics of strains believed to have degenerated (Adler and Crow, 1987; Jones et al., 1982a; Kashket and Cao, 1995; Kutzenok and Aschner, 1952; Woolley and Morris, 1990). Several unusual colony morphologies arose when colonies or liquid cultures were subcultured and when colonies were left to grow for extended periods of time (Adler and Crow, 1987; Kashket and Cao, 1995). Changes in cell physiology have also been observed when mutagens were used to isolate strains that showed the characteristics of a degenerate

(Jones et al., 1982a). It should be noted however that degeneration has been shown to be a passive process that does not require the use of mutagens (Adler and Crow, 1987; Kashket and Cao, 1995; Woolley and Morris, 1990). A greater understanding of why and how *Clostridium* strains undergo degeneration is needed to allow the phenomenon to be controlled or prevented. To allow for further study into clostridial degeneration, isolation and characterisation of a vast variety of degenerated strains is first needed.

3.2. Aims of this study

The aims of this study were to firstly to design and implement a reliable method for the isolation of degenerates from *C. beijerinckii* NCIMB 8052 so that characterisation of a variety of different clones and lineages could be attempted. *C. beijerinckii* NCIMB 8052 was chosen as it has been shown to be particularly prone to degeneration and has previously been studied extensively in degeneration studies (Kashket and Cao, 1995; Stephens et al., 1985; Woolley and Morris, 1990). Furthermore this strain has been used for industrial ABE fermentations and is the parental strain for the hyper butanol producer *C. beijerinckii* BA101 (Chen and Blaschek, 1999a). Once a consistent method had been established, degenerate isolates would be tested in regards to their solvent, acid, granulose and spore producing capabilities to assess the level of degeneration they had undergone. Following phenotypic characterisation, genetic analysis (Chapter 4) would allow comparisons between strains phenotypes and genotypes. It was also of interest to track the fermentation profile of a culture through repeated subculturing and to determine the composition of the culture population. The specific aims of this study were therefore as follows:

- To phenotypically characterise the *C. beijerinckii* NCIMB 8052 strain obtained from the NCIMB culture collection to establish its typical growth and product profile under the employed conditions. This would establish benchmark data for comparisons to potential degenerates
- To establish a robust, reproducible method of producing degenerates derived from *C. beijerinckii* NCIMB 8052
- To use this method to independently isolate a large number of different degenerate clones and to characterise them phenotypically as outlined above
- To repeatedly subculture *C. beijerinckii* NCIMB 8052 to observe the process of degeneration and its effects on the whole culture

3.3. Results

3.3.1. Phenotypic characterisation of wild type *Clostridium beijerinckii* NCIMB 8052

The overall goal of the work described in this chapter was to obtain degenerate mutant strains from wild type *C. beijerinckii* NCIMB 8052. This strain was chosen as it is a well-established model strain and also the immediate progenitor of a high-butanol producing variant (Adler and Crow, 1987; Kashket and Cao, 1995; Shi and Blaschek, 2008; Stephens et al., 1985; Woolley and Morris, 1990). It was decided that for this work a new stock culture should be obtained from the National Collection of Industrial Food and Marine Bacteria (NCIMB) to stay as close to the original source as possible. Once obtained, this new culture was still considered to represent a potentially heterogeneous population as its history in terms of cultivation was unknown and

mutant derivatives may have arisen at some point. To obtain a homogenous, clonal lineage it was decided to select a single colony whose phenotypic traits matched those described in the literature and that of other clones tested. A master spore stock would then be prepared so that all future degenerate comparisons could be traced to this tested, confirmed and sequenced clonal parental stock.

Various fermentation analyses of wild type *C. beijerinckii* NCIMB 8052 exist in the literature (Chen and Blaschek, 1999a; R. W. O'Brien & J. G. Morris, 1971; Woolley and Morris, 1990) but proper analysis of our stock under the conditions employed here was needed for future comparisons. Once obtained from the NCIMB, the stock of *C. beijerinckii* NCIMB 8052 was grown for five days to create a heterogeneous spore stock that was harvested by centrifugation. Isolation of our wild type from a spore stock was chosen as degenerate strains are known to have reduced or abolished sporulation (Kashket and Cao, 1995). From this spore stock, three individual colonies, germinated from independent spores were picked and inoculated in to 10 ml CBM-S broth 6% glucose and diluted as pre-cultures and left overnight. These pre-cultures were used to inoculate technical quadruplicates of 60 ml CBM-S broth 6% glucose for the main fermentation. Growth curves, pH profiles alongside acid and solvent concentrations produced over the fermentation period were analysed.

As can be seen from Figure 3.1, the growth and pH profiles for these cultures were highly similar.

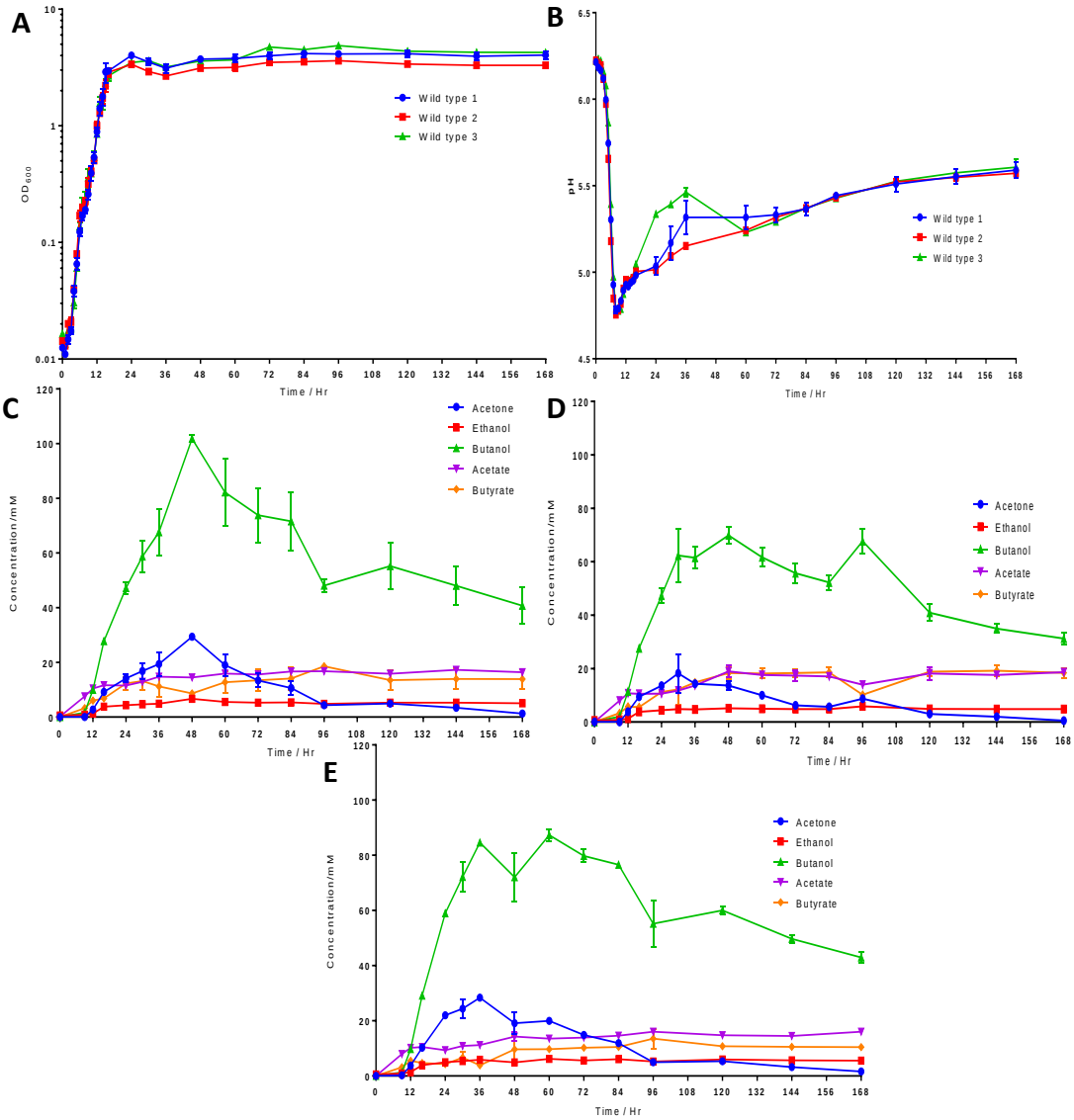


Figure 3.1 Growth, pH and solvent profiles for three independent clones of *C. beijerinckii* NCIMB 8052 wild type.

Growth (OD600) (A), pH profile (B) and main fermentation products produced by clone 1 (C), clone 2 (D), and clone 3 (E). Clones were obtained from a spore stock following germination and formation of single colonies. Clonal strains were grown in CBM-S liquid 6% glucose. Each colony was used to inoculate a separate starter culture which was then used to inoculate replicate main cultures. Technical replicates represented by standard error of the mean. n =4.

From hours 0 to 9, only acids were produced which corresponds to the drop in pH which reached a minimum of 4.75. After 9 hr, the production of butanol, acetone and ethanol was observed, indicating that the metabolic switch to solventogenesis had taken place. The maximum OD_{600} of each of the replicates is reached at around 24 h where they plateaued at around 4. A reduction of OD_{600} can be seen between hours 24-48, potentially due to cell aggregation or sporulation. At this point the OD_{600} stabilises as the cultures are in stationary phase. In the later stages of the fermentation the concentrations of solvents decreased, presumably because fermentation metabolism was completed and the solvents evaporated. All cultures reached a final pH of around 5.55.

Overall, the three biological replicates produced similar fermentation product profiles, although replicate 3 had the greatest initial increase in pH during the switch to solventogenesis and replicate 1 reached the highest butanol concentrations with a maximum of around 100 mM, which then declined more rapidly. All future work used replicate 3 as the parental strain as in terms of solvent production it appeared to have the least variation in the technical replicates.

3.3.2. Design and implementation of a subculturing protocol and the effects on *C. beijerinckii* NCIMB 8052 culture.

To gain a further understanding of the degeneration process it was first necessary to design and implement a robust and reliable method for producing degenerates of *C. beijerinckii* NCIMB 8052. The occurrence of degenerate strains during repeated sub-culturing is well documented in the literature (Adler and Crow, 1987; Kashket and Cao, 1995; Stephens et al., 1985; Woolley and Morris, 1990). A robust procedure was needed that allowed isolation of degenerates very soon after they had emerged to

avoid accumulation of additional mutations which may obscure the effects of those responsible for degeneration. In preliminary experiments two approaches were tested, the re-streaking of colonies on plates and repeated subculturing in liquid broth, followed by plating of dilutions (data not shown). Whilst more laborious, the latter allowed the screening of larger numbers of colonies and yielded degenerate colonies after very few subcultures depending on the inoculation regime. This procedure is outlined in Figure 3.2 and was used throughout this study to reproducibly produce degenerates in a relatively short space of time.

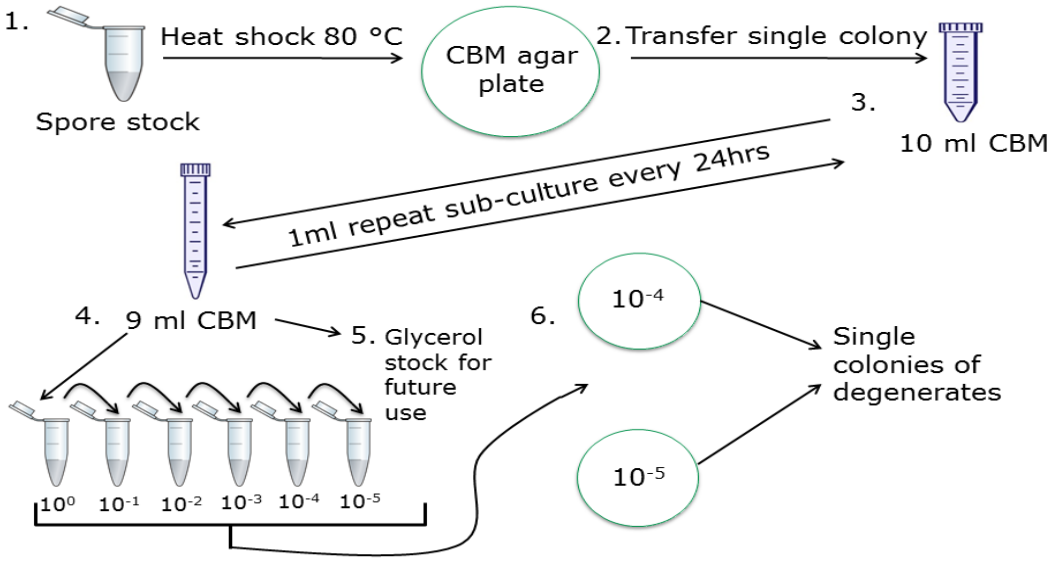


Figure 3.2 Schematic for the isolation of degenerates derived from wild type *Clostridium beijerinckii* NCIMB 8052.

The procedure is described in detail in Materials and Method section 2.5 but in brief; Stage 1 involved heat shocking wild type C. *beijerinckii* NCIMB 8052 spores and plating on CBM-S agar 6% glucose. Single colonies were transferred to 10 ml of CBM-S 6% glucose and left to grow for 24 hours at stage 2. Stage 3 saw the repeated subculturing (5 to 15) of the culture every 24 hours. Serial dilutions of the culture were carried out at stage 4 and stocks taken at stage 5. The serial dilutions were plated on CBM-S agar 6% glucose and colonies observed at stage 6.

A batch fermentation-style transfer every 24 hours was also employed for degenerate isolation as it is well documented that degeneration occurs during batch fermentations and the use of 24 hours has previously been described in the literature (Barbeau et al., 1988; Finn and Nowrey, 1959; Gapes et al., 1996; Hartmanis et al., 1986; Hayashida and Yoshino, 1989). Due to the lack of high-throughput screening facilities for strictly anaerobic bacteria, initial detection of degenerates could not be based on solvent profiles or lack of spore formation. Instead, changes in colony morphology were used as an indicator as previously described (Adler and Crow, 1987; Kashket and Cao, 1995).

The isolation process involved heat shocking a wild type spore stock (from clone 3) to both stimulate spore germination and remove any potential degenerates/vegetative cells that may be in the suspension. After germination, a single colony was picked and inoculated into fresh liquid CBM-S 6% glucose and left to grow for 24 hours. A single colony was used so that all subsequent degenerates isolated came from a single parental colony allowing for future genetic comparisons. After 24 hours of growth, the culture was subcultured every 24 hours in a 1:10 dilution to allow for the emergence of degenerates. Degenerates isolated for future analysis in this study were obtained after 5 subcultures and picked after serial dilutions of the culture. For the study of culture degeneration dynamics, subculturing was sometimes extended to up to 15 transfers.

When observing colony morphology at each successive subculture, a striking change was seen as an increasing array of different colony morphologies emerged that all appeared different to the original wild type colonies. Figure 3.3 shows examples for this mixture of colony types at various subculture stages.

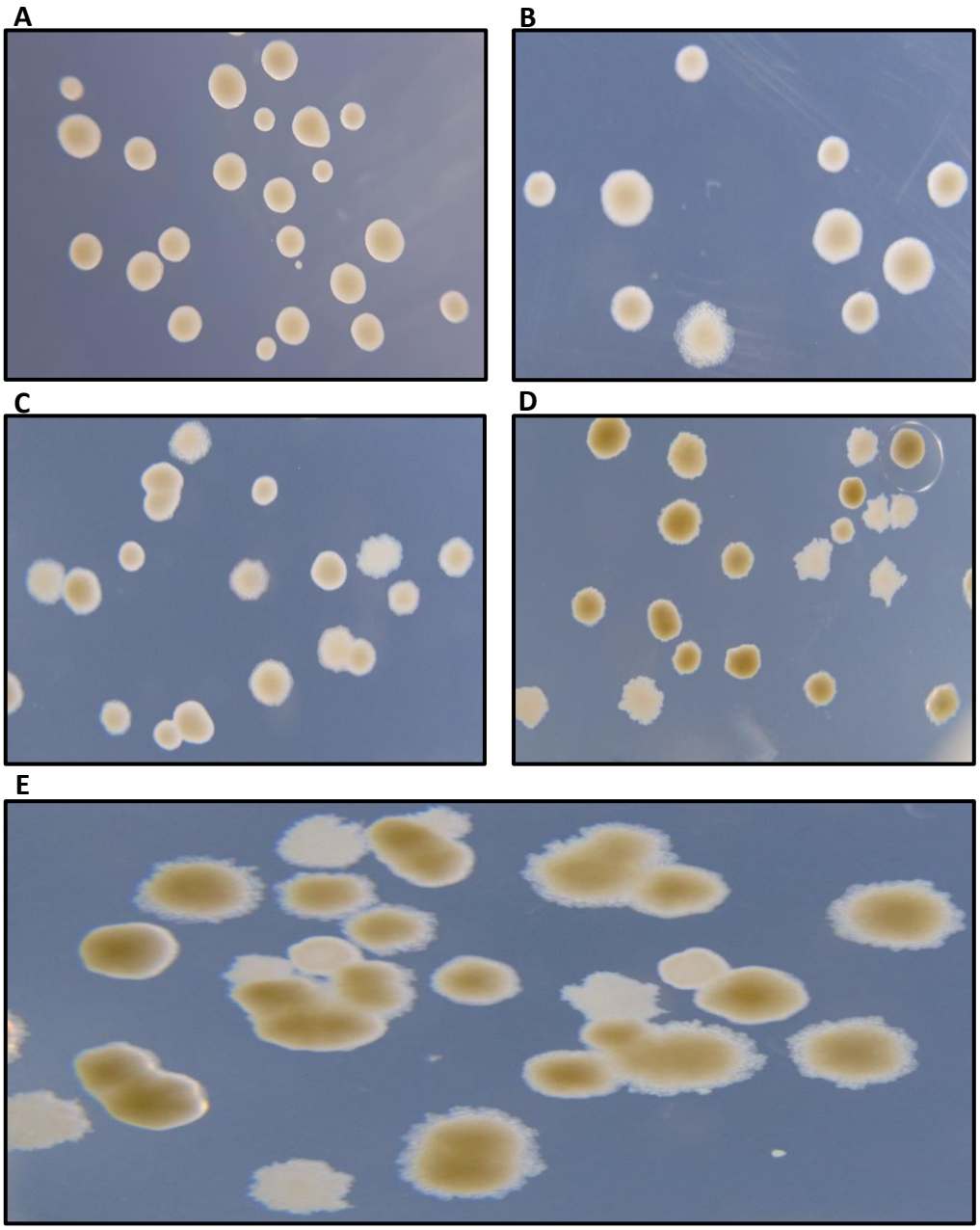


Figure 3.3 Colonies observed at several progressive subcultures visualised on CBM-S agar.

Culture heterogeneity was assessed at different stage of the subculturing regime as described in Figure 3.2 by plating on CBM-S agar lacking calcium carbonate to facilitate visual inspection. Colony morphologies obtained from (A) the starter culture (0 subcultures), (B) subculture 2, (C) subculture 4, (D) subculture 6 and (E) subculture 8.

It was found that successive subcultures led to varying numbers of different colony morphologies as the transfers progressed. It was also observed that different colony types emerged, took over the population and were subsequently replaced by another type as described in section 3.3.5.1.

Before subculturing the culture appeared homogeneous and comprised only of wild type morphologies that appeared round, smooth and slightly dark cream-like in colour (as can be seen Figure 3.3A). After one subculture, small numbers of colonies emerged that had a dark creamy centre with diffuse, translucent outgrowths that protruded out from the colony (Figure 3.3B). During the early stages of subculturing (1-3 subcultures), the number of colonies which contained these outgrowths increased rapidly and began to take over the population. This rapid take over subsequently saw a decline in the number of wild type morphologies. In addition, subculture 2 onwards saw the emergence of two other colony types alongside the outgrowth colony type but in relatively small numbers. One type appeared to have a translucent base layer topped with a dark donut ring shape and a second that appeared flat, white and diffuse. By 8 subcultures, as seen in Figure 3.3E, the culture population appeared as a 'zoo' of four different morphotypes all distinct from each other. When subculturing was continued past this point, the colonies that appeared flat, white and diffuse eventually took over the population leaving little to none of the other types. This complete takeover of the culture usually occurred when subculturing was continued past 12 subcultures and restoration of other types was never seen (Figure 3.11).

The four distinct colony types identified were named as "round and dark" (RD) which were assumed to be wild type, the "dark centre with outgrowths" (DCOG) which were the first colony type to emerge, the "caved in centre" (CIC) which appeared soon after

and the “flat and white” (FW). Several examples of the colony types found can be seen in Figure 3.4.

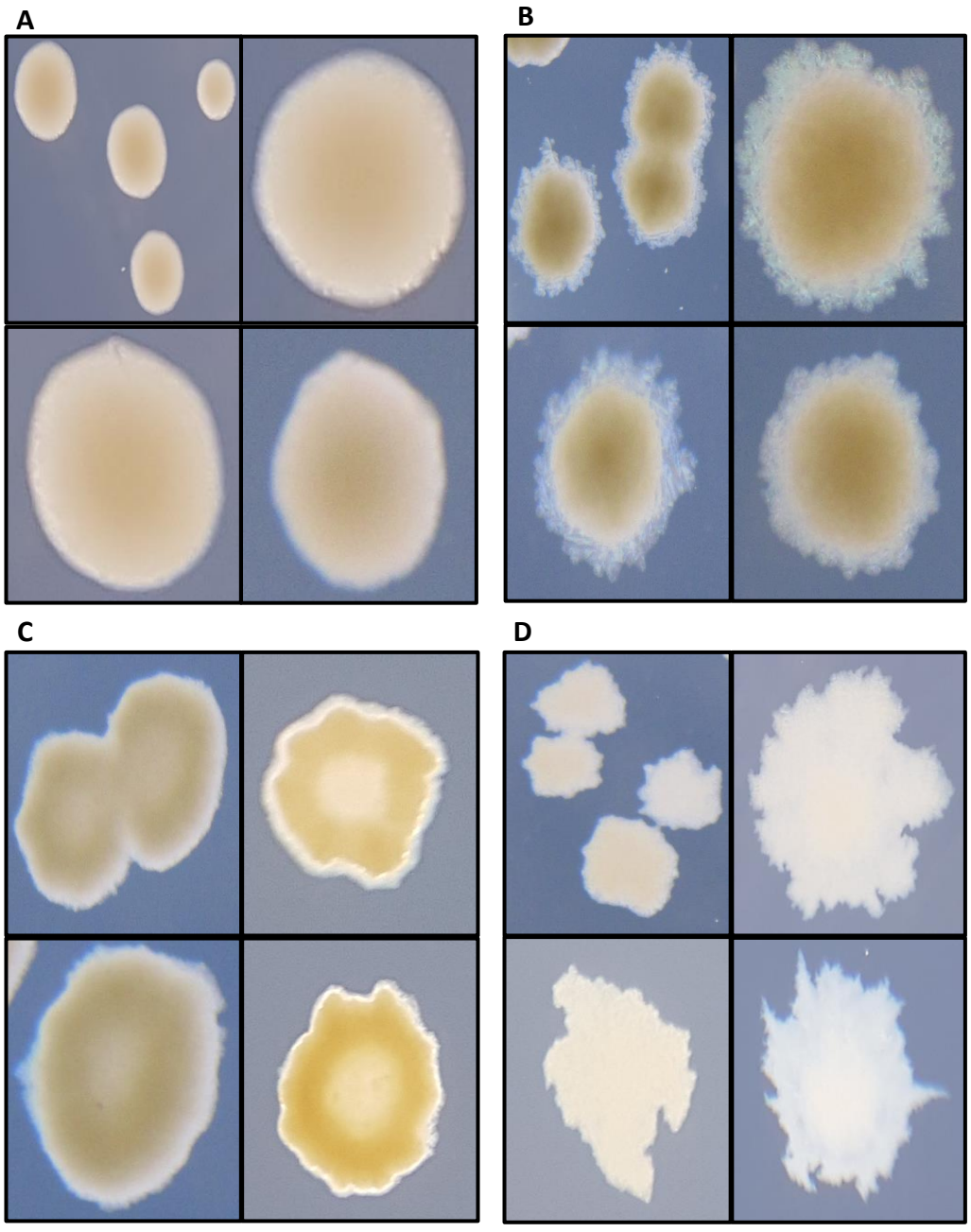


Figure 3.4 Colony types observed after repeated subculturing of *C. beijerinckii* NCIMB 8052

Four different colony types were observed: (A) “round and dark” (RD), (B) “dark centre and outgrowths” (DCOG), (C) “caved in centres” (CIC) and (D) “flat and white” (FW).

3.3.3. Isolation of degenerates of *C. beijerinckii* NCIMB 8052 for future analysis

Degenerates for future phenotypic and genotypic analysis were isolated in two separate rounds using the regime outlined in Figure 3.2. Each round of degenerate isolation used a single parental wild type colony, derived from a spore stock of clone 3, which was resuspended in liquid medium and distributed to 6 separate tubes for subculturing. Genomic DNA was isolated (as described in Materials and Method section 2.3.1.2) from each parent culture to allow for genomic comparisons. Potential degenerate colonies (from the four types previously described) were isolated after 5 subcultures. In the first round of subculturing 5 RD, 5 DCOG, 6 CIC and 7 FW morphology types were isolated. The second round yielded 12 RD, 12 DCOG, 12 CIC and 12 FW morphotypes. This gave a total of 17 RD, 17 DCOG, 18 CIC and 19 FW morphotypes that were to be used for the entirety of this PhD project.

To confirm that the mutants isolated were not contaminations, a diagnostic PCR was performed using the primers 8052 LAC_F and 8052 LAC_F. These primers amplified a lactate dehydrogenase gene (Cbei_1632) specific to *C. beijerinckii* NCIMB 8052 and therefore proving that the degenerates were derivatives of that strain. Control DNA of WT *C. beijerinckii* NCIMB 8052, *C. acetobutylicum* ATCC 824 and *C. beijerinckii* 59B were tested to confirm that the primers would only amplify *C. beijerinckii* NCIMB 8052 DNA (data not shown). All degenerate isolates were confirmed to be derivatives of *C. beijerinckii* NCIMB 8052 using these primers (data not shown).

3.3.4. Phenotypic characterisation of isolated degenerates

3.3.4.1. Solvent and acid production

Each isolated degenerate strain was grown alongside the parental wild type (clone 3) in 25 ml liquid CBM-S 6% glucose with solvent and acid samples taken after 2 days and 5 days post inoculation. Cultures were grown in technical triplicates and the experiment was repeated twice to yield a total of 6 data sets per isolate. This allowed for a fermentation profile for each of the isolated colony representatives to be created. The solvent and acids profiles are shown in Figures 3.5 to 3.8, for colony morphotypes RD, DCOG, CIC and FW, respectively. A summary of the statistical difference between degenerates and wild type fermentation products as determined by ANOVA can be found in the appendix section 7.2.

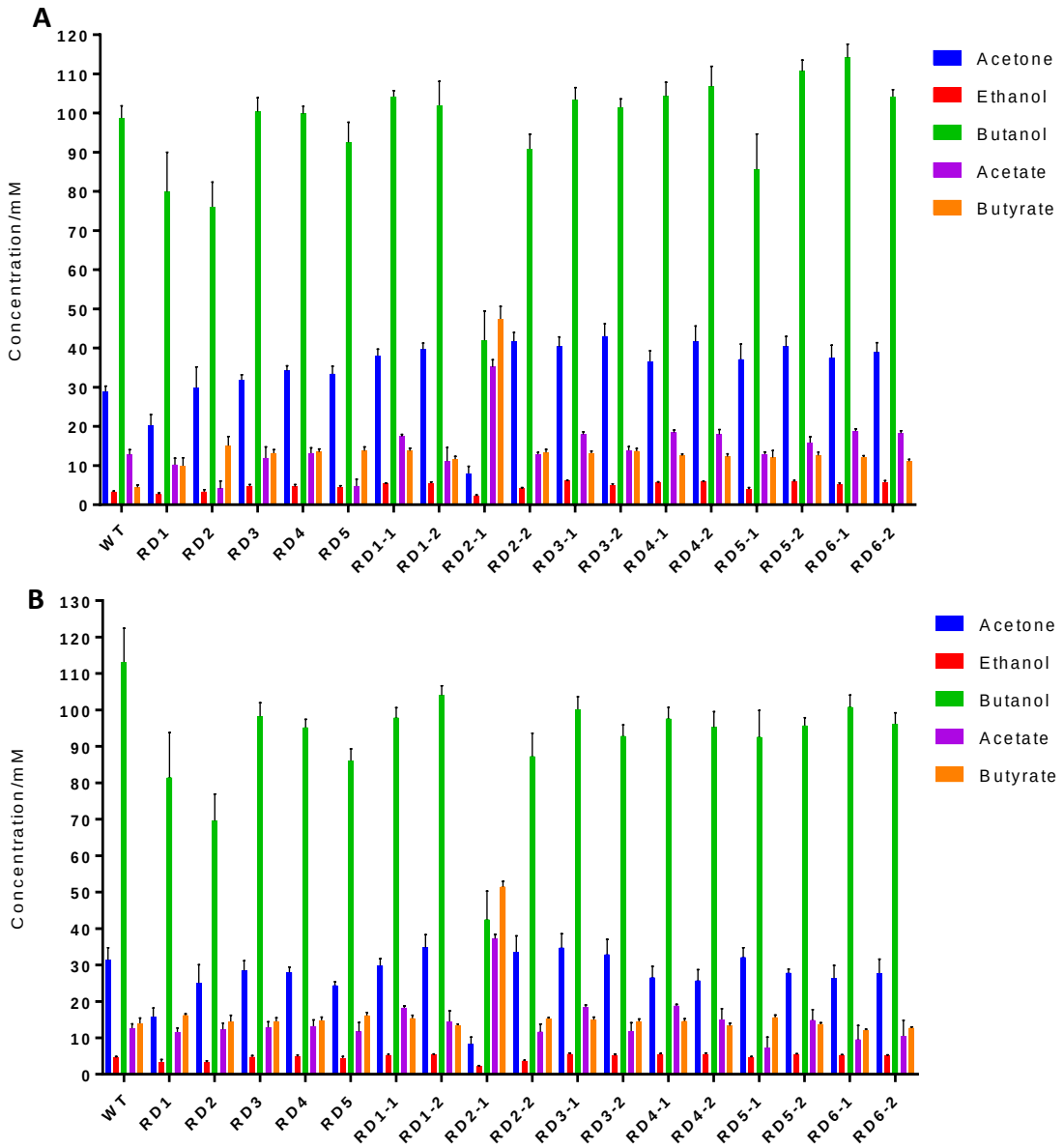


Figure 3.5 Solvent and acid concentrations at days 2 and 5 for cultures of the round and dark (RD) colony type.

Fermentation product profiles after (A) 2 days (48 hr) and (B) 5 days (120 hr) of growth in 25 ml CBM-S broth 6% glucose. Technical replicates represented by standard error of the mean. *n* = 6 for both graphs. Results were analysed using two-way ANOVA and can be found in appendix section 7.2.

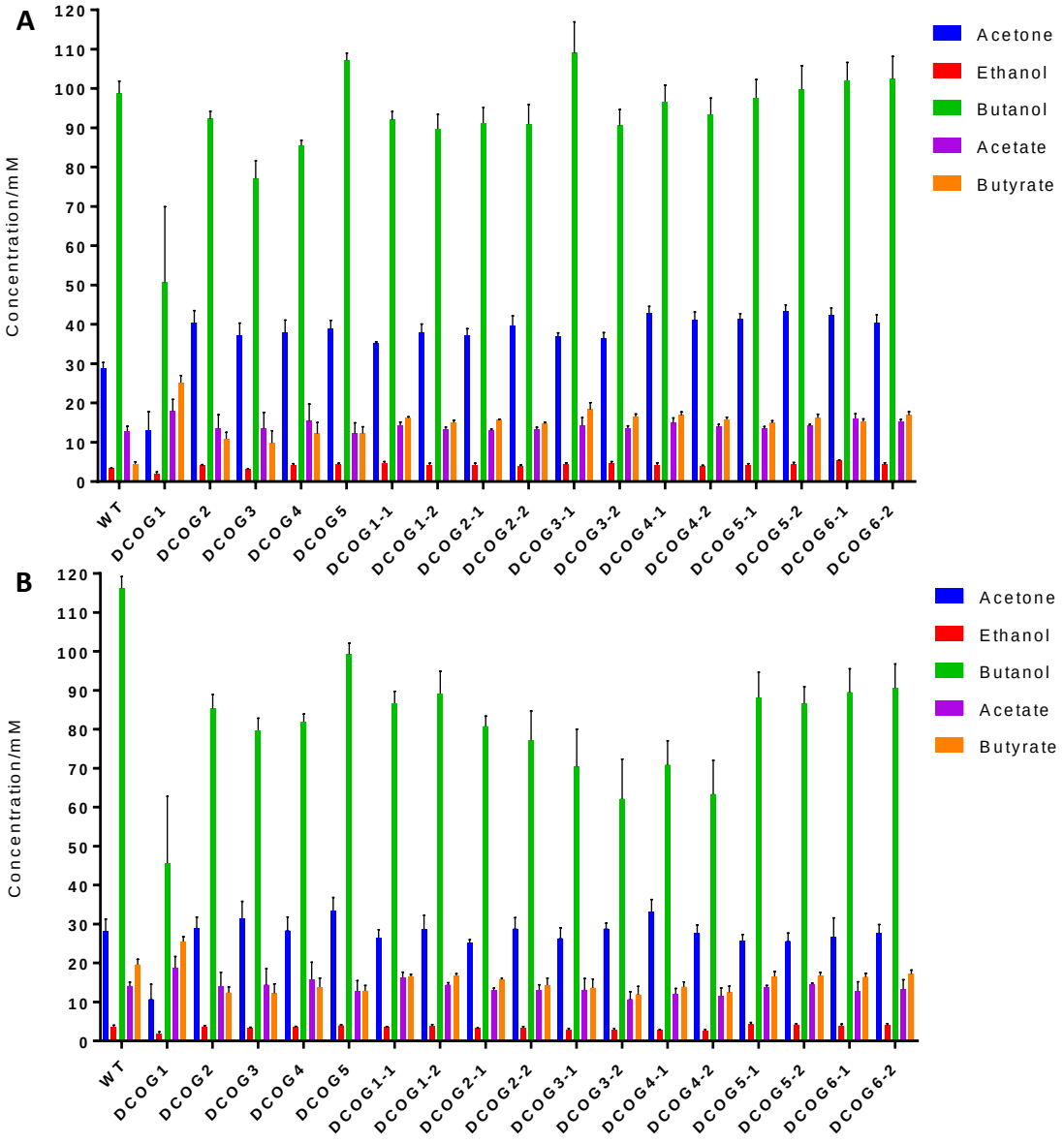


Figure 3.6 Solvent and acid concentrations at days 2 and 5 for cultures of the dark centre with outgrowths (DCOG) colony type.

Fermentation product profiles after (A) 2 days (48 hr) and (B) 5 days (120 hr) of growth in 25 ml CBM-S broth 6% glucose. Technical replicates represented by standard error of the mean. *n* = 6 for both graphs. Results were analysed using two-way ANOVA and can be found in appendix section 7.2.

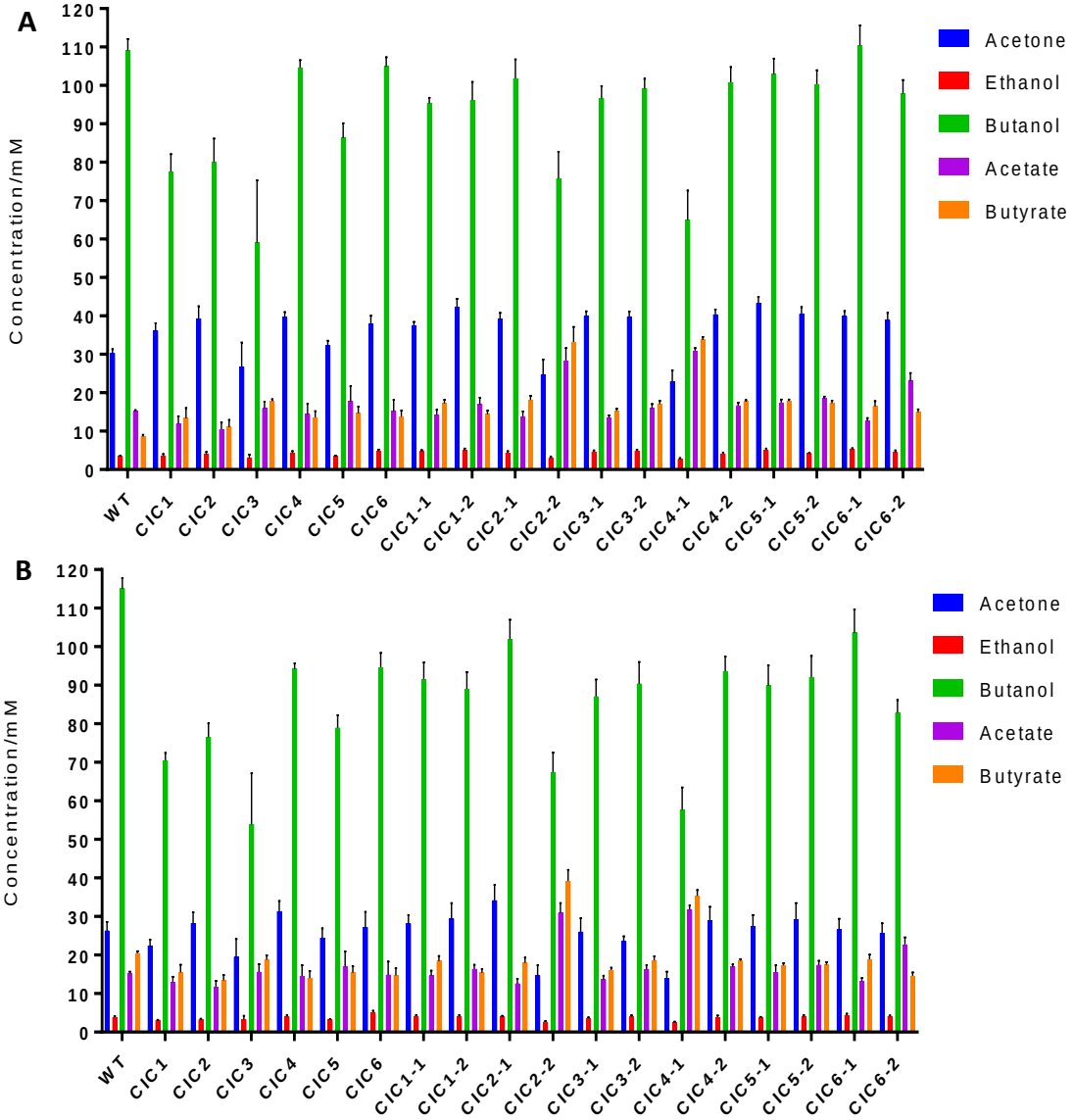


Figure 3.7 Solvent and acid concentrations at days 2 and 5 for cultures of the caved in centre (CIC) colony type.

Fermentation product profiles after (A) 2 days (48 hr) and (B) 5 days (120 hr) of growth in 25 ml CBM-S broth 6% glucose. Technical replicates represented by standard error of the mean. *n* = 6 for both graphs. Results were analysed using two-way ANOVA and can be found in appendix section 7.2.

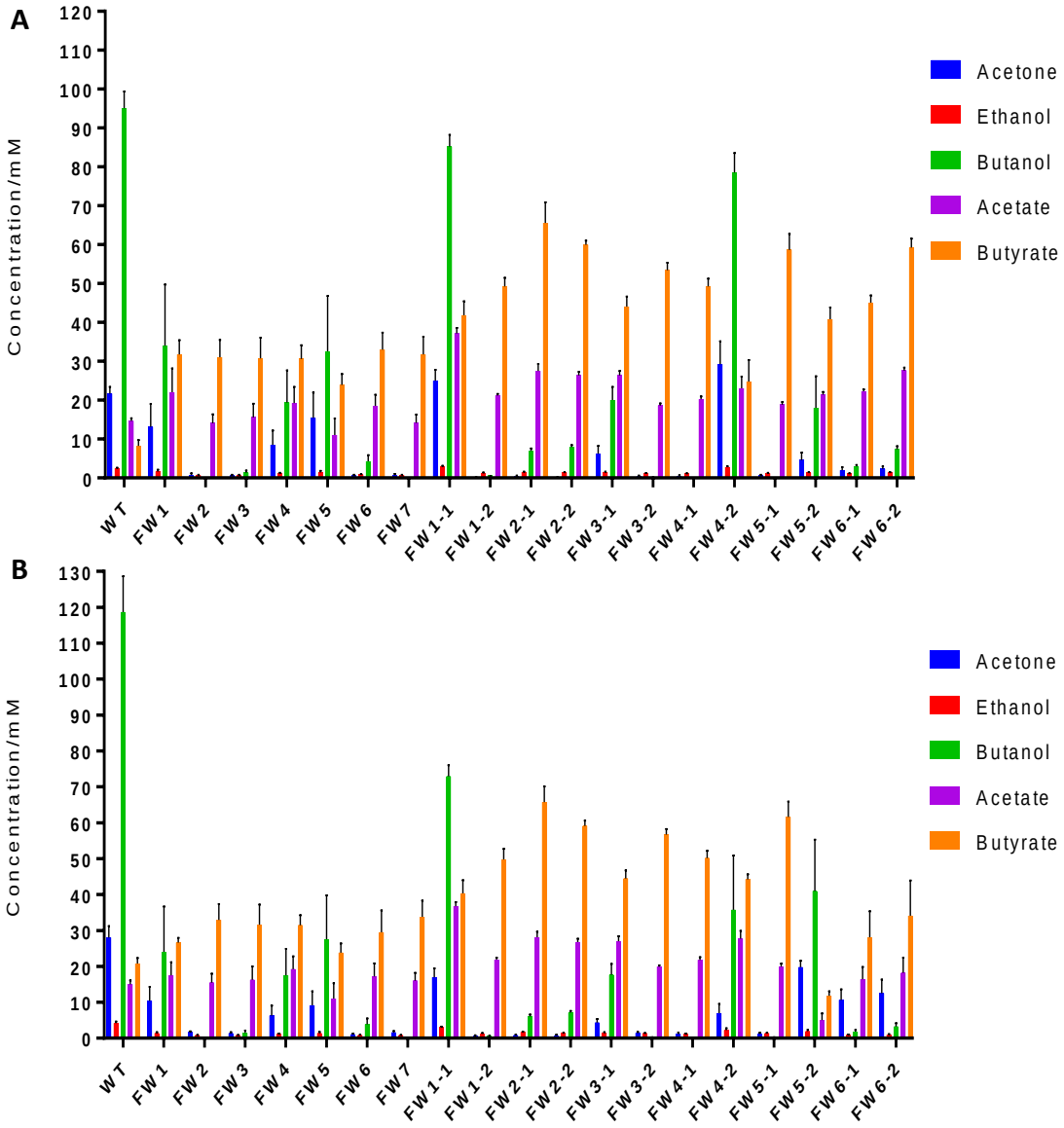


Figure 3.8 Solvent and acid concentrations at days 2 and 5 for cultures of the caved in centre (CIC) colony type

Fermentation product profiles after (A) 2 days (48 hr) and (B) 5 days (120 hr) of growth in 25 ml CBM-S broth 6% glucose. Technical replicates represented by standard error of the mean. *n* = 6 for both graphs. Results were analysed using two-way ANOVA and can be found in appendix section 7.2.

Solvent and acid profiles for each of the degenerate types were much more varied than expected. It was assumed that colonies that appeared like the wild type would produce similar concentrations of solvents and acids however this was not always the case. From Figure 3.5 several of the RD types produced varying butanol concentrations to the wild type after 2 days of fermentation but overall levels were similar for most isolates. RD1, RD2, RD2-1, RD5-1 all produced significantly less butanol whereas RD5-2 and RD6-1 produced more than the wild type, the latter having generated 114.2 ± 3.4 mM. Differences for most isolates did not exceed 10-20 mM, but in the case of RD2-1 the concentration of butanol was nearly 60 mM less than the wild type at the 2 day fermentation sample point. This reduction in butanol production coincided with the production of significantly more acids than the wild type suggesting a strain that had degenerated in some form. It was also interesting to observe that RD2-2, RD3-1, RD3-2, RD4-2 and RD5-2 all produced statistically more acetone (around 10 mM more) than the wild type, which had produced about 30 mM, even when the concentrations of butanol did not differ to that of the wild type. An interesting trend was apparent for 2-day acid concentrations: all RD isolates produced more butyrate (at least 2 times) than the wild type, whereas acetate concentrations for most isolates were similar or increased to a lesser extent. A notable exception were RD2 and RD5, which had produced less acetate.

At the 5 day end point of the fermentation, a distinct difference in butanol concentrations was observed. All strains except RD1-2 and RD6-1 had produced a reduced concentration of butanol to that of the wild type at this final sample point. This appears to be reflecting the fact that the wild type continued to produce around 10 mM more butanol after the 2 day point where as other strains did not. All isolates with the exception of RD1-2 and RD6-2 produced statistically different concentrations of butanol to the wild type. However, evaporative loss of butanol may have been a

contributing factor for some of the reduced concentrations observed, for instance the 10 mM reduction seen for RD6-1. The only strains to have statistically different acetone to the wild type were RD1 and RD2-1. Acid concentrations, with some exceptions (RD2-1) were largely similar for wild type and mutants.

Comparisons of the DCOG types isolated and the wild type yielded a picture very similar to that seen for RD isolates. With exception of DCOG1, DCOG3 and DCOG4 all other strains produced similar concentrations of butanol after 2 days of growth. DCOG1 showed a similar fermentation profile to that of RD2-1 whereby solvent concentrations were significantly reduced and acid concentrations increased compared to the wild type. DCOG3 and DCOG4 produced slightly less butanol than the wild type whereas all other strains were not significantly different. Generally, acetate levels were similar to those of the wild type or slightly higher, whereas butyrate was increased for all DCOG isolates, leading to increased overall acid production at this time point. Similarly, acetone concentrations were increased for all DCOG isolates, by about 8 to 10 mM depending on the isolate. After 5 days, as seen for of the RD strains, concentrations of butanol and acetone were reduced for most isolates when compared to the 2-day time point. Again, however, this appears to have been due to the wild type continuing to produce butanol when the other strains had stopped, thus allowing evaporative losses to become apparent.

For the CIC isolates there was more variation in butanol production at the 2-day point of fermentation, although levels for half of them were still comparable the wild type. CIC1, CIC2, CIC3, CIC5, CIC1-1, CIC1-2, CIC2-2, CIC3-1 and CIC4-1 all produced statistically significantly less butanol than the wild type. Less variation was seen with respect to strains acid producing capability. However, as for RD and DCOG isolates, 2-day butyrate concentrations were consistently increased, with some isolates also

showing increased acetate (CIC2-2, CIC4-1, CIC6-2) and there was a trend towards increased acetone production, although with notable exceptions (CIC3, CIC2-2, CIC4-1). Again, at the 5-day point of the fermentation, most isolates cultures showed significantly less butanol compared to the 2-day measurements. As for the RD and DCOG experiments, this appears due to the continued production of butanol by the wild type and evaporative loss.

The most striking fermentation profiles were observed when the FW types were grown alongside the wild type. FW1-1 was the only isolate to produce butanol that was not statistically different to the wild type however FW4-2 was only marginally smaller at the 2-day sample point. Many strains produced little to none butanol after 2 days. Acetone concentrations were also reduced for all but FW1, FW4, FW5, FW1-1 and FW4-2, for the latter two in line with their still comparatively high butanol production. All strains produced significantly higher concentrations of total acids. In particular butyrate was considerably increased, reaching concentrations of 60 mM and more for some strains (FW2-1, FW2-2, FW5-1 and FW6-2). It would appear the metabolism of these isolates favoured the production of butyrate over acetate. Again, the 5-day concentrations showed presumably evaporative losses for butanol and acetone whereas the wild type displayed slightly increased concentrations. As already seen in Figures 3.5 to 3.7, butyrate concentrations were higher for the WT after 5 days of cultivation and, as a result, the only isolates to have statistically different butyrate concentrations were FW1-1, FW1-2, FW2-1, FW2-2, FW3-1, FW3-2, FW4-1, FW4-2 and FW5-1. As suggested earlier for most RD, DCOG and CIC types, perhaps their fermentation had ended early whilst the wild type continued production of this acid beyond day 2. With the exception of some outliers, the FW isolates were clearly the most degenerate type as they produced the least amount of butanol and the largest amount of acids.

3.3.4.2. Spore formation

To ascertain the number of viable spores and therefore the level of sporulation of each isolate in the growth experiment described in 3.3.4.1, 1 ml samples were taken 5 days post inoculation for each strain including a wild type comparison. A 100 µl of each sample was heat shocked to stimulate the germination of spores, serially diluted and subsequently enumerated after 24 hours of growth on CGM agar as described in 2.7.6. The resulting spore profiles can be seen in Figure 3.9.

As with the fermentation product profiles, some general trends could be observed for the different isolate types. For the RD isolates (Figure 3.9A), the only isolate to have comparable spore numbers to the wild type was RD1 with approximately 10^8 heat-resistant CFU/ml. All other isolates produced considerably less spores, in most cases between 10^3 to 10^4/ml. It was interesting to find that two RD isolates, RD4-1 and RD4-2, produced no spores under the conditions used even when their solvent producing capability seemed normal (see Figure 3.5).

None of the DCOG, CIC and FW type isolates produced spores at a level comparable to the wild type. Otherwise, the general trend for DCOG (Figure 3.9B) and CIC isolates (Figure 3.9C) resembled that of the RD types, with most isolates showing spore concentrations of around 10^3 to 10^4/ml and a few that produced no detectable spores. For instance, DCOG2 and DCOG5 produced no detectable heat resistant CFUs and again even when their solvent profiles looked relatively normal with DCOG5 even producing slightly more butanol than the wild type at the 2 day sample point. Similarly, CIC3 and CIC4-1 did not produce any spores, although these isolates had produced reduced solvent concentrations and, in the case of CIC4-1, increased acids (Figure 3.7).

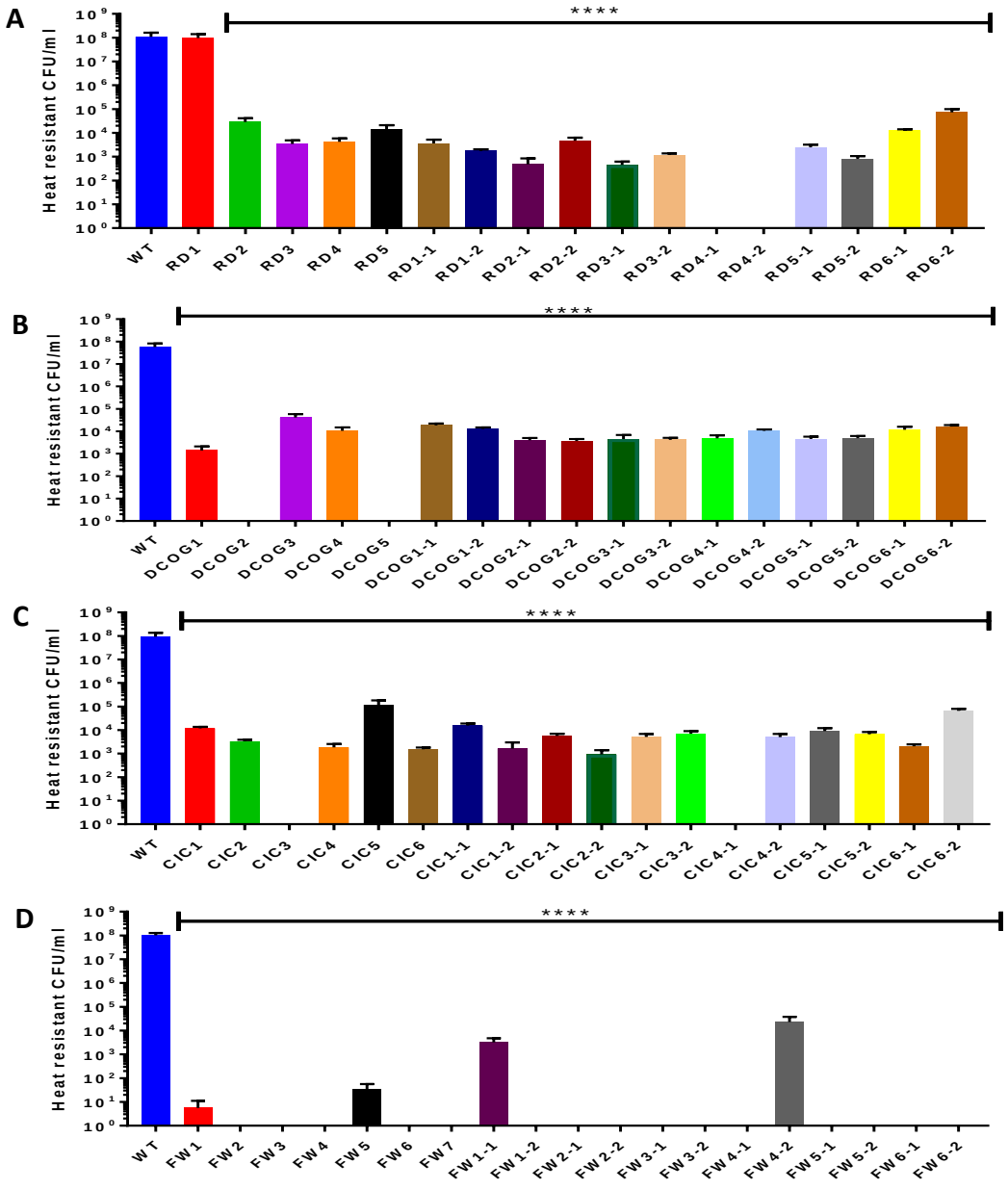


Figure 3.9 Sporulation capacity of degenerate isolates.

The number of heat-resistant CFUs/ml after 5 days of growth in CBM-S 6% broth glucose was taken as a measure for viable endospore formation for (A) RD, (B) DCOG, (C) CIC and (D) FW type isolates. Technical replicates represented by standard error of the mean. *n* = 6 for all graphs. Results were statistically analysed using one-way ANOVA against the wild type control giving P values of ≤ 0.0001 (****) for all isolates apart from RD1.

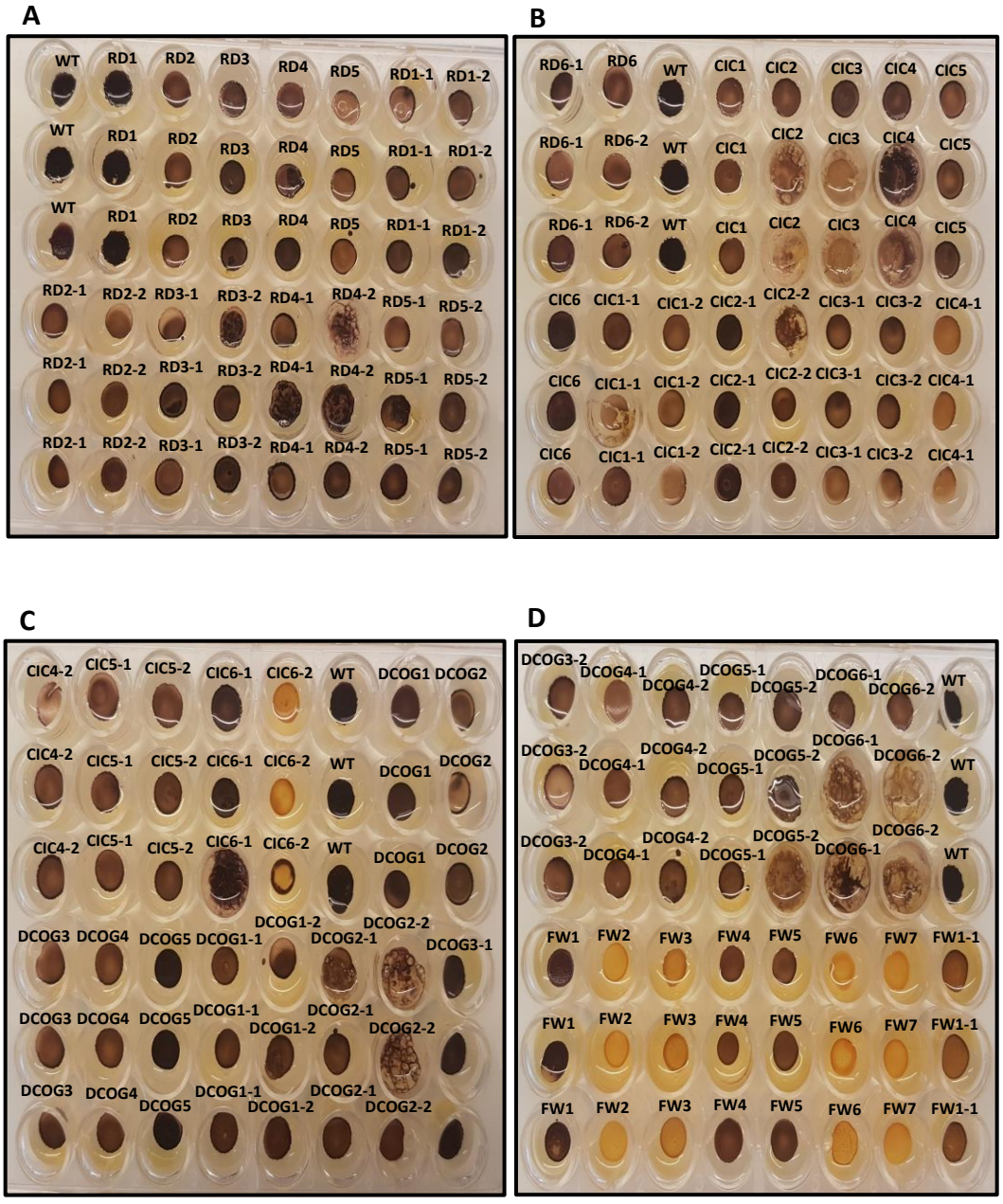
A different trend was observed for the FW type isolates (Figure 3.9D). Only four isolates, FW1, FW5, FW1-1 and FW4-2, produced a measurable number of spores but was however at a much reduced level. Interestingly, these isolates had also produced the highest amounts of butanol after 2 days of growth for the FW morphotypes (Figure 3.8). All other FW strains produced no detectable spores after 5 days of growth.

In summary, with the exception of RD1, it is clear from the spore profiles that all isolates were impaired in sporulation as heat resistant CFU counts were significantly reduced when compared to the wild type.

3.3.4.3. Granulose formation

To ascertain the ability of each isolate to produce the starch-like storage compound granulose, each strain was grown until the mid exponential phase of growth. At this point 10 µl droplets of culture were plated in triplicate onto 1.25 ml CBM-S 6% glucose agar lacking calcium carbonate, contained with the wells of a 48 well microplate. Cells contained within the droplets were left to grow for 48 hours for granulose to be accumulated and then subsequently stained with iodine solution. Each plate contained a wild type control to allow for a visual comparison (Figure 3.10). When iodine solution was added to the wild type control, the cells stained jet black due to the presence of large amounts of granulose. If no granulose was present, the cells would remain light orange as no staining would occur. Out of all of the RD isolates, RD1 was the only one to stain to the same degree as the wild type. All other RD isolates were able to produce a degree of granulose as each lawn of cells stained from pale black to brown. There was also clear variation within these strains, for example

RD3 stained darker than RD5. Overall all RD strains, with the exception of RD1, appeared to have a reduced ability to from granulose.



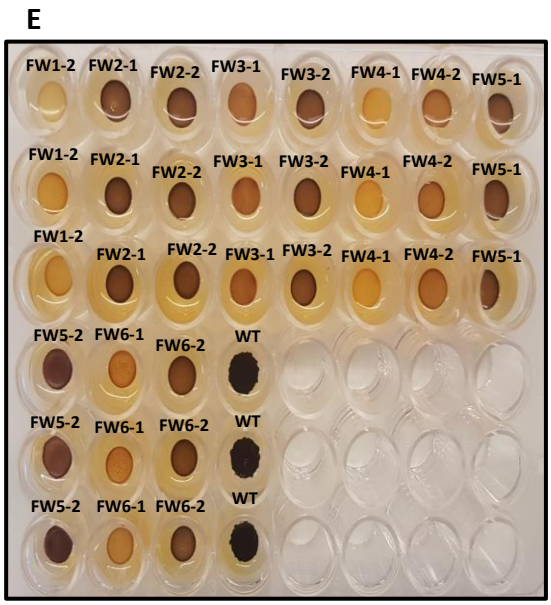


Figure 3.10 Accumulation of granulose in the isolated RD, DCOG, CIC and FW degenerates.

Each strain was assessed in triplicate by plating in different wells in a vertical row. Cultures were left to grow for 48 hours and granulose accumulation visualised by staining with iodine solution. Each plate contained triplicate wild type cultures for comparison. Figure (A) shows RD strains up to RD5-2, (B) shows the remaining RD strains continuing with CIC strains up to CIC 4-1, (C) shows the remaining CIC strains continuing with DCOG strains up to DCOG3-1, (D) shows the remaining DCOG strains continuing with the FW strains until FW1-1 and (E) shows the remaining FW strains.

Similar variation was seen within the CIC isolates which all appeared able to produce granulose, although 2 of the replicate drops of CIC6-2 stained dark orange. CIC4-1 appeared to produce the least amount of granulose as it was a very pale brown. CIC2-1 and CIC6-1 appeared the closest in colour to the wild type but still did not stain as fully.

Comparable results were seen for the DOCG isolates which were all able to form granulose as they stained to a degree. There appeared to be more consistency in colour between DCOG strains than the other colony types as most of the strains were

a pale shade of brown. DCOG5 and DCOG3-1 appeared the closest to the wild type in colour.

FW types showed the most variation in granulose producing ability with several isolates appearing to produce none or very little granulose according to the achieved staining. These included FW2, FW3, FW6, FW7, FW1-2, FW4-1 and FW6-1 as they all remained orange, the same colour as the iodine solution, when staining was attempted. FW1 appeared to produce more granulose than the other isolates but stained still less dark than the wild type. All other strains that were able to make granulose produced a pale brown colours similar to that of the DCOG strains. It was clear that overall the FW strains produced the least amount of granulose compared to the other colony types.

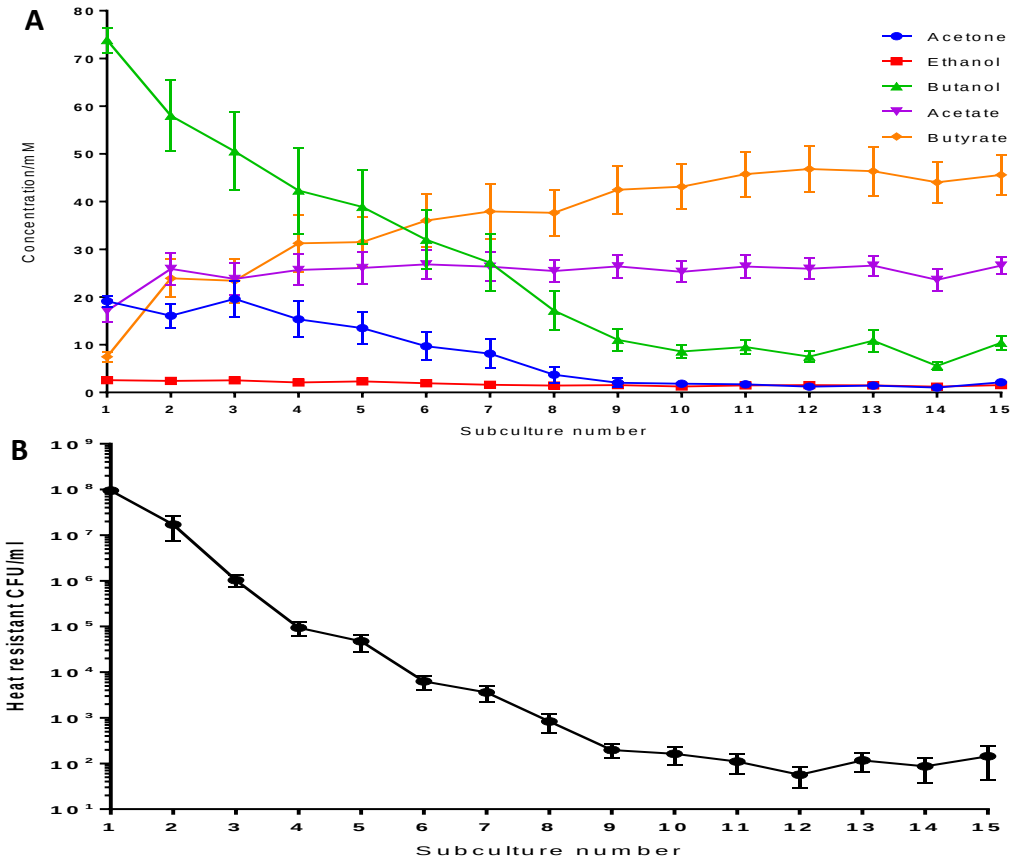
It should also be noted that several wells were broken up due to gas bubbles formed by the growing culture. Plates were left upside down to prevent moisture evaporation which as a result caused strains that produced lots of gas to disrupt the structure of the droplets. Several examples of this include wells from RD4-2, CIC2, DCO2-2 and DCOG6-2. However, it was still possible to see if there was granulose staining in these wells.

3.3.5. Dynamics of culture degeneration

3.3.5.1. Dynamics of culture degeneration for a 1:10 inoculation regime

Having observed the progressive accumulation of different colony types during repeated subculturing under the developed inoculation regime, it was of interest to

ascertain the dynamics of this process and how it affected productivity of the whole culture. As described in 3.3.3., degenerates were isolated using a 1 in 10 dilution at each subculture stage (i.e. 1 ml liquid from the previous subculture into 9 ml media for the next) and progressively took over the culture. To ascertain the effect of this process on the productivity of the whole culture and also which degenerate types emerged at each subculture, the experiment was repeated with CFU's enumerated after 24 hours of growth, solvent samples taken after 2 days of growth and spore samples taken after 5 days of growth at each subculture up to 15 subcultures. In this experiment, 10 individual wild type colonies were picked, inoculated into 10 ml CBM-S broth 6% glucose and independently subcultured. The results of this experiment showed a dynamic change in solvent productivity, colony morphology changes and sporulation which can be seen below in Figure 3.11.



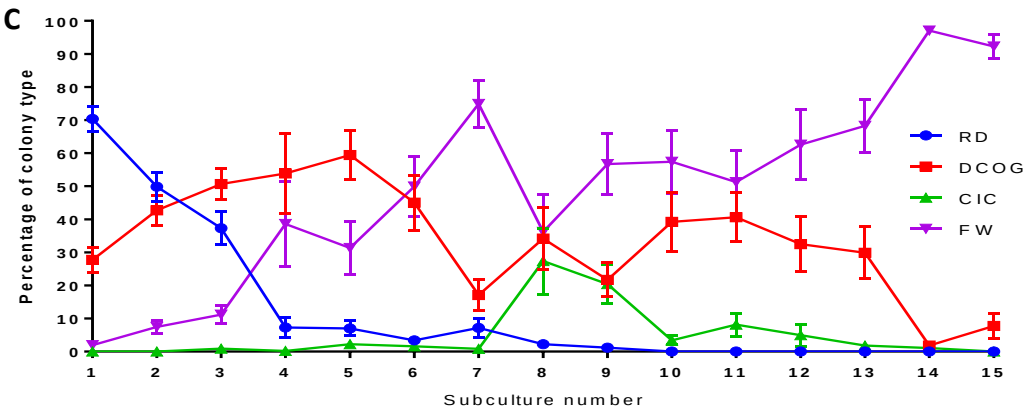


Figure 3.11 Dynamics of culture degeneration during repeated 1:10 subculturing with respect to fermentation products (A), sporulation (B), and percentage of observed colony types (C).

Replica cultures were grown in 10 ml of CBM-S broth 6% glucose and left to grow for 5 days. Solvent samples were taken 2 days after inoculation, spore samples after 5 days and samples for plating and colony type determination after 1 day. (A) Fermentation products as determined by GC; (B) viable heat resistant CFUs; (C) Percentage of each colony type for each subculture which was calculated from the total CFU/ml of each morphotype. Replicates represented by standard error of the mean. *n* = 10 for each graph. End point results were compared to the start and analysed using two way ANOVA and a summary table can be found in the appendix.

Figure 3.11A shows the fermentation product concentrations that were produced after 2 days of growth for each subculture, revealing some interesting trends. It is clear from the onset that at each subculture the concentration of solvents decline. The starting butanol concentration is at 73.9 ± 2.7 mM but by 15 subcultures it was reduced to only 10.4 ± 1.2 mM giving around an 86% reduction in the butanol producing capability of the culture. Each subculture saw the concentration of butanol drop by between 5-10 mM. Acetone concentrations showed a similar trend with a 95% reduction, starting at 19.1 ± 1.8 mM and dropping to a concentration of 2.1 ± 0.4 mM. The concentrations of both solvents were thus significantly less at the end of subculturing than the start (P values ≤ 0.0001, see appendix). Ethanol concentrations

appeared decrease slightly; however, there was no statistical difference between the ethanol concentration at the start and at the end. Statistically summaries of end point values can be found in the appendix.

Acid concentrations showed the reverse trend to the solvents with butyrate concentrations increasing from 7.5 ± 1.1 mM at subculture 1 to 46 ± 4.2 mM at subculture 15 leading to a 542% increase. Acetate concentrations increased from 17.1 ± 2.3 mM at subculture 1 to 26.6 ± 1.8 mM at subculture 15 giving a 52% increase. The concentration of butyrate saw the greatest rise out of the two organic acids. At the start of subculturing the concentration of butyrate was considerably lower than that of acetate. However after the third subculture the concentration of butyrate rapidly over took the concentration of acetate. It appears that by the end of subculturing, the culture was dominated by butyrate production as there was nearly twice as much as acetate (the second highest product) and nearly 4 times as much as butanol after 15 subcultures. It was also very interesting that acetate concentrations appeared to be stable from the second subculture onwards, whereas butyrate continued to increase and was only relatively stable from subculture 9 onwards. Butyrate concentrations were significantly more at the end of subculturing (P values ≤ 0.0001) however acetate was not statistically different. Statistically summaries of end point values can be found in the appendix.

The overall ability of cells within the culture to sporulate also decreased at each subculture matching the ability of the culture to produce solvents. For the first 9 subcultures there was a steep and steady decline in spore numbers resulting in an almost 10^6 fold decrease in heat resistant CFU/ml from 10^8 in subculture to just over 10^2 in subculture 9. After this point the decrease began to tail off with the total heat resistant CFU/ml remaining at around 10^2.

Relative frequencies of colony morphologies, obtained by plating samples of each subculture after 24 hours, showed dynamic changes as the subculturing regime progressed through its iterations. As to be expected, the RD type morphology dominated the culture at the beginning (70.4%) although it was interesting to observe a considerable proportion (27.7%) of DCOG types even after 1 subculture and a low percentage (1.9%) of FW types. The percentage of the RD types, however, rapidly decreased with numbers dropping to less than 10% after 4 subcultures. From the tenth subculture the RD types were no longer detected in the culture. The first 5 subcultures saw the DCOG colony types take over the population with their percentages increasing from just under 30% to nearly 60% of the total culture. From this point, however, the percentage of DCOG types began to decrease, fluctuating between approximately 20% and 40% from subculture 6 to 13, before declining to below 10% for the last two subcultures. The CIC colony type only emerged after 4 subcultures and remained at levels below 2.5% up to subculture 7. It then drastically increased to a maximal relative frequency of just over 27% of the population at subculture 8 before declining; no colonies of this type were found for the final subculture.

It is interesting that from the fourth subculture onwards, FW and DCOG relative frequencies appeared to be inversely correlated, with one exception at the tenth subculture, coinciding with the demise of the CIC types. As a general trend FW type frequencies increased over the course of the experiment until it made up more than 90% of the population in the final two subcultures.

3.3.5.2. Dynamics of culture degeneration for a 1:100 inoculation regime

It has been theorised that continued rapid growth of a culture without sufficient time to induce the solventogenic pathway may play a factor in degeneration (Kashket and Cao, 1995). It was therefore hypothesised for the applied 24 hour transfer regime that enrichment of sporulation and solvent-negative mutants may be driven by faster growth rather than increased survival during stationary phase. A 1:10 inoculation will allow for far fewer generations compared to a 1:100 inoculation. It was therefore hypothesised that smaller inocula should lead to more rapid culture degeneration (in terms of subculture number), as faster growing cells would show stronger enrichment during a single subculture. To see whether such an effect would be observed, the previous experiment (3.3.5.1) was repeated with an inoculum volume of 100 µl instead of 1 ml for each subculture. Only fermentation samples were taken at 2 days and spore samples 5 days post inoculation and the results can be seen below Figure 3.12.

Figure 3.12A displays the fermentation products at each subculture when the inoculum volume was changed to 1 in 100. Comparing to a 1 in 10 dilution, there was a vastly different product profile for the first 10 subcultures. Up to subculture 3, the concentrations of butanol and acetone both increased until they reached a more or less constant concentration of around 80 mM and 35 mM, respectively, having started in subculture 1 from 60 mM butanol and 15 mM acetone. From subculture 11-15, however, the concentrations of both solvents began to decline which coincided with error bars for each of the solvents becoming much larger than for the previous subcultures. It was found that some of the replicate cultures began to degenerate much earlier which therefore affected the average solvent concentrations of the

replicates. Interestingly, concentrations of acids remained steady throughout the first 10 subcultures, with around 10mM for both butyrate and acetate. After this point, along with solvents declining, the concentrations of acids began to increase, particularly for butyrate however end points were not statistically different for any product compared to the start.

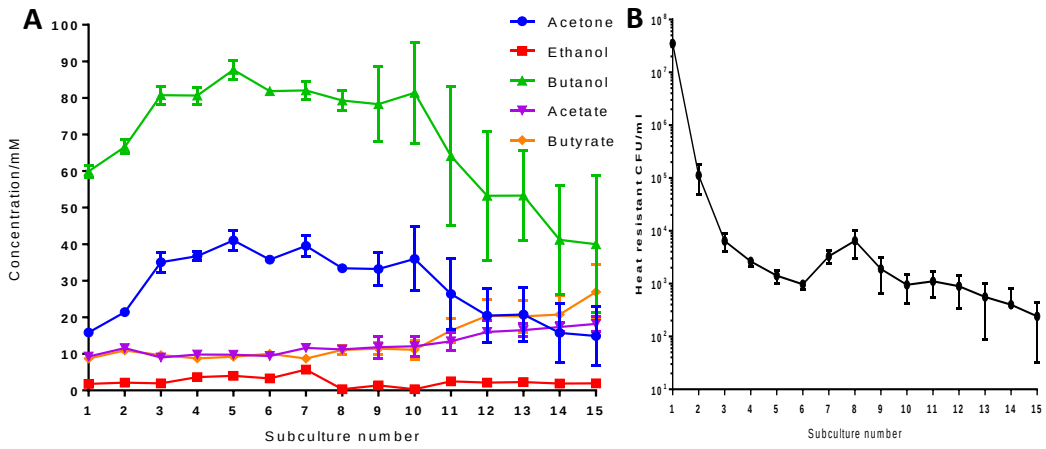


Figure 3.12 Dynamics of culture degeneration during repeated 1:100 subculturing with respect to fermentation products (A) and sporulation (B).

Replica cultures were grown in 10 ml of CBM-S 6% glucose and left to grow for 5 days. Solvent samples were taken 2 days after inoculation and spore 5 days after inoculation. (A) Fermentation products as determined by GC; (B) viable heat resistant CFUs. Technical replicates represented by standard error of the mean; *n* = 4 for each graph. End point results were compared to the start and analysed using two way ANOVA and a summary table can be found in the appendix.

It appears that having a lower dilution for each subculture gave a protective effect on the culture degenerating which lasted for around 10 subcultures. This protective effect could however not be maintained as some of the replicates began to degenerate earlier than others, however, all replicate cultures did begin to or had

degenerated by the end. Although colony type counting was not performed throughout this experiment, colony morphology observations were made at the 5th, 10th and 15th subculture. It was found that at the 5th subculture, the culture population consisted of both RD (52%) and DCOG (48%) types with very little, if at all, of the other two types. By 10 subcultures there was a sharp decrease in the RD type to 22% and an increase in DCOG (46%), CIC (26%) and FW (5%) types. By 15 subcultures, the FW type had begun increase in numbers (33%), as seen with the 1 in 10 dilution, but was still less than the DCOG type (46%). RD and CIC were less than both at 18% and 2% respectively.

The protective effect seen on solvent concentrations was however not observed in regards to the level of culture sporulation. The heat resistant CFU/ml declined until subculture 6 after which point the level of sporulation appeared to stabilise. The initial decline in sporulation appeared to be somewhat faster compared to the 1:10 inoculation regime, reaching average spore levels of below 10^4/ml already for the third subculture. From then on, average spore numbers stayed at around 10^3, with a transient increase for subcultures 7-9, and a final decrease to an average of $3x10^3$/ml. However, two of the replicates produced little or no spores after the subculture 10 point which resulted in larger error bars for these data points. Similarly, it was these replicates that produced much less solvents and increased acids after subculture 10.

3.4. Discussion

C. beijerinckii NCIMB 8052 was chosen as the model organism to study the mechanisms of strain degeneration as it has been found to degenerate rapidly, even in as little as 24 hours (Adler and Crow, 1987; Kashket and Cao, 1995; Stephens et al.,

1985; Woolley and Morris, 1990). Because of this property, it was imperative to start the work with a version of the strain that was as close to the wild type which we obtained from the NCIMB culture collection. It was also necessary to understand how this wild type strain would perform in our laboratory, as specific conditions will vary from the literature, so that it would be possible to pick out any abnormalities in growth when studying degenerates. In terms of strain storage, it was decided to maintain the obtained strain in the form of a spore stock. Degenerated variants are known to show reduced or absent sporulation (Kashket and Cao, 1995) and spore stocks have been used in the past to maintain high solvent formation in industrial strains (Jones and Woods, 1986). A spore stock was therefore generated from the obtained *C. beijerinckii* NCIMB 8052 by leaving a culture to grow in liquid CBM-S 6% glucose for five days and then harvesting the spores by centrifugation. The biological replicates used for the initial phenotypic characterisation originated from a separate, individual colony grown from a single germinated spore from the aforementioned stock which were inoculated in liquid CBM-S 6% glucose to make precultures for the main fermentation. The three biological replicates (also grown in technical quadruplicates) performed as to be expected for this species with acid production in the initial phase of growth followed by solvent production (Lee et al., 2008b). All three biological replicates grew similarly and it was decided to move forward with biological replicate 3 as it appeared to have the least variation in the technical replicates. A spore stock was generated for replicate 3 by harvesting the spores formed during this experiment so that future cells were as close to what was seen from this original phenotypic characterisation. Furthermore, as the spores for this stock came originally from a single colony (the preculture stage) this served to further reduce the amount of genotypic variation within the population for future comparison with isolated

degenerates. A master spore stock was created from this strain and all future work used this stock as the parental wild type.

To study degeneration a reliable protocol was required to promote the emergence of degenerates. As it is well established that degeneration occurs when cultures are repeatedly subcultured in liquid media (Hartmanis et al., 1986; Jones and Woods, 1986; Kashket and Cao, 1995; Kutzenok and Aschner, 1952) this became the starting point for the development of an effective and robust method to isolate degenerates. A heat shocked spore suspension was always used at the beginning as any germinated spores would be assumed to be wild type and not a degenerate. Subsequent inoculation of a single colony was used as to reduce the amount of variation that would be found in a heterogeneous spore suspension population.

Colony morphology became a key indicator in determining the emergence of degenerates which was based on the work of Alder and Crow (1987) and Kashket and Cao (1995). Using strains of *C. acetobutylicum*, Alder and Crow observed progressive changes in colony morphology with each abnormal colony having varying solvent and spore producing capabilities. They characterised 4 types of colony including dense colonies, colonies with outgrowths and colonies that appeared grey (Adler and Crow, 1987). They suggested that the process was sequential and that they never saw colony type reversion. Unlike Adler and Crow, Kasket and Cao studied *C. beijerinckii* NCIMB 8052 directly. Whilst observing degeneration solid media, they took a culture of colonies with a high fraction of degenerate cells (it is unclear how they determined this), diluted, plated on rich agar and left for 1 week. They observed 3 colony types which they described as being beige, thin outgrowth consisting of degenerate cells, a chestnut brown, granular, dry looking, spore-containing centre that did not show a degenerate outgrowths and a third colony type from which no cells could be cultured

(Kashket and Cao, 1995). Both of these studies provided clear evidence that colony morphology is an important tool in determining the presence of degenerates and therefore it became key to the development of the isolation protocol used in this study.

Although Kashket and Cao (1995) presented similar findings to the ones obtained in this study, in terms of identified colony morphologies, an important difference was the quality and resolution in the observations presented here. For instance, they only described 3 colony types, one of which they could not culture, compared to the 4 found in this study. They also published agar plate images containing their colony types however the resolution of the image was insufficient to fully appreciate the variation in morphology. Their work, however, served as a good grounding to begin more detailed investigations into degeneration as carried out here.

The images presented in Figure 3.4 clearly show 4 distinct colony morphotypes labelled based on their appearance. As mentioned previously, the 4 types being the round and dark (RD), dark centre with outgrowths (DCOG), caved in centre (CIC) and flat and white (FW). When observing these colony types through a series of 1:10 inoculated subcultures, a clear, progressive change in the presence and frequency of each type could be observed (Figure 3.3 and Figure 3.11C), which can be summarised as follows: before the fourth subculture the most dramatic changes occurred for DCOG and RD types, the relative frequencies of which also appeared to be inversely related, with DCOG types increasing in relative frequency and RD types rapidly declining. In a second phase, from subculture 4 to 7, when RD and CIC relative frequencies were very low, percentages for FW types rapidly increased whereas those for DCOG types decreased. This trend was reversed in subculture 8 which saw the sudden, but transient, rise in CIC types. Following CIC decline, subcultures 10-13

showed comparatively little change for DCOG and FW types before the latter became by far the dominating type by reaching >90% at the end.

In these sequence of events, it would seem the DCOG initially outcompetes the RD with increasing FW contributing to reduced percentages of the RD. Once RD numbers are sufficient low, the FW is then able to outcompete the DCOG, although interrupted by the CIC which also appeared to promote DCOG competiveness. This is however short lived and FW eventually outcompetes all types. Although there are clear images of each of the colony types (Figure 3.4), assigning each type can be subjective. With the exception of the FW, the other types do share similarities such as dark centres (RD, DCOG and CIC) and small outgrowths (DCOG and CIC). This subjectiveness seems to be reflected by the range of phenotypic variation seen from the analysis of the isolates (Figure 3.5-8) making interpretation difficult as to how culture productivity links to type frequency. Moreover, most colony types are probably still evolving throughout the subcultures so that a specific type isolated at the start of the subculturing will most likely differ phenotypically/genotypically to same type isolated in later subcultures. The population is however progressing to one which is dominated by types that have reduced, or abolished, solvent and spore forming capabilities. Perhaps the use of ultra-deep sequencing, as employed with industrial *E. coli* (Rugbjerg et al., 2018), could be used to track the distribution of genetic defects and trends at the genomic level.

As mentioned, various colony types of *C. acetobutylicum* have been observed when subcultured (Adler and Crow, 1987). The initial colonies isolated produced relatively high levels of solvents and spores compared to the colonies isolated at the end of subculturing which produce no solvents or spores. Furthermore, two intermediate colony types were also isolated which had varying solvents and spore producing

capabilities (Adler and Crow, 1987). The colony types found in this study support what was found for *C. acetobutylicum* as the RD seem to represent the first type, the FW the last and the DCOG/CIC the intermediate. Furthermore the data here also supports evidence that degeneration of this strain can occur after a small amount of subcultures (Kashket and Cao, 1995).

Phenotypic analysis of each given type revealed more variation than expected. Assuming the isolated RD types were still exhibiting wild type characteristics, it would be expected that this type would produce comparable levels of solvents, granulose and spores. This, however, was not the case with only one isolate (RD1) having a comparable profile to that of the wild type. Although most of the RD types did produce comparable solvents, the ability to sporulate was significantly lower (Figure 3.5 and Figure 3.9A). Similar results were seen with the DCOG and CIC types as many of these morphologies producing normal levels of butanol but lacked the ability to sporulate at number to that of the wild type. These intermediate types further support the assertion that although solvent formation and spore formation are linked (Dürre, 2014), the ability to form viable spores is not necessarily linked to the relative amount of solvent production (Adler and Crow, 1987; Gottschal and Morris, 1982; Jones et al., 1982a; Woolley and Morris, 1990). Comparing these three types with each other, generally the RD, DCOG and CIC types all looked phenotypically similar with the exception of some outliers. Whilst colony type indicated degeneration, there seemed to be no standard phenotype that could be assigned to colony type, apart from perhaps the FW.

The most drastic phenotypes observed were those of the FW type. Most isolates of this morphotype produced little to no solvents and spores, the characteristics of a strain that has fully degenerated (Jones and Woods, 1986; Kashket and Cao, 1995;

Kutzenok and Aschner, 1952; Woolley and Morris, 1990). From data presented in Figure 3.11C, this type eventually took over the population leaving little to none of the other types. It would be reasonable to assume, therefore, that by becoming fully degenerated, the FW types have gained a fitness advantage over the other types, even those that had already lost the ability to form spores but still produced solvents. This would imply that under the employed conditions the ability to produce solvents was not necessary for survival of a strain and its loss improved fitness, allowing a mutant to eventually take over the population. Adler and Crow (1987) theorised that the cells which do not sporulate have the tendency to take over the population as they can continue to divide while other cells have committed to sporulation. This may be true for continuous cultures; however, for cultures that are consistently regrown every 24 hours, an alternative explanation might be that it is the faster growing cells that will dominate (a hypothesis will be further looked at in Chapter 5).

As mentioned above, the ability to produce solvents at a comparable level to the wild type appears not to be an indicator of a strain's ability to sporulate. At least for *C. acetobutylicum,* the onset of solventogenesis is assumed to be coupled with sporulation (Dürre, 2014) however isolates such as DCOG2 were able to produce solvents but did not produce any apparent spores. This is perhaps not surprising, given that *C. beijerinckii*, in contrast to the more commonly studied model species *C. acetobutylicum,* produces solvents whilst actively growing, as can be seen in Figure 3.1. Furthermore, the DCOD2 strain was still able to produce the storage molecule granulose which is usually accumulated before the onset of sporulation (Reysenbach et al., 1986) implying different pathways may be involved. Similar results have been observed in the literature with several examples of strains that can produce sufficient solvents without the need to sporulate (Adler and Crow, 1987; Jones et al., 1982b; Kashket and Cao, 1995; Long et al., 1984; Woolley and Morris, 1990). Mutants

displaying such characteristics are entirely expected, based on our current understanding of sporulation and solventogenesis in clostridial species (Dürre, 2014; Dürre and Hollergschwandner, 2004; Ravagnani et al., 2000): both are controlled via the master regulator Spo0A and inactivation of the encoding gene or those required for its full functionality will affect both phenotypes. However, mutational interference with regulatory processes or structural genes further down in the sporulation cascade will lead to solvent producing, asporogenous mutants (Jones et al., 1982a). It follows that the morphotype observed does not always indicate a strain's phenotypic behaviour and merely serves as a suggestion to the solvent and spore producing capabilities of that strain. It is clear however that practically all of the types were degenerated in some capacity after 5 transfers with the only exception of RD1.

An interesting theme also emerged when comparing the fermentation products at 2 days and 5 days post inoculation. It was found that by the end of the 5 day fermentation wild type samples contained more solvents than samples at 2 days. This phenotype was not observed for the obtained isolates. At day 5, isolate samples either contained equal or lower solvent concentrations than the 2 day samples. The cause for this decrease in solvent concentrations was most likely evaporative loss; however the differing behaviour of the wild type still implies differing growth and (or) production rates. Isolated strains may grow faster, switch to solventogenesis earlier and therefore end their growth earlier than the wild type. This would result in higher solvent concentrations in the 2 day sample compared to the 5 day sample. The implication of this is that the isolated strains have an initial growth advantage in the early stages of growth, allowing them to reach the stationary phase quicker than the wild type (a hypothesis that will be addressed in Chapter 5). Interestingly, extended solvent production by the wild type beyond day 2 was not observed in the original growth experiments shown in Fig 3.1. Characterisation of isolates was undertaken at

smaller scale in Falcon tubes and it is possible that this change in condition affected the growth and fermentation kinetics of the wild type.

An experiment similar to that displayed in Figure 3.11A was carried out by Kutzenok and Aschner (1952). They studied the degeneration of a strain of *C. butylicum* and monitored the fermentation profile at each subculture. Similar results were obtained, i.e. successive subculturing caused concentrations of solvents to decrease and concentrations of acids to increase as degenerates accumulated in the culture however this only began after subculture 7. They did not observe spore numbers however and the data presented in this study found that spores decreased significantly until they were maintained at 10^2 heat resistant CFU/ml. The low levels of spores are most likely due to a very small subpopulation that are still able to produce spores, possibly the remaining DCOG.

It is apparent that the FW type favours the production of butyrate over acetate (Figure 3.8) which explains why butyrate is increasingly produced through the subcultures. It should be noted that production of acetate from glucose generates more ATP than butyrate i.e. 4 ATP compared to 3 ATP from 1 molecule glucose, respectively (Jones and Woods, 1986). Producing acetate however creates a net gain in NADH which needs to be reduced to continue glycolysis (Jones and Woods, 1986). Perhaps these organisms favour butyrate production as this is redox neutral and will allow sufficient ATP generation without causing unfavourable changes in the redox balance. For many years, NADH ferredoxin oxidoreductase and hydrogenase have been considered important enzymes in controlling the direction of electron and carbon flow through the two acid producing pathways. It was believed that NADH ferredoxin oxidoreductase catalysed the transfer of electrons from NADH to ferredoxin, which was subsequently re-oxidised in the hydrogenase reaction, leading

to H_2 production (Jones and Woods, 1986). As the first reaction is thermodynamically very unfavourable, it was thought to occur only at low H_2 partial pressures, thus limiting the amount of NADH that could be oxidised via this pathway. Thus, when the respective enzyme complex is inhibited or partial pressures are too high, excess NADH must be utilized in the production of butyrate to maintain the correct redox balance (Jones and Woods, 1986). This was seen for *C. butyricum* when excess mannitol was used as a more reduced substrate and increased H_2 pressure caused the organism to produce more butyrate. This ensured that NADH was utilised in the production of butyrate as the enzyme complex could not maintain redox balances (Crabbendam et al., 1985). With the recent discovery of the Rnf complex (Biegel et al., 2011; Dürre, 2014; Herrmann et al., 2008; Lütke-Eversloh and Bahl, 2011), the biochemical basis of this process has become much clearer. One of the steps in the C_4 pathway leading to butyrate and butanol formation, the butyryl-CoA dehydrogenase reaction (see introduction Figure 1.1), is highly exergonic. Many bacteria conserve some of the released energy through a process termed electron bifurcation (Buckel and Thauer, 2013). In essence, this involves the use of 2 NADH, which are used to reduce both crotonyl-CoA and ferredoxin in an overall thermodynamically favourable reaction. Electrons from the latter can then be used for hydrogen generation, allowing energy conservation through increased acetate production. Alternatively, some solventogenic clostridia, including *C. beijerinckii* can make use of the so-called Rnf complex, a membrane-bound ferredoxin:NAD oxidoreductase. This complex uses the ferredoxin-derived electrons for NAD reduction whilst conserving the released energy in form of an ion (proton) gradient across the cytoplasmic membrane (Biegel et al., 2011; Dürre, 2014; Herrmann et al., 2008; Lütke-Eversloh and Bahl, 2011). It thus seems that degenerate isolates make use of butyrate formation as this is a redox neutral pathway when grown on glucose and allows for additional energy

conservation. Whilst butyrate is considered more toxic than acetate, for mutants limited to acid formation use of this pathway makes sense as it limits overall acid concentration (i.e. generating 1 C_4 acid rather than 2 C_2 acids from 1 glucose).

The above discussion focused on the original 1:10 inoculation regime. It was interesting to observe that the rate of culture degeneration appeared to be slowed for the 1:100 regime. In fact, solvent concentrations initially increased before stabilising. This ‘‘protective’’ effect lasted around 10 subcultures but was not maintained past this point due to the emergence of the FW morphotype. It appears that once this type emerges, the culture were destined to completely degenerate as they take over the population. The dramatic changes observed in the “kinetics” of degeneration suggests that minor changes of conditions can have strong impacts on the relative fitness of the different mutants.

From the findings in this study it is clear that the developed subculturing protocol provides a reliable, quick and robust method to produce degenerates. Degenerates that emerge are diverse in their morphology, solvent, spore and granulose producing capability. It appears that the first degenerate mutants to emerge have lose their ability to sporulate efficiently, which is then followed by mutants that do not produce solvents. FW colonies were the most severe morphotype which eventually took over the population. The genes involved in degeneration will subsequently be explored in Chapter 4 making use of the pool of isolated degenerates described here.

4. Genotypic characterisation of the isolated degenerates and subsequent genetic manipulation of *C. beijerinckii* NCIMB 8052

4.1. Introduction

There has been little to no study of the causes of degeneration through genetic methods such as whole genome sequencing and comparative genomics. Suggestions have been made that degeneration may be via disruption of a global regulatory gene (Kashket and Cao, 1995; Woolley and Morris, 1990) but no further genetic investigations into this area have been carried out. With the advent of Next Generation Sequencing (NGS) it may now be possible to compare large genetic data sets against each other to study the course and outcomes of the degeneration process.

Since the first bacterial genomes were fully sequenced in 1995 (Fleischmann et al., 1995; Fraser et al., 1995), the field of whole genome sequencing has expanded rapidly bringing vast amounts of genomic data to individual research groups rather than large genome centres (Land et al., 2015; Shendure and Ji, 2008). In the early years of bacterial genome sequencing, a finished bacterial genome could cost up to $50,000; however, costs have now fallen to nearly $1 to generate raw sequences for a draft genome (Land et al., 2015). The reduction in cost is owed to the advent of new technologies known as Next Generation Sequencing (NGS) which has allowed for a greater number of sequence reads than traditional sequencing methods in a shorter space of time (Edwards and Holt, 2013; Land et al., 2015; Loman et al., 2012; Metzker, 2005; Shendure and Ji, 2008).

The earlier forms of high through put sequencing used semi-automated implementations of Sanger biochemistry (Shendure and Ji, 2008). These methods required synthesis of single strands of DNA templates while randomly incorporating chain terminators through PCR amplification. The chain terminators would be

fluorescently labelled dideoxynucleotides (ddNTPS) giving a range of different fragment sizes with terminal labels. Fragments would be separated using capillaries with the fluorescent labels excited to give Sanger 'traces' which would finally be read using software to create DNA sequences (Hall, 2007; Mardis, 2013; Sanger et al., 1977; Shendure and Ji, 2008). Problems with the use of this technology for whole genome sequencing was that it required dedicated genome centres to manage the cost and time needed for sequencing (Hall, 2007). A further drawback of this technology was the need for fragmented DNA to be cloned into high copy number plasmids and transformed into *E. coli* before cycling could be completed (Hall, 2007; Shendure and Ji, 2008). This became problematic as fragments may not replicate well in the *E. coli* hosts due to their length or if they code for toxic compounds (Hall, 2007). Furthermore, the cloning process could take up to a week for DNA to be ready for sequencing adding to sequence completion times (Mardis, 2017).

Unlike the high through put Sanger sequencing technologies, the NGS technologies can prepare DNA for sequences in a much short time as there is significantly less sample preparation (Mardis, 2017). There are many NGS variants but the general principles are as follows: DNA used for sequencing is prepared into a fragment library, either by enzyme cutting or sonication, with universal DNA adapters subsequently ligated to the ends of the fragments. These adapters allow binding of the fragments to either beads or a solid surfaces to provide the templates needed for polymerase amplification of the amplicons. Fluorescently labelled nucleotides are added, incorporated and then washed away in a cycling PCR amplification of the fragments which are then read by the sequencing machine (Mardis, 2013, 2017; Shendure and Ji, 2008). Thousands of sequences are created which are aligned together using bioinformatic software to piece together a genome (Behjati and Tarpey, 2013). The above outline is an incredibly simplified version of the techniques used by several NGS

companies and comparisons of their technologies can be found in the literature (van Dijk et al., 2014; Mardis, 2017; Reuter et al., 2015; Shendure and Ji, 2008).

The main advantages of NGS over Sanger sequencing include the ability to not have to create fragment libraries in *E. coli,* much less reagents and time needed and hundreds of millions more reads (Shendure and Ji, 2008). Sanger sequencing does however provide longer sequence reads and reduction in incorrect base calling than the traditional NGS technologies. To overcome sequence length, newer technologies such as PacBio (Pacific Bioscience) and Nanopore (Oxford Nanopore technologies) sequencing can create significantly longer reads (>15,000 for PacBio) than other NGS technologies (Mardis, 2017).

After the introduction of cheap and rapid NGS in 2005, genome sequencing has provided a powerful tool in the world of microbiology. The field of medial microbiology has benefited significantly from NGS providing large pools of metadata for clinical laboratories (Bertelli and Greub, 2013). This has given rise to considerable improvements in diagnostics for the detection of pathogenic isolates, epidemiological typing of pathogens, the tracking of clinical isolates and the ability to study pathogen evolution (Bertelli and Greub, 2013; Hasman et al., 2014; Klemm and Dougan, 2016). NGS has further been used to study evolution of engineered *E. coli* industrial strains to improve product yields when undesired mutants arise within fermentations (Rugbjerg et al., 2018).

Although NGS has proven to be an incredibly useful tool in the study of bacterial genomes, it provides less information on the role and function of selected genes. By genetically manipulating bacterial genes directly, it is possible to observe phenotypes of the resulting mutants to provide evidence of a gene's function. There is now an extensive set of genetic tools used to manipulate *Clostridia* which has been covered

in section 1.6. One important technology employed is the use of ClosTrons which provide a quick and robust method for gene disruption (Heap et al., 2007, 2010; Kuehne et al., 2012). ClosTron technology uses mobile group II introns to create insertional disruptions in genes in an effort to disable them. Draw backs when using this technology include downstream polar effects and incorrect insertions within the genome (Heap et al., 2010; Kuehne et al., 2012; Pyne et al., 2014; Shao et al., 2007). These can be overcome by the use of both sense and anti-sense insertions and with complementation of genes back on the chromosome to restore function and thus prove that the ClosTron was the cause of the phenotype. Complementation can easily be achieved through the use of Allele Coupled Exchange (ACE) (Kuehne et al., 2012; Ng et al., 2013). This technique allows large fragments of the DNA to be integrated downstream of the *pyrE* gene through use of homologous recombination and restoration of the *pyrE* gene (Heap et al., 2012).

NGS was employed here as a tool to study degeneration through comparative genomics as this has not been attempted before in solventogenic *Clostridia*. It was hoped that any important genetic factors would be eluded to. These would then be investigated further to prove their significance.

4.2. Aims of this study

The main aims of this study was to provide a greater understanding of the genes involved in strain degeneration of solventogenic *Clostridia*. This would be done through the use of NGS to provide genomic data for the 71 degenerate strains isolated in Chapter 3. Each genome would be subsequently compared with that of the parental wild type they were derived from. Any genes believed to be important would be

characterised through ClosTron gene disruption to observe if they provided the same phenotypes as the naturally evolved degenerates. The specific aims were as follows:

- To obtain and compare the genome sequences of the 71 isolated degenerates from Chapter 3 with that of the parental strain they were derived from
- To create a genomic map to visualise the distributions of mutations and see if there are any 'hot spot' genes which contained more mutations than any other regions
- To disrupt any genes believed to be involved in degeneration using ClosTron technology
- To create a Δ*pyrE* mutant to allow for complementation of disrupted genes using Allele Coupled Exchange (ACE)
- To characterise the created ClosTron mutants and complemented mutants via their fermentation products and spore profiles

4.3. Results

4.3.1. Genomic comparisons of 71 isolated degenerates with their parental wild type

Comparative genomics has been employed to study mutation rates and spectra in a variety of organisms including *Burkholderia cenocepacia* (Dillon et al., 2015), yeasts (Lynch et al., 2008) *Caenorhabditis elegans* (Denver et al., 2009), *Arabidopsis thaliana* (Ossowski et al., 2010) and *E. coli* (Lee et al., 2012) to name a few. These studies began with a single clonal ancestor (whether it was a colony, cell or seed) which is subsequently passaged through several replicative lineages either with or without

bottlenecks (Dillon et al., 2015). The genomic sequences of each lineage are then compared to the original ancestor to determine both the spectrum and rate of mutations. This is technique is known as mutation accumulation (MA) with whole genome sequencing (WGS) (MA-WGS).

In the previous study, 71 degenerate strains of *C. beijerinckii* NCIMB 8052 were isolated and phenotypically characterised. Each of the strains had their genomic DNA extracted when they were isolated so that it could be compared to the parental wild type they were derived from. As explained in Chapter 3, there were 2 rounds of degenerate isolation which began with a single colony distributed to 6 Falcon tubes and subcultured for 5 days. The single colonies used (from clone 3) became the parental strains that would be used for genomic comparisons. Genomic sequencing was carried out using the Illumina MiSeq platform by MicrobesNG (Birmingham, UK). Both parental wild types and isolated degenerate reads were assembled against the most up to date reference genome at the time found on the NCIMB database (Wang et al., 2012) using CLC Genomics Workbench 11 (Qiagen). Once read mapping had been completed, each of the degenerate isolates were compared against their respective parental wild type (WT1 for the first round of isolated degenerates and WT2 for the second round of isolated degenerates). Any of the same mutations were discarded as these were assumed to be the common starting mutations present in the ancestral strain (again using CLC Genomics Workbench 11 (Qiagen)). Below is a comprehensive list of all of the mutations found from the comparative analysis including those of the parental wild types (Table 4.1). The table shows the position of the mutation, the nature of the genetic change (SNV [single nucleotide variant], deletion or insertion), if there was an amino acid change and which gene was affected including both old and new locus tags. Only mutations that were present in above 90% of the reads were counted.

Table 4.1 The full list of mutations for each of the isolated degenerates.

Strain	Position	SNV	Amino acid change	Gene affected
Parental WT1	1099984	Deletion -> AAATA	-	**Cbei_RS04935** Phage tail tape measure protein
	1255087	T -> C	Val -> Ala	**Cbei_RS05570/Cbei_1046** Paraslipin
	33369627	Insertion of T	-	**Cbei_RS14880/Cbei_2885** Lactate permease
	5108804	G -> T	Gln -> Lys	**Cbei_RS22635/Cbei_4430** Histidine kinase
	5170894	T -> C	-	non-coding region before phosohydrolase
Parental WT2	1099984	Deletion of AAATA	-	**Cbei_RS04935** Phage tail tape measure protein
	1255087	T -> C	Val -> Ala	**Cbei_RS05570/Cbei_1046** Paraslipin
	3369627	Insertion of T	-	**Cbei_RS14880/Cbei_2885** Lactate permease
	5170894	T -> C	-	non-coding region before phosohydrolase
	5769732	T -> C	-	non-coding region
RD1	2457002	G -> T	Asp -> Tyr	**Cbei_RS10935/Cbei_2108** Transcriptional regulator
RD2	30222347	G -> T	Pro -> His	**Cbei_RS13425/Cbei_2598** Amine oxidase
	3590871	C -> A	Arg -> Ile	**Cbei_RS15850/Cbei_3077** TetR family transcriptional regulator
	4940233	C -> A	Glu -> STOP	**Cbei_RS21905/Cbei_4283** Hypothetical protein
	5817138	G -> T	Thr -> Asn	**Cbei_RS25280/Cbei_4941** ABC transporter substrate binding protein
RD3	5968574	G -> A	-	**Cbei_RS26000** tRNA repeat region
RD4	28760	G -> T	Glu -> STOP	**Cbei_RS00130/Cbei_0017** Two component sensor histidine kinase
	679651	G -> T	Arg -> Ile	**Cbei_RS03160/bei_0568** Stage V sporulation protein R
	4068435	C -> A	-	Intergenic region between hydantoinase and RpiR family transcriptional regulator
RD5	1981714	G -> T	Val -> Leu	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
	4981609	C -> A	Asp -> Tyr	**Cbei_RS22105/Cbei_4323** DNA gyrase subunit A
	5931540	C -> A	-	**Cbei_RS25795** repeat region rRNA
RD1-1	28469	G -> T	Glu -> STOP	**Cbei_RS00130/Cbei_0017** Two-component sensor histidine kinase
	2084065	G -> T	Glu -> STOP	**Cbei_RS09370/Cbei_1801** Hypothetical protein
	2270633	G -> T	-	**Cbei_RS10160** repeat region rRNA
	3973931	C -> A	Glu -> STOP	**Cbei_RS17565/Cbei_3423** LysR family transcriptional regulator
	4253038	G -> T	Asp -> Tyr	**Cbei_RS18960/Cbei_3698** MarR family transcriptional regulator
	4797168	C -> A	Leu -> Phe	**Cbei_RS21310/Cbei_4166** Hypothetical protein
	6000108	C -> A	Asp -> Tyr	**Cbei_RS26170/Cbei_5103** Ribonuclease P protein component
RD1-2	28469	G -> T	Glu -> STOP	**Cbei_RS00130/Cbei_0017** Two-component sensor histidine kinase
	2084065	G -> T	Glu -> STOP	**Cbei_RS09370/Cbei_1801** Hypothetical protein
	4253038	G -> T	Asp -> Tyr	**Cbei_RS18960/Cbei_3698** MarR family transcriptional regulator
RD2-1	3050567	C -> A	Glu -> STOP	**Cbei_RS13575** Hypothetical protein
	4718705	C -> A	Gly -> Cys	**Cbei_RS21005/Cbei_4106** Hypothetical protein

	5537882	C -> A	Asp -> Tyr	**Cbei_RS24215/Cbei_4748** Hypothetical protein
	5707961	C -> A	Asp -> Tyr	**Cbei_RS24840/Cbei_4873** Hypothetical protein
	5720359	C -> A	-	Non-coding region between **Cbei_RS24895/Cbei_4884** Hypothetical protein and abrB Family transcriptional regulator
RD2-2	1754001	G -> T	No change	**Cbei_RS07805/Cbei_1488** Hypothetical protein
	2109257	G -> T	Gly -> Val	**Cbei_RS09480/Cbei_1823** Methionine aminopeptidase
	3591533	A -> T	Met -> Lys	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3591535	T -> C	-	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3591539	A -> G	Ile -> Thr	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3591533 - 3591717	Deletion 185bp		**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
RD3-1	29210	G -> T	Glu -> STOP	**Cbei_RS00130/Cbei_0017** Two-component sensor histidine kinase
	5422515	C -> A	Asp -> Tyr	**Cbei_RS23845/Cbei_4674** Glycosyl hydrolase
RD3-2	28790	G -> T	Glu -> STOP	**Cbei_RS00130/Cbei_0017** Two-component sensor histidine kinase
	2322996	G -> T	Gly -> Val	**Cbei_RS10390/Cbei_1998** Hypothetical protein
	5532027	G -> T	Gly -> Cys	**Cbei_RS24200/Cbei_4745** Cell wall-binding repeat-containing protein
RD4-1	5711308	C -> A	-	Intergenic region before hypothetical protein
RD4-2	29225	G -> T	Asp -> Tyr	**Cbei_RS00130/Cbei_0017** Two-component sensor histidine kinase
	50881	C -> A	Pro -> His	**Cbei_RS00230/Cbei_0036** Peptidoglycan-binding protein
	3017983	C -> T	Trp -> STOP	**Cbei_RS13400/Cbei_2593** Hypothetical protein
	4439332	G -> T	Met -> Ile	**Cbei_RS19800/Cbei_3867** DNA-binding response regulator
RD5-1	3591585	C -> A	Gly -> STOP	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
RD5-2	29033	G -> T	Gly -> Trp	**Cbei_RS00130/Cbei_0017** Two-component sensor histidine kinase
	920068	G -> T	Glu -> STOP	**Cbei_RS04125/Cbei_0760** Membrance protein
	2134513	G -> T	Glu -> STOP	**Cbei_RS09580/Cbei_1843** 1-phosphofructokinase
RD6-1	418185	C -> A	Ser -> Tyr	**Cbei_RS01900/Cbei_0348** MarR family transcriptional regulator
	883108	G -> T	Val -> Leu	**Cbei_RS03990/Cbei_0734** ABC transporter permease
	2732900	G -> T	-	Non-coding region
	2860132	C -> A	-	Non-coding region
	3454643	C -> A	Arg -> Met	**Cbei_RS15205/Cbei_2947** ATPase AAA
	3592500	C -> A	Glu -> STOP	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3680423	G -> T	Ser -> STOP	**Cbei_RS16225/Cbei_3151** DNA helicase UvrD
RD6-2	3706328	C -> A	Gly -> Val	**Cbei_RS16310/Cbei_3170** Hypothetical protein
	3928585	C -> A	Ala -> Ser	**Cbei_RS17320/Cbei_3377** PAS domain-containing histidine kinase
	4577768 - 4577785	Deletion 18bp	-	**Cbei_RS20405/Cbei_3986** Serine protease
DCOG1	2401284	C -> A	Leu -> Ile	**Cbei_RS10705/Cei_2062** Hypothetical protein
	5720570	G -> T		Intergenic region before AbrB family transcriptional regulator
	5874958	C -> A	Val -> Phe	**Cbei_RS25530/Cbei_4991** Phage minor structural protein
DCOG2	817796	C -> A	Ala -> Glu	**Cbei_RS03705/Cbei_0679** Heme ABC transporter ATP-binding protein
	1163369	G -> T	Ser -> Tyr	**Cbei_RS05205/Cbei_0973** RpiR family transcriptional regulator

	1982142	G -> T	Gly -> Val	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A (at end of gene)
	4015790	C -> A	Ala -> Ser	**Cbei_RS17805/Cbei_3469** Two component sensor histidine kinase
	4686712	C -> A	Gly -> Val	**Cbei_RS20880/Cbei_4080** Histidine triad (HIT) protein
DCOG3	2360809	C -> T	Gln -> STOP	**Cbei_RS10540/Cbei_2029** Potassium transporter Kup
	2807017	G -> T	Gly -> Val	**Cbei_RS12540/Cbei_2425** Hypothetical protein
	3705198	C -> A	Val -> Phe	**Cbei_RS16305/Cbei_3169** Histidine kinase
	5414424	C -> A		Intergenic region between ABC transporter substrate binding protein and sulfatase
	5931540	C -> T		**Cbei_RS25795** repeat region rRNA
DCOG4	338501	G -> T	-	Intergenic region after hypothetical protein
	1758232	G -> T	Gly -> Cys	**Cbei_RS07815/Cbei_1490** Hypothetical protein
	3586973	C -> A	Glu -> STOP	**Cbei_RS15835/Cbei_3074** Hybrid sensor histidine kinase/response regulator
	3590854	C -> A	Val -> Phe	**Cbei_RS15850/Cbei_3077** TetR family transcriptional regulator
	4897862	C -> A	silent - no change	**Cbei_RS21690/Cbei_4241** DNA polymerase IV
DCOG5	57110020	C -> A	Asp -> Tyr	**Cbei_RS24860/Cbei_4877** Hypothetical protein
DCOG1-1	3591533	A -> T	Met -> Lys	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3591535	T -> C	-	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3591539	A -> G	Ile -> Thr	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3591533 - 3591717	Deletion 185bp		**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
DCOG1-2	3460894	C -> A	Arg -> Ile	**Cbei_RS15235/Cbei_2953** Methyl-accepting chemotaxis protein
	3591533	A -> T	Met -> Lys	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3591535	T -> C	-	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3591539	A -> G	Ile -> Thr	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3591533 - 3591717	Deletion 185bp		**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	5360378	C -> A	Val -> Phe	**Cbei_RS23615/Cbei_4628** Hypothetical protein
DCOG2-1	3592522	G -> T	Ser -> Tyr	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3685250	C -> A	Glu -> STOP	**Cbei_RS16235/Cbei_3153** Hypothetical protein
DCOG2-2	3591854	G -> T	Ser -> STOP	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	5140000	C -> A	Met -> Ile	**Cbei_RS22740/Cbei_4451** Hypothetical protein
DCOG3-1	5979710	C -> A	Gly -> STOP	**Cbei_RS26055/Cbei_5080** ATPase AAA
DCOG3-2	2130349	G -> T	Asp -> Tyr	**Cbei_RS09560/Cbei_1839** Altronase hydrolase
	3355706	G -> T	Gln -> Lys	**Cbei_RS14825/Cbei_2875** Acriflavin resistance protein
	3590667	G -> T	Ser -> STOP	**Cbei_RS15850/Cbei_3077** TetR family transcriptional regulator
	3693457	C -> A	Glu -> STOP	**Cbei_RS16265/Cbei_3159** Hypothetical protein
	5463951	G -> T	Gln -> Lys	**Cbei_RS23970/Cbei_4699** Cell wall hydrolase
DCOG4-1	502632	G -> T	Arg -> Ile	**Cbei_RS02360/Cbei_0414** ATP synthase subunit B F1F0?
	3591584	C -> A	Gly -> Val	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
DCOG4-2	740972	G -> T	Ser -> Ile	**Cbei_RS03425/Cbei_0621** Hydrogenase
	2937610	G -> T	Ser -> Arg	**Cbei_RS13070/Cbei_2530** Cell division protein FtsH

	3593522	G -> T	Ser -> Tyr	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	4061814	C -> A	Lys -> Asn	**Cbei_RS18040/Cbei_3516** Phosphoenolpyruvate synthase
	4088692	C -> A	Glu -> Asp	**Cbei_RS18170/Cbei_3541** Glycine/betaine ABC transporter substrate-binding protein
	5012968	C -> A	Val -> Leu	**Cbei_RS22225/Cbei_4347** 4-aminobuytrate aminotransferase
	5380087	G -> T	Ala -> Asp	**Cbei_RS23695/Cbei_4644** Transketolase
DCOG5-1	2968728	G -> T	-	Non-coding region
	3592439	G -> T	Ser -> STOP	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	4742319	C -> A	Asp -> Tyr	**Cbei_RS21095/Cbei_4124** Methyl-accepting chemotaxis sensory transducer
DCOG6-1	3593531	Deletion	-	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	5644970	C -> A	Ala -> Ser	**Cbei_RS24620/Cbei_4828** Chemotaxis protein
DCOG6-2	3593235	C -> A	Glu -> STOP	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
CIC1	3591369	G -> T	Pro -> Thr	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3922555-3922558	Deletion -> ATAA	-	**Cbei_RS17295** AraC transcriptional regulator
CIC2	82285	G -> T	Lys -> Asn	**Cbei_RS00360/Cbei_0062** Prolipoprotein diacylglyceryl transferase
	2124238	C -> A	Ala -> Asp	**Cbei_RS09540/Cbei_1835** MFS transporter
	3375032	C -> A	Glu -> STOP	**Cbei_RS14905/Cbei_2890** ABC transporter substrate binding protein
	3592697	G -> T	Ser -> STOP	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	4646727	G -> T	Non-coding region	Intergenic region just before 2-oxoacod:ferredoxin oxidoreductase subunit alpha
CIC3	1981516	G -> T	Ala -> Ser	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
	3011101	G -> T	Gly -> STOP	**Cbei_RS26260/Cbei_2588** Hypothetical protein - collagen triple helix repeat
CIC4	28867	C -> A	Phe -> Leu	**Cbei_RS00130/Cbei_0017** Two component sensor histidine kinase
	3972442	C -> A	Gly -> Val	**Cbei_RS17560/Cbei_3422** Galactose ABC transporter substrate binding protein
CIC5	3706256	C -> A	Ser -> STOP	**Cbei_RS16310/Cbei_3170** Conserved hypothetical protein between accessory gene regulator B and multi-sensor signal transduction histidine kinase
CIC6	28718	G -> T	Glu -> STOP	**Cbei_RS00130/Cbei_0017** Two component sensor histidine kinase
	1142639	G -> T	Ala -> Ser	**Cbei_RS05125/Cbei_0957** PTS sugar transporter
	3217052	C -> A	Gly -> Trp	**Cbei_RS14285/Cbei_2767** ABC transporter
CIC1-1	334943	G -> T	-	Non coding region
	2551452	C -> A	Ser -> STOP	**Cbei_RS11310/Cbei_2184** Histidine kinase
	3591533	A -> T	Met -> Lys	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3591535	T -> C	-	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3591539	A -> G	Ile -> Thr	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	3591533 - 3591717	Deletion 185bp		**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	4183310	C -> A	Gly -> Cys	**Cbei_RS18625/Cbei_3633** Short-chain dehydrogenase
	5508820	C -> A	Ala -> Ser	**Cbei_RS24110/Cbei_4727** Cell wall-binding protein
CIC1-2	1809360	C -> A	Leu -> Ile	**Cbei_RS08045/Cbei_1534** Preprotein translocase subunit YajC
	1981107	Deletion	-	Non-coding region 280 bp before Spo0A

	45118061	C -> A	-	Non-coding region before cell wall-binding repeat-containing protein
CIC2-1	5979332	G -> T	Ser -> STOP	**Cbei_RS26055/Cbei_5080** ATPase AAA
CIC2-2	4221171	G -> T	-	Non-coding region
	5537882	C -> A	Asp -> Tyr	**Cbei_RS24215/Cbei_4748** Hypothetical protein
	5720359	C -> A	-	Non-coding region
CIC3-1	3417009	C -> A	Ala -> Ser	**Cbei_RS15065/Cbei_2917** Acriflavin resistance protein
	3591584	C -> A	Gly -> Val	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	4471097	C -> A	Val -> Leu	**Cbei_RS19900/Cbei_3887** Hypothetical protein
CIC3-2	3591584	C -> A	Gly -> Val	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
CIC4-1	1981493	G -> T	Gly -> Val	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
	2148216	G -> T	-	Non-coding region
	3972887	G -> T	His -> Asn	**Cbei_RS17560/Cbei_3422** Galactose ABC transporter substrate-binding protein
CIC4-2	3593522	G -> T	Ser -> Tyr	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	5380087	G -> T	Ala -> Asp	**Cbei_RS23695/Cbei_4644** Transketolase
CIC5-1	3591437	G -> T	Thr -> Asn	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
CIC5-2	595319	G -> T	Lys -> Asn	**Cbei_RS02805/Cbei_0494** Penicillin-binding protein
	3591437	G -> T	Thr -> Asn	**Cbei_RS15855/Cbei_3078** Hybrid sensor histidine kinase/response regulator
	5799598	G -> A	-	Non-coding region
CIC6-1	5980548	C -> A	Gly -> STOP	**Cbei_RS26055/Cbei_5080** ATPase AAA
CIC6-2	3706328	C -> A	Gyl -> Val	**Cbei_RS16310/Cbei_3170** Hypothetical protein
	3928585	C -> A	Ala -> Ser	**Cbei_RS17320/Cbei_3377** PAS domain-containing histidine kinase
FW1	-	-	-	NO SNPs
FW2	812638	C -> T	Arg -> Cys	**Cbei_RS03690** Transposase
	1981504	G -> T	Asp -> Tyr	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
FW3	1981661	G - > T	Gly -> Val	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
	3402338	C -> A		**Cbei_RS15025** repeat region rRNA
FW4	1982059	C -> A	His-> Asn	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
FW5	325050	G -> T	-	Non-coding region between histidine kinase and hypothetical protein
	1982189	G -> T	Arg -> Ile	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
	2779914	C -> A	Pro -> Thr	**Cbei_RS12440/Cbei_2405** 2-Hydroxyglutaryl-CoA dehydratase
FW6	1100481	G -> T	Arg -> Leu	**Cbei_RS04935** Phage tail tape measure protein
	1982179	G -> T	Asp -> Tyr	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
	2920250	G -> T	Lys -> Asn	**Cbei_RS13000/Cbei_2616** Histidine Kinase
FW7	1982057	G -> T	Arg -> Ile	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
	3422197	G -> T	No change	**Cbei_RS15080/Cbei2920** Two component sensor histidine kinase
FW1-1	1283334	C -> A	Pro -> Thr	**Cbei_RS05695/Cbei_1070** ACP S-malonyltransferase
	1955903	G -> T	Ala -> Ser	**Cbei_RS08755/Cbei_1679** Methyltransferase
	2003933	G -> T	Glu -> STOP	**Cbei_RS09010/Cbei_1731** DNA-binding response regulator
	5720653	G -> T	-	Non-coding region

FW1-2	1288499	G -> T	Ala -> Ser	**Cbei_RS05720/Cbei_1075** Acetyl-CoA carboxylase biotin carboxylase subunit
	1915854	G -> T	Val -> Phe	**Cbei_RS08535/Cbei_1634** Hypothetical protein
	1982002	C -> A	Pro -> Thr	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
FW2-1	3046536	G -> T	-	Non-coding region
	5720577	G -> T	-	Non-coding region
FW2-2	3046536	G -> T	-	Non-coding region
	5720577	G -> T	-	Non-coding region
FW3-1	1982167	G -> T	Ala -> Ser	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
FW3-2	531609^531610	Insertion of A	-	Non-coding region
	5122426	G -> T	Asp -> Glu	**Cbei_RS22675/Cbei_4438** Chemotaxis protein
	5720108^5720109	Insertion	Introduced Asp	**Cbei_RS24895/Cbei_4884** Hypothetical protein
	5720114	C -> A	Phe -> Asp	**Cbei_RS24895/Cbei_4884** Hypothetical protein
	5720115	A -> T	Leu -> Tyr	**Cbei_RS24895/Cbei_4884** Hypothetical protein
	5720109-5720407	Deletion 299bp	-	Part of at start **Cbei_RS24895/Cbei_4884** Hypothetical protein and then non-coding
	5720407	A -> T	-	Non-coding region
FW4-1	572953	G -> T	Ser -> Ile	**Cbei_RS02710/Cbei_0475** Phosphoesterase
	1981939	G -> T	Glu -> STOP	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
FW4-2	1736196	G -> T	Ala -> Asp	**Cbei_RS07735/Cbei_1473** Transposase
	1981426	G -> T	Asp -> Tyr	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
FW5-1	669358	C -> T	No change	**Cbei_RS03120/Cbei_0558** Transposase
	4873316	C -> A	-	Non-coding region
	4908300	C -> A	Glu -> STOP	**Cbei_RS21735/Cbei_4249** RNA polymerase sigma factor
	5439105	A -> T	-	Non-coding region
	5439106	C -> T	-	Non-coding region
	5439107	T -> A	-	Non-coding region
	5720174	Deletion G	-	Non-coding region between **Cbei_RS24895/Cbei_4884** Hypothetical protein and abrB Family transcriptional regulator
	5720181-5720489	Deletion 309 bp	-	Non-coding region between **Cbei_RS24895/Cbei_4884** Hypothetical protein and abrB Family transcriptional regulator
	5720494	G -> T	-	Non-coding region between **Cbei_RS24895/Cbei_4884** Hypothetical protein and abrB Family transcriptional regulator
	5720497	Deletion	-	Non-coding region between **Cbei_RS24895/Cbei_4884** Hypothetical protein and abrB Family transcriptional regulator
	5720500	T -> A	-	Non-coding region between **Cbei_RS24895/Cbei_4884** Hypothetical protein and abrB Family transcriptional regulator
	5720504	C -> T	-	Non-coding region between **Cbei_RS24895/Cbei_4884** Hypothetical protein and abrB Family transcriptional regulator
	5720506	A -> T	-	Non-coding region between **Cbei_RS24895/Cbei_4884** Hypothetical protein and abrB Family transcriptional regulator
FW5-2	501672	G -> T	Gly -> Val	**Cbei_RS02350/Cbei_0412** ATP synthase F0F1 subunit A
	513931	G -> T	Glu -> STOP	**Cbei_RS02415/Cbei_0425** Rod shape-determining protein
	1037715	G -> T	No change	**Cbei_RS04620/Cbei_0854** SAM-dependent methyltransferase

	1982059	C -> A	His -> Asn	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
FW6-1	1657287	C -> T	Ala -> Thr	**Cbei_RS07410/Cbei_1410** Phosphotase
	1981661	G -> T	Gly -> Val	**Cbei_RS08920/Cbei_1712** Sporulation transcription factor Spo0A
	2853156	G -> T	Met -> Ile	**Cbei_RS12735/Cbei_2464** GTPase
	3100336	G -> T	Gly -> Cys	**Cbei_RS13780/Cbei_2670** Hypothetical protein
	3674732	G -> T	Met -> Ile	**Cbei_RS16190/Cbei_3145** MATE family efflux transporter
	4390681	C -> A	Gly -> Val	**Cbei_RS19575/Cbei_3823** Chemotaxis protein
FW6-2	2641823	G -> T	Arg -> Ile	**Cbei_RS11805/Cbei_2280** Hydroxylamine reductase
	3013492	C -> A	Phe -> Leu	**Cbei_RS13380/Cbei_2589** Beta 1,4 glycosyltransferase
	5720577	G -> T	-	Non-coding region

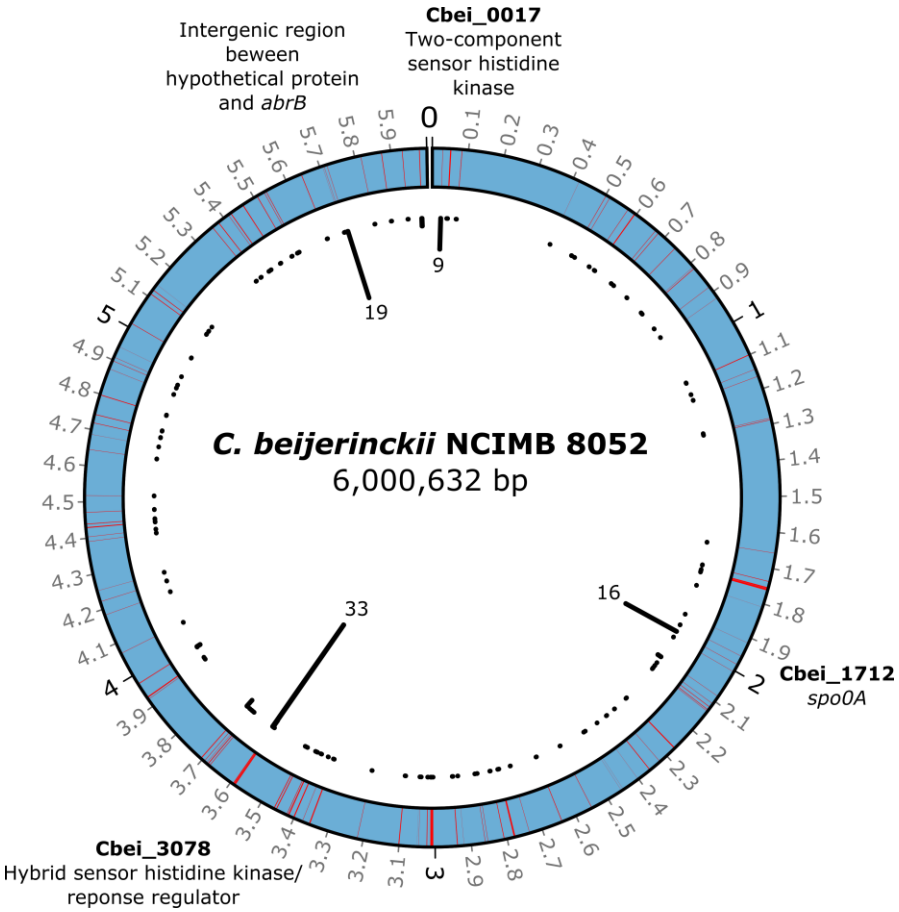


Figure 4.1 Non-uniform chromosomal distribution of mutations in degenerate strains.

A chromosomal image representing the distributions of mutations in degenerate isolates throughout the genome. Red lines indicate where a mutation was found and black lines in the centre of the image show the number of times that region was hit. Genomic regions that contained significantly more mutations than others are highlighted. The genome image was kindly created by Cynthia Chibani, Göttingen Genomics Laboratories, using the software Circus.

The parental wild types contained several mutations that were different to that of the reference genome. These included a deletion in a phage tail protein, a SNV in a paraslipin gene, an insertion in a lactate permease and a SNV in a non-coding region before a phosphohydrolase. Interestingly, the first parental wild type contained an SNV in a histidine kinase (C.bei_4430) which was not present in the second parental wild type. The second parental wild type however had an SNV in a non-coding region that was not present in the first. Chapter 3 showed normal phenotypes for our wild type and so the mutations were not believed to be related to degeneration. As expected, all of the mutations found in the parental wild types were found in the respective degenerates that were derived from them. These common mutations were filtered out as they were not relevant to the comparisons.

From genomic comparisons of all of the degenerates, a total of 211 SNVs, 12 deletions and 2 insertions were found. This gave an average genetic mutation rate of 3.16 per degenerate after five serial transfers (equating to approximately 16-17 generations). There was an interesting difference in the type of nucleotide base change that occurred. The most common SNV was a guanine to thymine (G -> T) base change which totalled 104 out of the 211 SNVs. The second was cytosine to adenosine (C -> A) which equalled 80 out of the 211 SNVs. Thus, given that there are 12 possible changes (4 different transitions and 8 different transversions), these two transversions were over-represented as they made up 87% of the observed SNVs. Following from these two SNVs, there were 8 A -> T, 7 C -> T, 4 A -> G, 4 T -> C, 2 G -> A and 2 T -> A making up 3.8%, 3.3%, 1.9%, 1.9%, 1% and 1% of the remaining SNVs respectively.

Although there were many mutations, many of these were in non-coding regions that did not cause any amino acid change. It was found however that several genomic

regions contained significantly more mutations than other areas of the genome. The most common area to be hit was that of a gene labelled as a hybrid sensor histidine kinase/response regulator (locus tag Cbei_3078). This region was hit a total of 33 times with 21 out of 71 of the isolates containing mutations in this region. Interestingly, several isolates contained the same mutations. One such example of this is a deletion of 185 bp found in RD2-2, CIC1-1, DCOG1-1 and DCOG1-2. CIC1-1, DCOG1-1 and DCOG1-2 were also isolated from the tube so the same mutation seems appropriate however RD2-2 was isolated separately making the exact same mutation unusual. Several other SNVs can be seen which are shared (Figure 4.2).

The region to contain the second most mutations was one containing a hypothetical protein (Cbei_4884), the intergenic region between the hypothetical protein and that of an *abrB* transcriptional regulator gene and the *abrB* (Cbei_4885) gene itself. The total number of mutations was 19 however only 9 out 71 isolates had mutations in this region. FW3-2 and FW5-1 made up the majority of mutations in this region as they had 5 and 7 of the 19 mutations respectively. It was interesting to find that most of the mutations in this region were in the non-coding area between the genes. Perhaps the promoters of both the genes were affected or they are linked in some other way.

The third region to be hit the most was that of the master regulator gene *spo0A* (Cbei_1712). The number of mutations totalled 16 with 16 out of 71 isolates having mutations in this gene. s*po0A* is known to control solventogenesis and spore formation making mutations in this region certain to affect both (Ravagnani et al., 2000).

The final hotspot region was that of a two-component sensor histidine kinase (Cbei_0017) which had a total of 9 SNVs from 9 out of 71 isolates. There appeared to

be no direct relation between the regions as each of the four hotspot regions were distributed throughout the genome and not clustered together (Figure 4.1). To confirm that the observed SNVs were real and not an artefact of the Illumina sequencing or the assembly and SNV calling pipelines, relevant sections of the four hot spot regions were PCR-amplified and subjected to Sanger sequencing. All mutations in these regions were thus independently confirmed. A more in depth analysis of the distribution of mutations for each of these regions can be seen in Figure 4.2.

The distribution of SNVs in these four regions was not centralised to one location as several mutations were found throughout each of the hotspot genes (Figure 4.2). For Cbei_0017, *spo0A* and Cbei_3078, most mutations were in the known domains of the proteins which may affect specific function (Figure 4.2A, B, C). Several SNVs were found in the specific DNA binding motif of *spo0A* which if affected can cause serious defects in Spo0A (Ravagnani et al., 2000). Several isolates (RD2-2, CIC1-1, DCOG1-1 and DCOG1-2) all had the same SNV and deletion in and around the CheY-like domain of the protein, a domain family that has been linked to Spo0A (Volz, 1993). Most mutations for Cbei_4884 and Cbei_4885 were in the intergenic region separating these two genes (Figure 4.2D).

Although there were four main mutation regions, several other regions were of particular interest. One being a TetR family transcriptional regulator (Cbei_3077) next to the histidine kinase Cbei_3078 mentioned previously. This gene was hit 3 times in the isolates RD2, DCOG3-2 and DCOG4. Another being the hypothetical protein Cbei_3170 which CIC5, RD6-2 and CIC6-2 all had mutations in. This gene may have a link to Agr quorum sensing which has been shown to affect the level of sporulation in

C. acetobutylicum ATCC 824 (Kotte et al., 2017; Steiner et al., 2012), as it is found next to an *agrB* homologue.

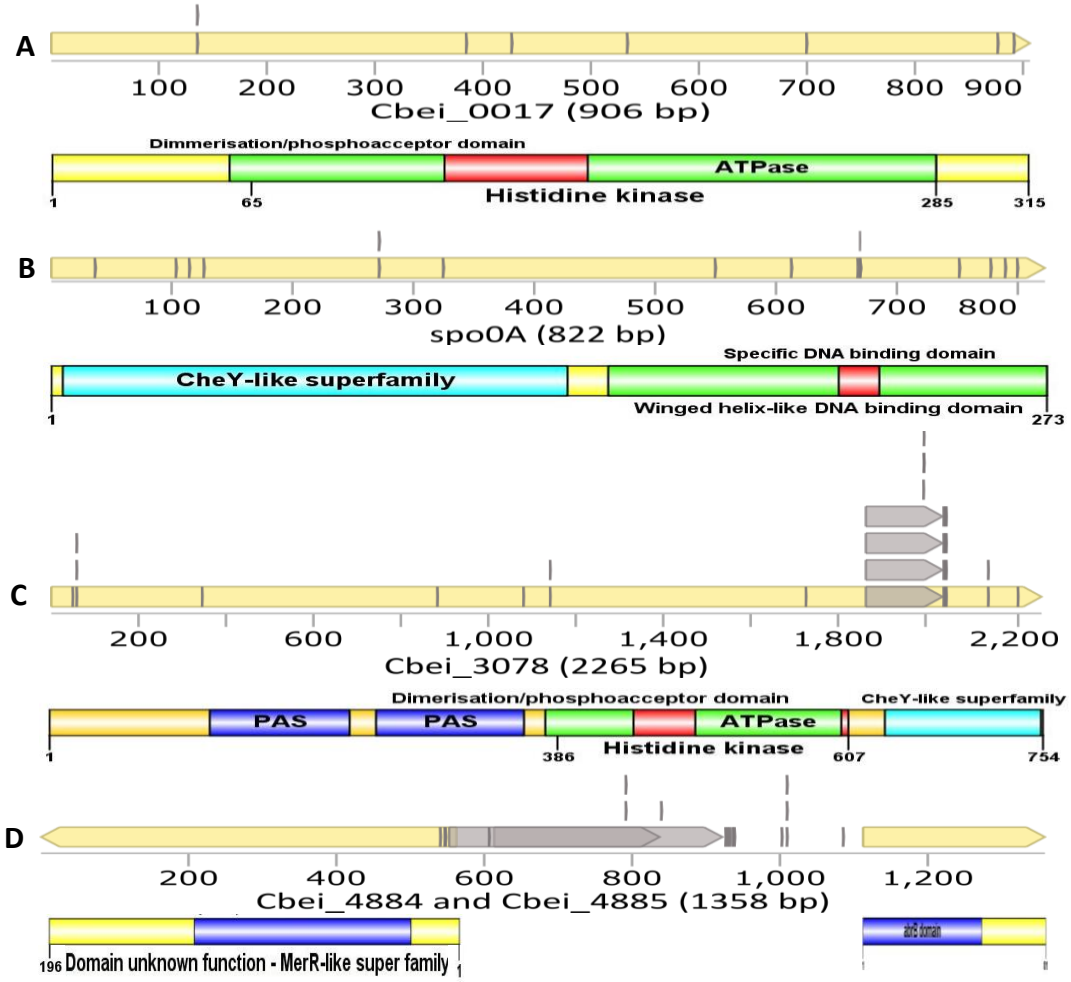


Figure 4.2 In depth analysis of the mutation position in the nucleotide sequence for the four most hit regions and the protein domains for three of the genes.

Positions of mutations are highlighted in grey. Single lines represent SNVs or base insertions and long regions represent deletions. If the same position was hit twice, the line was placed above. Protein sequences show domains found using InterPro (www.ebi.ac.uk/interpro). Cbei_0017 (A), spo0A (B), Cbei_3078 (C) and Cbei_4884, Cbei_4885 and the intergenic region between them (D) are shown. Image created using Benchling (www.benchling.com). Protein images created using Illustrator for Biological Sequences (Liu et al., 2015a).

Isolates FW1, DCOG5, RD1, RD3, CIC1-2, CIC2-1 CIC6-1, DCOG3-1 and FW6-2 contained no apparent mutations that would cause their phenotype. As mentioned earlier only mutations that were present in at least 90% of the reads were counted and so potential mutations may have been missed. Many isolates including RD3, DCOG1, DCOG2, DCOG3, DCOG4, DCOG5, DCOG1-1, DCOG2-2, DCOG3-2, DCOG5-2 DCOG6-1, DCOG6-2, CIC1, CIC2, CIC4, CIC5, CIC6, CIC1-1, CIC3-2, CIC4-2, CIC6-2, and FW3 contained SNVs between a diguanylate cyclase and an integrase. All of these SNVs were only ever present in around 50% of the reads with most of these SNVs found in the same place at the same location. It is unclear whether these mutations are significant or merely errors in the sequencing/alignment process. As is to be discussed further on, diguanylate cyclase are known to regulate several important cellular processes which may explain the lack of evidence for a cause of some degenerate phenotypes.

4.3.2. Creation of a Δ*pyrE* mutant of *C. beijerinckii* NCIMB 8052

The original intention for the creation of a Δ*pyrE* mutant of *C. beijerinckii* NCIMB 8052 was to perform in-frame deletions of the four hotspot genes. This is achieved by using the newly acquired FOA resistance, once the *pyrE* has been truncated, as a counter selection marker for in-frame deletions (Ehsaan et al., 2016; Ng et al., 2013). This was attempted but proved ineffective and time consuming for the chosen strain and so was not achieved within the available time frame. ClosTron mediated gene disruption was therefore employed to allow for quick gene inactivation to ascertain phenotypes (4.4.3). The Δ*pyrE* counter selection system, however was still required for

chromosomal integration of complementation cassettes encoding the genes disrupted by the ClosTrons.

To create a Δ*pyrE* mutant the plasmid pMTL-JRH1 was created (see methods 2.3.3.1). This plasmid contained two homology arms that would truncate the *C. beijerinckii pyrE* gene rendering it inactive, providing subsequent resistance to FOA (5-fluoroorotic acid). The left homology arm on the plasmid contained 300 bp of the *pyrE* gene but with a TAA stop codon at the end of the 300 bp sequence (primers pyrE KO LHA_F and pyrE KO LHA_R, see 2.1). The right homology arm contained the 1200 bp downstream region of the *pyrE* gene (primers pyrE KO RHA_F and pyrE KO RHA_F). Each homology arm was cloned into the shuttle vector pMTL83251 (Heap et al., 2009). When *C. beijerinckii* NCIMB 8052 clone 3 was transformed with pMTL-JRH1 (see methods section 2.25), the resulting transformants were selected by growth in 10 µg/ml erythromycin for plasmid uptake. These transformants were further restreaked and selected for by the ability to grow on CBM-S 6% glucose supplemented with 600 µg/ml FOA and 2 µg/ml uracil. Any colonies that grew were expected be uracil auxotrophs and, due to their Δ*pyrE* genotype, FOA resistant. Colonies were tested for *pyrE* truncation using the flanking primers *pyrE* KO confirmation_F and *pyrE* KO confirmation_R (2.1). The results for 12 isolated colonies can be seen below in Figure 4.3.

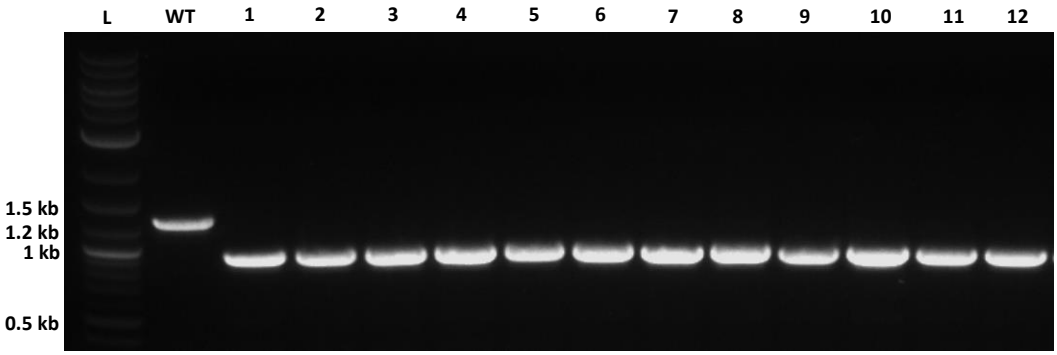


Figure 4.3 PCR confirmation of *pyrE* truncation in candidate *C. beijerinckii* Δ*pyrE* strains.

PCR amplification of the *C. beijerinckii pyrE* region in wild type (WT) and 12 candidate (1-12) mutant strains using primer pair pyrE KO confirmation_F/pyrE KO confirmation_R. All 12 colonies selected had a truncated gene as the PCR product showed the expected decrease in size compared to the wild type. The size of PCR products was estimated by 1% agarose gel electrophoresis using a 2-log ladder (L).

The wild type DNA should give a PCR product of 1330 bp and that of the *pyrE* mutant should be 955 bp. This can clearly be seen for all of the candidate strains, proving that they now had a truncated *pyrE* gene. All of these mutants could grow on media supplemented with FOA, again confirming the truncation of the *pyrE* gene.

Plasmid loss was tested via PCR amplification of the pMTL-JRH1 plasmid and isolates 3, 4, 5 and 8 were taken forward for phenotypic testing as they had lost the pMTL-JRH1 knock-out plasmid (data not shown). Each isolate was tested for its solvent, acid and spore producing capabilities and the isolate that performed most similar to the wild type (considering that the media now contained uracil) would be taken forward in future experiments.

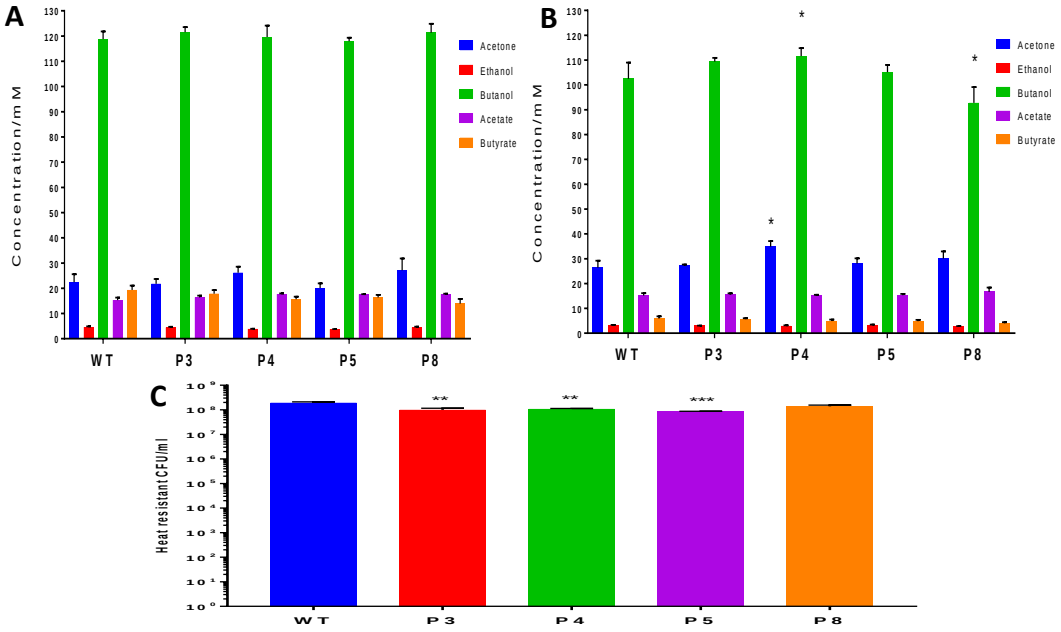


Figure 4.4 Fermentation products and spore counts of 4 confirmed Δ*pyrE* mutants.

Fermentation products after (A) 2 days (48 hours) and (B) 5 days (120 hours) of growth in CBM-S 6% glucose. Heat resistant CFU/ml (C) obtained after 5 days (120 hours) of fermentation. Technical replicates represent standard error of the mean. *n* = 3 for each of the graphs. Fermentation products are compared against the wild type using two way ANNOVA. Heat resistant CFU/ml is compared using one way ANNOVA against the wild type. Results show P values were of: * ≤ 0.05, ** ≤ 0.01 and *** ≤ 0.001.

The fermentation profiles of the 4 isolates proved to very similar to the wild type. Strain P4 produced slightly more solvents than the wild type and P8 produced less butanol at the 2 day sample point. This difference was not observed at the end point sampling. These differences were only minor and could be attributed to slightly different growth on the day. There was more variation to the wild type for the spore counts as P3, P4 and P5 had produced statistically less spores. Spore numbers were still higher than all degenerate isolates so this difference was not assumed to be an issue with the strain.

Each isolate was further sent for whole genome sequencing to compare against the wild type to assess the genetic variation. This was again completed by Illumina sequencing by MicrobesNG (Birmingham, UK) and the sequences were processed as before.

Table 4.2 The list of mutations the isolated Δ*pyrE* had compared to the wild type.

Strain	Position	Genetic change	Amino acid change	Gene affected
ΔpyrE 3	1173259	G -> T	Gly -> Val	**Cbei_RS05245/Cbei_0981** Chromosomal segregation protein SMC
ΔpyrE 4	586393	C -> A	Asp -> Tyr	**Cbei_RS02760/Cbei_0485** Pyruvate kinase
	2412990	G -> T	Ser -> STOP	**Cbei_RS10755/Cbei_2072** ABC transporter-substrate binding protein
	2494683	G -> T	Gln -> His	**Cbei_RS11080/Cbei_2138** Ferrichrome ABC transporter-substrate binding protein
ΔpyrE 5	1543942	G -> T	-	Intergenic region before UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase
	2252433	C -> A	No change	**Cbei_RS10100/Cbei_1950** Hypothetical protein
ΔpyrE 8	586393	C -> A	Asp -> Tyr	**Cbei_RS02760/Cbei_0485** Pyruvate kinase
	1593053	G -> T	Glu -> STOP	**Cbei_RS07120/Cbei_1353** Hypothetical protein
	2412990	G -> T	Ser -> STOP	**Cbei_RS10755/Cbei_2072** ABC transporter-substrate binding protein
	2494683	G -> T	Gln -> His	**Cbei_RS11080/Cbei_2138** Ferrichrome ABC transporter-substrate binding protein

It was unfortunate that several of the isolates had acquired undesirable mutations. P4 and P8 had gained mutations in the same genes (Cbei_0485, Cbei_2072 and Cbei_2138) and had clearly been isolated as daughter colonies. P3 had a mutation in Cbei_0981 which was a chromosomal segregation protein making use of this isolate potentially problematic. Although isolate P5 had two mutations, one was in a non-coding region and the second caused a silent mutation. It was therefore decided that all future work would be carried out with strain P5.

4.3.3. Creation of ClosTron mutants, repair of the *pyrE* gene and mutant complementation

As mentioned previously, the intended use of a *ΔpyrE* strain was to create in-frame deletions as employed in the literature (Ehsaan et al., 2016; Ng et al., 2013). Attempts were made to create in-frame deletion mutants of the four hotspot regions which included the genes for the identified two-component sensor histidine kinase (Cbei-_0017), *spo0A* (Cbei_1712), hybrid sensor histidine kinase/response regulator (Cbei_3078), hypothetical protein (Cbei_4884) and the AbrB transcriptional regulator (Cbei_4885). It was further attempted to make an in-frame deletion of the *TetR*-type transcriptional regulator gene (Cbei_3077). Making the in-frame deletions using this method proved to be ineffective and time consuming in this *C. beijerinckii* strain and it was therefore deemed necessary to use alternative methods.

ClosTron gene disruption was subsequently chosen as this has been proven to be quick, robust and already effective in *C. beijerinckii* NCIMB 8052 (Heap et al., 2010; Kuehne et al., 2012). ClosTron target sites were designed using the algorithms found at www.clostron.com for all of the previously mentioned genes and the highest target scores were selected. The following ClosTron plasmids were synthesised by ATUM (formerly DNA 2.0): pMTL007S-E2::cbei_0017-694|695a, pMTL007S-E2::cbei_3078-969|970s, pMTL007S-E2::cbei_3077-138|139a, pMTL007S-E2::cbei_4884-159|160s and pMTL007S-E2::cbei_4885-108|109s. pMTL007S-E2::Cbe-spo0A-407a already existed in our culture collection (Heap et al., 2010) and was used to inactivate *spo0A*.

The created *C. beijerinckii* NCIMB 8052 *ΔpyrE* strain (P5) was transformed with each of the plasmids and plasmid uptake was initially selected for by plating on CGM agar containing 750 µg/ml spectinomycin. The resulting transformants were selected for ClosTron insertion by restreaking onto CGM containing 10 µg/ml erythromycin.

Suspected mutants were picked and insertions screened using flanking primers of the gene of interest (2.1). Within the available timeframe, it was not possible to obtain ClosTron mutants for the *abrB* gene, Cbei_4884 or the TetR family transcriptional regulator gene. Each of these genes had produced low scores on the used ClosTron targeting algorithm and therefore the absence of mutants was not entirely unexpected. Three independent ClosTron mutants were isolated for Cbei_0017, Cbei_3078 and *spo0A* by two BMedSci students under my supervision, Charmaine Chan and Panicha Pusiritanachote. Initial confirmation of insertion was carried out by PCR amplification using primers that flanked the target genes (2.1) (data not shown). PCR products were sent for Sanger sequencing to further confirm insertion (data not shown).

Once ClosTron mutants had been confirmed through the flanking primer PCR and that isolates had lost the ClosTron plasmid, Δ*pyrE* repair and complementation was undertaken. The plasmid pMTL-JRH4 contained two homology arms, the left homology arm contained an intact *pyrE* gene (primers pyrE repair LHA_F and pyrE repair LHA_R) and the right homology arm contained a region of 1200 bp starting immediately downstream of the *pyrE* gene (primers pyrE repair RHA_F and pyrE repair RHA_R). Each homology arm was cloned into the pMTL84351 vector (Heap et al., 2009) either side of a lacZα multiple cloning site to allow for complementation genes to be inserted. Complementation vectors were created with the addition of the native gene between the two homology arms, i.e. the gene that had been disrupted by ClosTron mutagenesis together with approximately 200 bp upstream region so as to include the native promoter. Complement vectors were created for pMTL-JRH4_Cbei_0017, pMTL-JRH4_ Cbei_3078 and pMTL-JRH4_ *spo0A.*

Each of the 3 independent ClosTron mutants per gene were transformed with pMTL-JRH4 and the derived respective complementation vector. The P5 strain was also transformed with pMTL-JRH4 to repair the Δ*pyrE* mutation of this strain. This was to isolate a comparative "wild type" strain that would be used for fair phenotypic comparisons as it was derived from the same P5 starting clone as the *pyrE* repaired ClosTron mutants. Isolates were selected for based on the ability to grow purely on CBM-S agar 6% glucose that had not been supplemented with either uracil or FOA. Colonies that could grow in the absence of uracil should have their *pyrE* gene functionality restored. The same primers used to check for *pyrE* knock out were used to check for restoration and/or complementation. Figure 4.5B shows the *ΔpyrE* repair and complementation.

If the *pyrE* gene had been repaired in the ClosTron mutants, a product size of approximately 1600 bp long would be observed compared to the 1330 bp found of the wild type. This was indeed the case and due to the addition of the *lacZ*α multiple cloning site on the repair vector. Chromosomal gene complementation was confirmed through increased *pyrE* PCR fragment sizes, as this led to the addition of 1113 bp for Cbei_0017, 2314 bp for Cbei_3078 and 1143 bp for *spo0A.* The PCR products for the complements were checked by Sanger sequencing to confirm the correct gene sequence had been inserted.

A final confirmation of the ClosTron insertion for the final *pyrE* repaired and complemented strains was obtained by PCR amplification again using primers that flanked the target genes (Figure 4.5A). It is clear from Figure 4.5A and B that *pyrE* repair had been successful and that the original ClosTron insertion was still present.

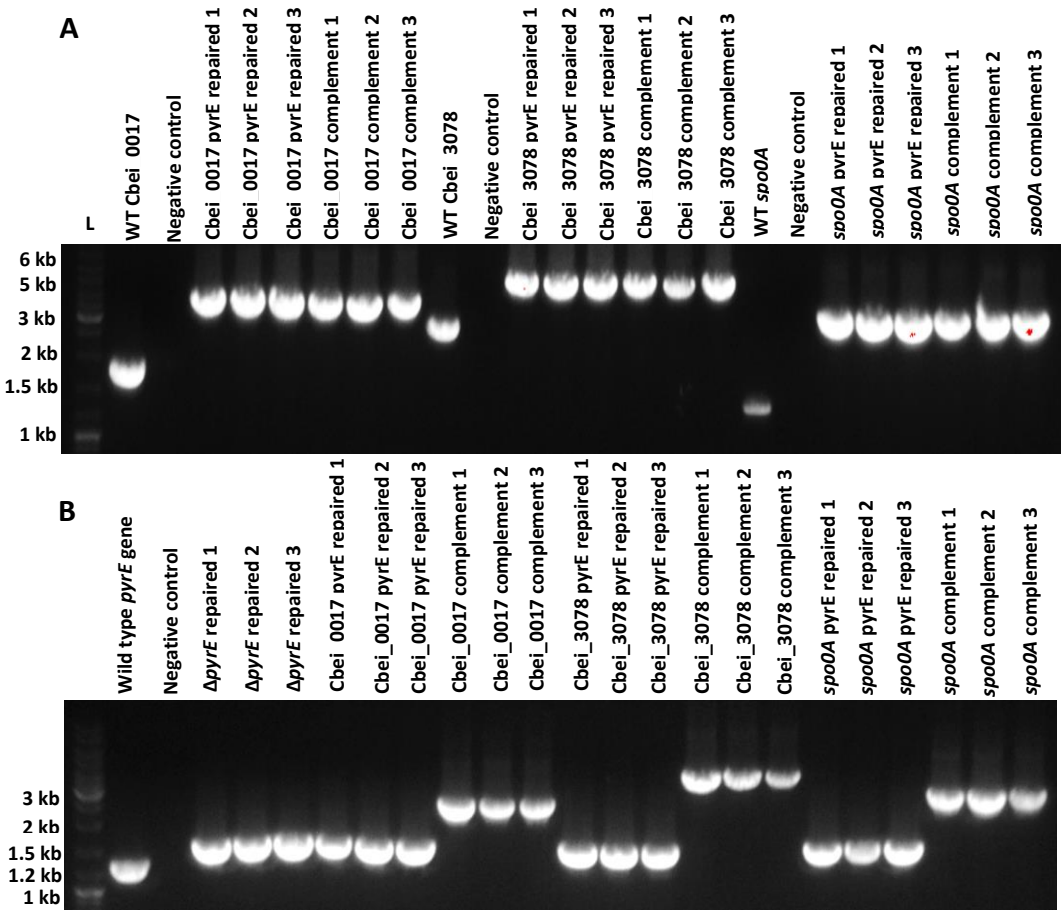
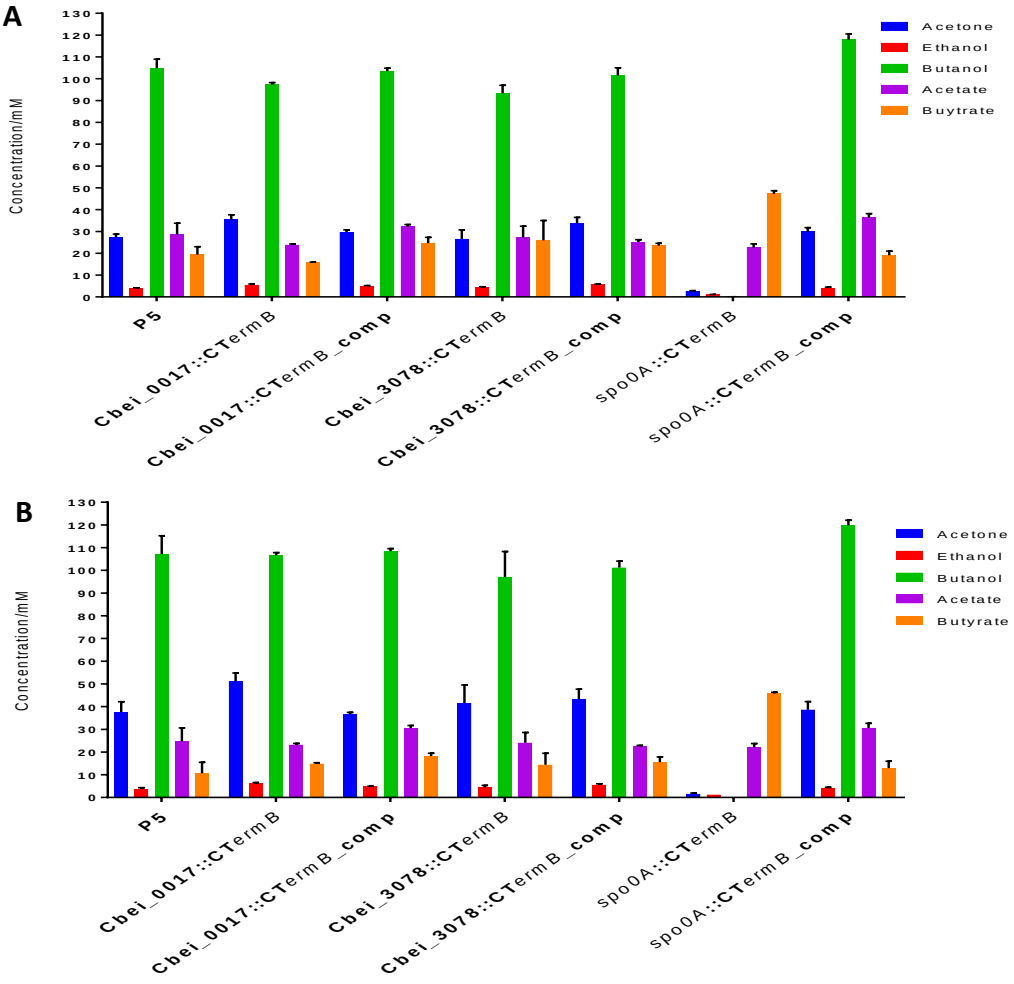


Figure 4.5 PCR confirmation of ClosTron mediated gene inactivation in *pyrE* repaired Cbei_0017, *spo0A* and Cbei_3078 and complemented mutants (A) and PCR confirmation of Δ*pyrE* repair with or without gene complementation (B).

PCR amplification of the Cbei_0017, *spo0A* and Cbei_3078 gene regions from both the *pyrE* repaired mutants and complement mutants using primers 0017 check_F/R, spo0A check_F/R and 3078 check_F/R (A) (2.1). PCR amplification of the *pyrE* gene regions of *pyrE* repaired parental strain derivatives (Δ*pyrE* repaired 1-3), mutants (Cbei_0017 repaired 1-3, *spo0A* repaired 1-3 and Cbei_3078 repaired 1-3) and complemented mutants (Cbei_0017 complement 1-3, *spo0A* complement 1-3 and Cbei_3078 complement 1-3) using primers *pyrE* confirmation KO_F/R (B) (2.1). Wild type (WT) and negative controls (water) were also amplified. PCR fragments were separated by 1% agarose gel and compared to a 2-log ladder (L).

4.3.4. Phenotypic characterisation of the created mutants

Once all isolates had been confirmed for *pyrE* repair, complementation and ClosTron insertion they were grown in triplicate in 25 ml CBM-S broth 6% glucose alongside *C. beijerinckii* NCIMB 8052 Δ*pyrE*::pMTL-JRH4 repaired which here after will be referred to as P5. Solvent samples were taken at 2 days and 5 days post inoculation (Figure 4.6A+B). Spore samples were taken after 5 days of growth (Figure 4.6C).



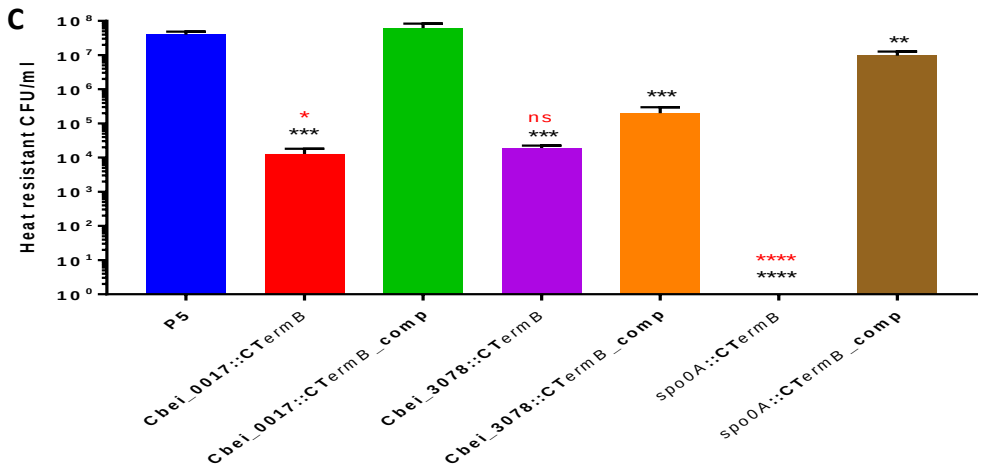


Figure 4.6 Fermentation products and sporulation capacity of Cbei_0017, Cbei_3078 and *spo0A* ClosTron mutants.

Fermentation products after (A) 2 days (48 hours) and (B) 5 days (120 hours) of growth in CBM-S 6% glucose. Sporulation was assessed as heat resistant CFU/ml (C) obtained after 5 days (120 hours) of fermentation. Mutants with repaired *pyrE* are labelled CT::*ermB* and complemented mutants named CT::*ermB*_comp. Biological replicates represent standard error of the mean. *n = 3* which is the average of each of the technical triplicates used to make the 3 biological replicates. A summary of the statistical differences between strains fermentation products can be found in the appendix. The spore counts were analysed using T-test against P5 and against the respective complemented strain. Results above show P values of: * ≤0.05, **≤0.01 ***≤0.001 and **** ≤ 0.0001. Black indicates statistically differences against P5 and red against the respective complemented strain.

With the exception of *spo0A*::CT*ermB* solvent samples of the other ClosTron mutants with and without complementation were not that dissimilar from that of P5. Concentrations of butanol remained high for all strains, besides *spo0A*::CT*ermB pyrE* repaired, at around 100 mM with the *spo0A*::CT*ermB*_comp appearing to produce the highest concentration of butanol at 119.8 ± 2.3 mM and 117.8 ± 2.7 mM at the 2 and

5 day point respectively. This however was not statistically different to the other strains. The Cbei_0017::CT*ermB pyrE* repaired strain appeared to produce more acetone at the 2 day sample point than P5 however this was not statically significant. This was also seen from some of the natural degenerate strains. The *spo0A*::CT*ermB pyrE* repaired strains produced no butanol and very little acetone and ethanol.

The *spo0A*::CT*ermB* strains produced significantly more acids, namely butyrate, than the other strains, a phenotype that was previously observed from the isolated FW strains. It should be noted that the *spo0A*::CT*ermB* strains looked like flat and white colonies (Figure 4.7F). It was also possible to restore colony morphology by complementing the *spo0A* gene (Figure 4.7G). All other strains produced acid concentration that were not statistically different to P5.

Even though solvent production appears to be unaffected, besides *spo0A*::CT*ermB pyrE* repaired*,* the level of sporulation was significantly different to P5 for all of the uncomplemented strains. Cbei_0017::CT*ermB pyrE* repaired produced significantly less spores than both P5 and the complemented strain with heat resistant CFU/ml decreasing from 10^7 to 10^4. Complementation clearly saw a restoration of sporulation. Similarly the Cbei_3078::CT*ermB* saw heat resistant CFU/ml drop to around 10^4 however the complementation did not restore completely sporulation. This is most likely due to a lack of a native promoter and, potentially, polar effects of the ClosTron insertion, which will be discussed later. The most striking change in sporulation was that seen with the *spo0A*::CT*ermB* as no spores could be detected. However, complementation was able to restore sporulation back levels near that of P5. It appears that the histidine kinase genes Cbei_0017 and Cbei_3078 do not affect solventogenesis but clearly are involved in spore formation. A selection of the colony morphologies of the discussed ClosTron mutants can be seen below (Figure 4.7).

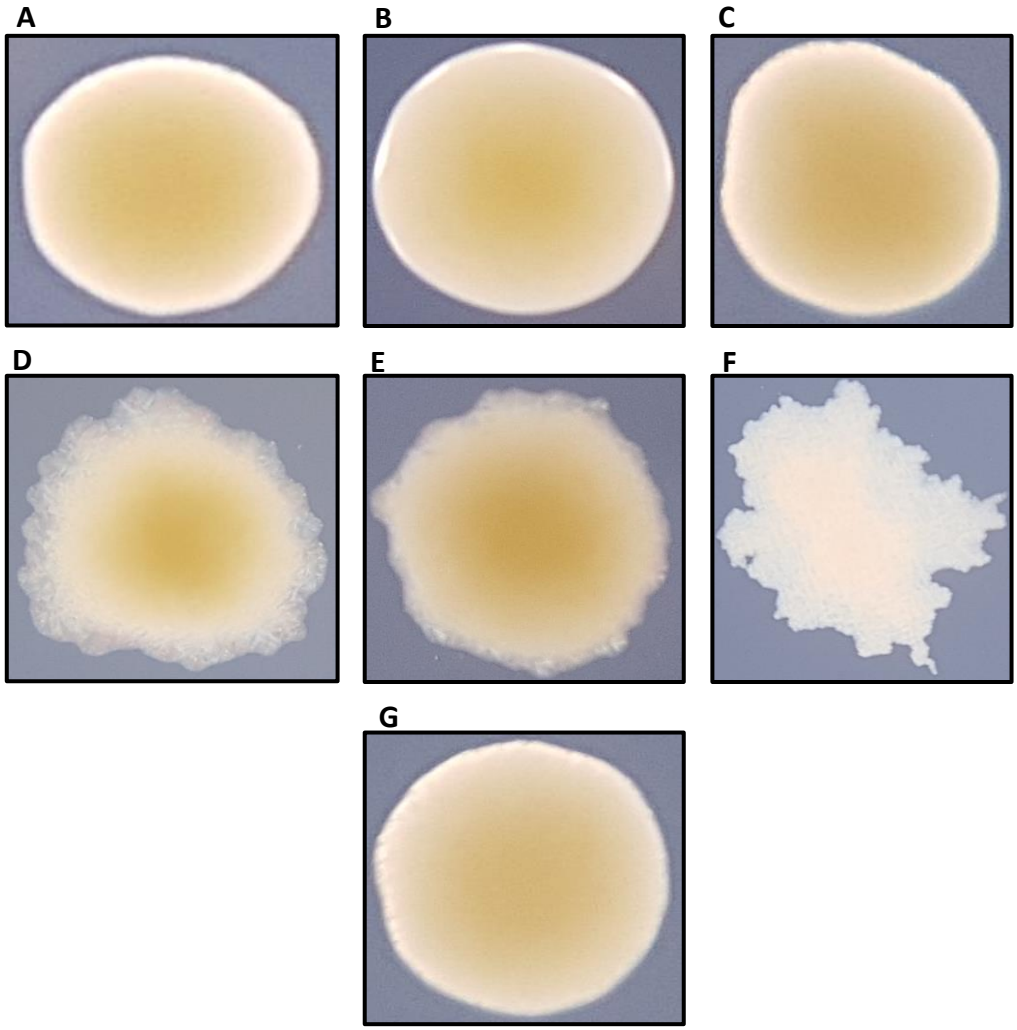


Figure 4.7 Colony morphologies of P5, the *pyrE* repaired ClosTron mutants and their complemented derivatives.

Single colonies of P5 (A), Cbei_0017::CT*ermB* (B), Cbei_0017::CT*ermB*_comp (C), Cbei_3078::CT*ermB* (D), Cbei_3078::CT*ermB*_comp (E), *spo0A*::CT*ermB* (F) and *spo0A*::CT*ermB*_comp (G).

4.4. Discussion

As described in Chapter 3, a total of 71 degenerates were isolated from two descendants of the *C. beijerinckii* NCIMB 8052 clone 3 lineage which were allowed to grow in 5 rounds of subculturing without any mutational bias (Chapter 3). i.e. there were no added mutagens or specific bottlenecks (Foster et al., 2015). The whole genome analysis of all the degenerates against their respective parental wild type revealed a whole array of different mutations found throughout the genome (Table 4.1 and Figure 4.1). Many of the mutations found were in genes not suspected to be related to either solvent formation or sporulation however a clear pattern emerged in which four distinct regions contained many more mutations than other areas of the genome (Figure 4.1). These regions included two histidine kinases, the master regulator gene *spo0A* and a region containing a hypothetical protein and the transcriptional regulator gene *abrB*.

Spo0A is a master regulator of paramount importance for solventogenic *Clostridium* species as it controls the switch to solventogenesis and the formation of spores when concentrations of acids reach toxic levels (Dürre, 2014; Jones and Woods, 1986; Ravagnani et al., 2000). It also acts as a master regulator of sporulation in aerobic endospore forming bacteria, for instance in the genus *Bacillus* (Piggot and Hilbert, 2004; Strauch et al., 1990; Traag et al., 2013). Gene regulation is achieved via binding of phosphorylated Spo0A (Spo0A~P) to specific DNA sequences known as 0A boxes. These sequences are found within the 5' promoter region and allow Spo0A~P to bind, causing gene transcription or repression (Ravagnani et al., 2000; Steiner et al., 2011; Strauch et al., 1990; Wilkinson et al., 1995). In *B. subtilis* the 0A box sequence was found to be TGNCGAA with most 0A boxes TGTCGAA as it gives optimal binding (Baldus et al., 1994; Strauch et al., 1990). 0A boxes in the genus *Clostridium* follow a

similar sequence, however greater variation has been observed for the second and fourth nucleotide. These are often substituted for an A or a T which reflects the AT rich DNA sequence of *Clostridia* (Ravagnani et al., 2000; Wilkinson et al., 1995). Several genes have multiple 0A boxes in the 5' region implying that a certain number of Spo0A~P molecules are required for full gene regulation (Ravagnani et al., 2000). The sequence of the 0A box is highly important for Spo0A regulation as it allows for varying binding affinities. This was seen when several of the nucleotide sequences of the *ptb* 0A box were changed giving varying amounts of Spo0A affinity (Ravagnani et al., 2000). It would thus seem that Spo0A regulation is in part down to how well the protein can bind to specific 0A boxes.

The Spo0A protein contains an N- terminal phosphoacceptor domain linked to the C-terminal effector (DNA binding) domain by a variable connector region (Brown et al., 1994; Escobar et al., 2009; Grimsley et al., 1994; Ravagnani et al., 2000). The putative DNA recognition helix of Spo0A has been found to contain the amino acid sequence SRVERAIRHAIE and is highly conserved amongst *Clostridia* and *Bacillus* (Brown et al., 1994; Ravagnani et al., 2000). Sequence analysis revealed that strain FW7 had mutation in the DNA binding region causing the amino acid sequence to change to SRVERAI(**R**->**I)**HAIE through a guanine to a thymine change at position 1982057. This single base change caused the protein to become completely inactive as the strain produced no solvents or spores (Chapter 3). As arginine is both larger and positively charged compared to isoleucine perhaps the lack of positive charge renders FW7 Spo0A incapable of DNA binding and thus able to regulate gene transcription. Although there is a significant change in the binding region structure, it is interesting that a single amino acid change in the DNA binding region can completely inactivate Spo0A.

Another mutation in the DNA binding region was found for both FW4 and FW5-2 causing the amino sequence to change to SRVERAIR(**H**->**N**)AIE. This mutation saw a positively charged histidine change to a non-charged asparagine but did not render Spo0A completely inactive as both strains could produce solvents, albeit significantly less than the wild type, but did lack the ability to produce spores. It would appear that DNA binding, or at least the affinity to bind, is also residue specific and reduced binding can allow for regulation of some genes but not others.

In *B. subtilis* it has been found that varying levels of Spo0A~P will either activate or repress gene expression depending on the relative concentrations of Spo0A~P (Fujita et al., 2005). This system exists so that gradual increases in Spo0A~P will allow for certain 'low-threshold genes' to be switched on or off before the activation/repression of 'high-threshold genes' to allow for differential gene expression in the cell (Fujita and Losick, 2005). Furthermore, some genes have several 0A boxes linked to them which may exist so that a certain amount of Spo0A~P must be present within the cell before full gene transcription can occur (Lewis et al., 2000). In solventogenic *Clostridia*, the specificity of the DNA binding motif and the 0A boxes may be due to the tight regulation needed for both the switch to solventogenesis and sporulation so that sufficient cellular growth is achieved before cells produce solvents and ultimately metabolically inactive spores. Unlike *C. acetobutylicum* ATCC 824, *C. beijerinckii* NCIMB 8052 produces solvents during the exponential phase of growth. As induction of the solvent genes requires Spo0A~P (Ravagnani et al., 2000), to allow for enough cellular growth the use of 0A boxes with differing strength may be one possible mechanism for differential gene expression to allow for early solventogenesis and late sporulation.

It was interesting to observe several strains who had mutations in *spo0A* outside of the binding motifs which still affected normal function. At the 2 day sample point, DCOG2 produced butanol at a concentration not significantly different to the wild type but produced no detectable heat resistant CFU/ml. Genetic analysis revealed a mutation towards the end of the gene at position 1982142. This mutation was outside of the specific DNA binding motif but was still within the DNA binding domain. Perhaps this mutation allowed for sufficient binding and thus induction of the solvent genes but could not for the genes involved in sporulation. Similar results were seen with CIC4-1 however this strain did produce significantly less butanol than the wild type and more acids. The mutation this isolate had acquired was near the N- terminus of the gene which, as previously mentioned, is the phosphoacceptor domain (Brown et al., 1994; Grimsley et al., 1994). Perhaps mutations in this region meant insufficient phosphorylation of Spo0A thus conferring partial activity giving no spores, reduced solvents and increased acids. Similarly, two isolates, FW2 and FW4-2, had mutations that caused Aspartate (Asp) residues to change, at positions 1981504 and 1981426 respectively. The Asp residues are known to accept the phosphoryl group to phosphorylate Spo0A causing a conformational change in the protein to activate it (Brown et al., 1994; Escobar et al., 2009; Grimsley et al., 1994; Lewis et al., 2000). Perhaps the reduced solvent and spore (Figure 3.8 and 3.9) capacity of these strains is due to a reduced phosphorylation of Spo0A. It would seem that several regions of Spo0A are important for its function and that mutations do not only affect DNA binding.

In *Bacillus* species regulation of Spo0A is through a phosphorelay system that begins with histidine kinases. In this system, Spo0F is first phosphorylated and then Spo0B accepts the phosphoryl group from Spo0F~P at a histidine residue. It is then transferred to Spo0A activating the protein by allowing DNA binding (see Figure 4.8A)

(Dürre, 2014; Perego, 1998; Piggot and Hilbert, 2004; Steiner et al., 2011; Strauch and Hoch, 1993).

In *Clostridium* however there has been no evidence of a phosphorelay system as no *spo0B* or *spo0F* genes have been found (Dürre, 2014; Dürre and Hollergschwandner, 2004; Steiner et al., 2011). In *C. acetobutylicum* ATCC 824 it was found that direct phosphorylation of Spo0A by orphan histidine kinases was responsible for Spo0A activation to initiate sporulation (Steiner et al., 2011).

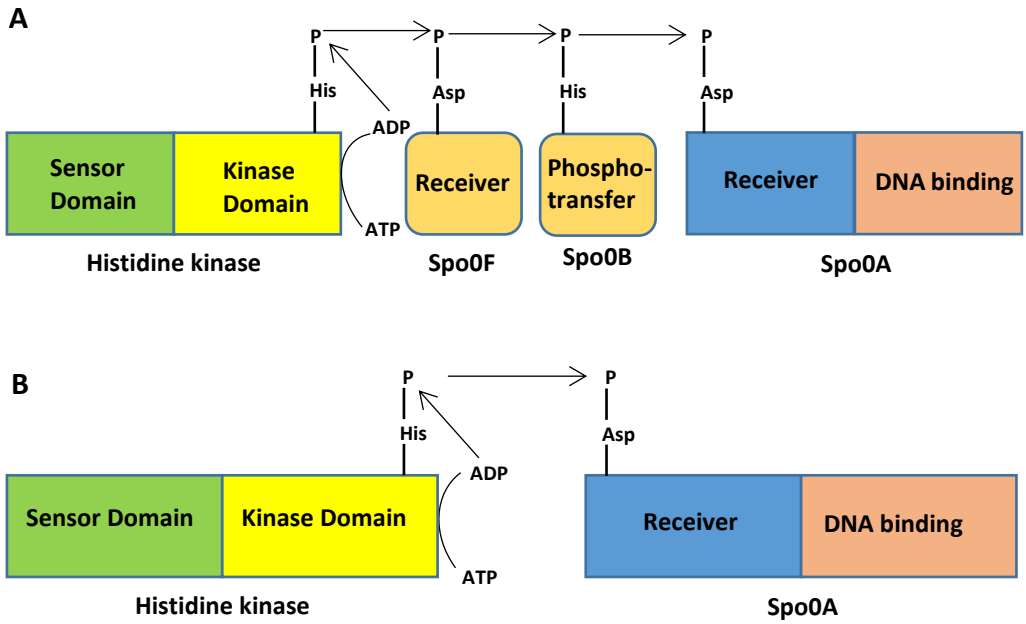


Figure 4.8 Signal transduction for Spo0A phosphorylation in *B. subtilis* and *C. acetobutylicum*.

A summary of the phosphorelay system for Spo0A phosphorylation in the *B. subtilis* model (A) and the direct phosphorylation of Spo0A in the *C. acetobutylicum* model (B). In the *B. subtilis* model, histidine kinases phosphorylate Spo0F which in turn phosphorylates Spo0B which subsequently phosphorylates Spo0A and activates the protein. In *C. acetobutylicum* activation of Spo0A is achieved through direct phosphorylation by orphan histidine kinases. Image adapted from Dürre, 2014.

A similar system has not been identified in *C. beijerinckii* NCIMB 8052 however mutants of the two histidine kinases Cbei_0017 and Cbei_3078 showed a reduced capability to form spores, although not solvents. It was therefore hypothesised that these two kinases may play a role in Spo0A activation via phosphorylation in *C. beijerinckii* NCIMB 8052. Cbei_0017 has been annotated as a two-component sensor histidine kinase and Cbei_3078 as a hybrid sensor histidine kinase/response regulator. Two-component sensor histidine kinases serve to adapt gene expression to a variety of internal and external stimuli and are centred around two proteins, the histidine protein kinase and the response regulator protein (Groisman, 2016; Laub, 2011; West and Stock, 2001). There is a sensor domain of the kinase that regulates the histidine kinase activity. Once a specific stimuli has been detected, the kinase catalyses ATP-dependent autophosphorylation of a specific histidine (His) residue within the histidine kinase dimerisation domain. The phosphorylated kinase domain then serves as a phosphodonor to the regulator domain and the transfer of the phosphoryl group from the phospho-His to the Asp residue on the regulator occurs. The phosphorylation of the regulator subsequently activates the effector domain that causes the response (Groisman, 2016; West and Stock, 2001). In contrast to the classical system is which the kinase and regulator domain are separate proteins, hybrid systems, as annotated for Cbei_3078, has the two domains fused as one polypeptide meaning that the His and Asp residues are found within the single protein (Groisman, 2016; West and Stock, 2001).

Domain analysis using InterPro (www.ebi.ac.uk/interpro) of Cbei_0017 and Cbei_3078 revealed interesting results. For Cbei_0017, within the overarching histidine kinase domain that was annotated, two sub-domains were also annotated as a dimerisation/phosphoacceptor domain and an ATPase domain (Figure 4.2A).

Presumably the first domain is used to accept phosphoryl groups from the sensor and then subsequently the ATPase will transfer the phosphoryl groups to illicit a response.

For Cbei_3078, both these domains were also found within the main histidine kinase domain however three other domains were also annotated (Figure 4.2C). The first two domains were annotated as PAS (Period Clock protein, Ah receptor nuclear translocator protein, Single-minded protein) domains. These domains are known as signal sensor domains of proteins and sense environmental factors such as oxygen and light and are often found in pairs (Taylor and Zhulin, 1999). It would seem fitting that Cbei_3078 contained these domains as it is a hybrid system that should contain all the necessary domains for sensing and kinase activity. What is most interesting is the final domain found after kinase which is annotated as a CheY-like superfamily. The CheY-like superfamily are known response regulators, first characterised in chemotaxis systems as P~CheY was found to interact with the flagella motor (Stock et al., 1989; Volz, 1993; Wolanin et al., 2003). CheY proteins contain the necessary Asp residues to receive phosphoryl groups but are also able to catalyse their own dephosphorlyation to transfer phosphoryl groups to other proteins via CheY-like receiver domains (Parkinson and Kofoid, 1992). Several structural analysis of CheY have shown a remarkable similarity to the Spo0F protein of *B. subtilis* (Madhusudan et al., 1996; Parkinson and Kofoid, 1992). Domain analysis of *spo0A* revealed an N-terminal CheY-like domain (Figure 4.2B). It may well be plausible that the CheY-like domain of Cbei_3078 is able to phosphorylate Spo0A to cause activation of this protein.

Transcriptome analysis of *C. beijerinckii* NCIMB 8052 showed that expression of the Cbei_0017 gene increased to its highest at the end of exponential growth as the cells entered stationary phase (Wang et al., 2012). Expression of Cbei_0017 was continued

into the stationary phase but was not as high as when the cells entered. Expression of Cbei_0017 during the transition to stationary phase implies that the gene is involved in stationary phase phenotypes which are controlled by Spo0A. There was unfortunately no expression data for Cbei_3078.

It has been found that some response regulators have effector domains that are involved in cyclic-di-GMP signalling (Laub, 2011). These domains are known as GGDEF domains and are diguanylate cyclases which catalyse formation of cyclic-di-GMP (Jenal and Malone, 2006). Cyclic-di-GMP signalling has been shown to be involved in the transition of motile to sessile growth, i.e. from exponential to stationary growth, in variety of different bacteria (Jenal and Malone, 2006; Laub, 2011). When looking at the position of Cbei_3078 in the genome it appears to be in an operon with two other hybrid histidine kinases, the TetR-type transcriptional regulator mentioned earlier and two diguanylate cyclases (Figure 4.9).

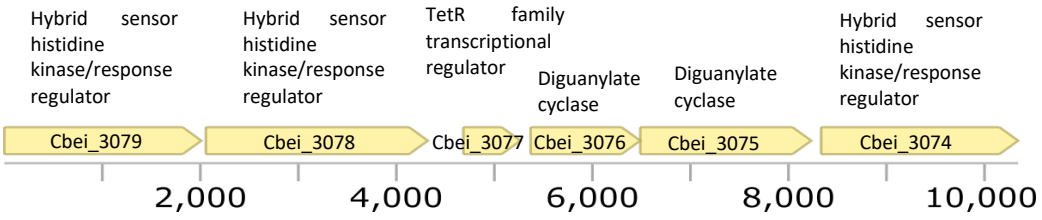


Figure 4.9 Genetic organisation of the Cbei_3078 gene region in *C. beijerinckii* NCIMB 8052.

Shows the gene orientation of the surrounding genes of Cbei_3078 which includes the two diguanylate cyclases. The putative hybrid sensor histidine kinase/response regulator Cbei_3078 appears to be part of a six gene operon also containing the TetR family regulator Cbei_3077, two diguanylate cyclases and two other hybrid sensor histidine kinase/response regulators.

Perhaps this operon serves to sense external stimuli, such as high acid concentrations, causing cellular changes for the onset of solventogenesis and spore formation. It remains unclear whether the role of Cbei_3078 is to phosphorylate Spo0A directly or to aid cyclic-di-GMP signalling in the switch to stationary phase.

Comparison of Cbei_0017 and Cbei_3078 protein sequences with the known orphan histidine kinases of *C. acetobutylicum* that phosphorylate Spo0A revealed little similarity. These orphan histidine kinases in *C. acetobutylicum* are annotated as CA_C0323, CA_C0903, CA_C3319 and CA_C0437. Cbei_0017 similarity to these proteins was 37%, 40%, 26% and 39% and Cbei_3078 was 31%, 29%, 26% and 29% respectively. Although they show low similarity to the *C. acetobutylicum* histidine kinases, the importance of Cbei_0017 and Cbei_3078 in late stage phenotypes was proved by ClosTron mutagenesis.

It is clear from the phenotypic data of the isolated degenerates and from the ClosTron mutants that both of these genes do play a role in spore formation. All of the 3 independent ClosTron mutants of Cbei_0017 and Cbei_3078 produced significantly less spores than the P5 control. It was also possible to restore spore formation in the Cbei_0017 complement strain back to wild type numbers. There was only partial restoration for the Cbei_3078 ClosTron mutant which is believed to be caused by incorrect promoter use. Only the small intergenic region between Cbei_3078 and Cbei_3079 was used in the complementation construct which was far smaller than the usual 200 bp regions used for other constructs. Assuming both genes are part of an operon as shown in Figure 4.9, the native promoter region is most likely located upstream of Cbei_3079. Thus, only RNA polymerase read-through from the *pyrE* promoter would have caused gene expression of Cbei_3078 which was insufficient to restore sporulation completely. Alternatively, or in addition, polar effects of the

ClosTron mutation on the downstream located regulator genes including Cbei_3077 may have been responsible.

Although sporulation was affected by the gene disruption, there was little effect on solvent formation. Both strains produced comparable solvent concentrations to that of the P5 and so evidence suggests these histidine kinases play a role in spore formation and not solventogenesis. Furthermore, isolates that contained mutations in this gene (CIC1, CIC5-1, CIC5-2, RD2-2, CIC1-1, DCOG1-1, DCOG1-2, CIC3-1, CIC3-2, DCOG4-1, RD5-1, DCOG2-2, DCOG5-1, DCOG5-2, RD6-1, CIC4-1, DCOG2-1, CIC2, CIC6-1, CIC6-2 and DCOG4-2) all were still able to produce solvents even when sporulation was reduced. Moreover, several isolates that had mutations in these two genes produced higher concentrations of acids than the wild type (Chapter 3) but this phenotype was not observed for the ClosTron mutants (Figure 4.6). This could be attributed to other mutations that the original isolates had acquired which may affect acidogenesis or that the burden of having a ClosTron insertion meant slightly less growth and thus less acid production. It would be interesting to further study the role of these kinases to the extent of Steiner et al. (2011) through the use of purified protein to observe whether phosphorylation of Spo0A could be achieved *in vitro*.

The region between Cbei_4884 (hypothetical protein) and Cbei_4885 (*abrB* transcriptional regulator) accumulated a significant amount of mutations and clearly plays a role in degeneration. In *B. subtilis* AbrB acts as a repressor of sporulation to prevent premature spore formation until the cells reach the stationary phase (Perego et al., 1988; Strauch et al., 1989). In the *B. subtilis* system, Spo0A~P binds to the 0A box in the *abrB* promoter region and represses the expression of *abrB.* The decrease in expression of the *abrB* allows for activation of the sporulation genes through increased expression of *sigH* and subsequently *spo0A* (Al-Hinai et al., 2015; Perego et

al., 1988; Piggot and Hilbert, 2004; Scotcher et al., 2005; Strauch and Hoch, 1993; Strauch et al., 1989). In *Clostridium* however it seems that the role of AbrB is somewhat different although several homologous do exist. Three *abrB* genes have been identified in *C. acetobutylicum* ATCC 824 (CA_C0310, CA_C1941 and CA_C3647) (Xue et al., 2016). It was found that disruption of CA_C1941 produced no obvious phenotype however both CA_C0310 and CA_C3647 mutants produced significant differences in solvent production. The *abrB3647* mutant produced during that study was able to grow quickly and produce solvents earlier implicating that it has a repressive role on solventogenesis (Xue et al., 2016). Contrasting to this was the *abrB0130* disrupted mutant which produced far less solvents than the wild type suggesting that this AbrB aids solventogenesis (Xue et al., 2016). This data supported findings which used antisense RNA to block the expression of the *abrB0310* gene (CA_C0310) from *C. acetobutylicum* causing a delay in solventogenesis and sporulation which led to the accumulation of acids in the medium (Scotcher et al., 2005). Unlike *B. subtilis* where repression of the *abrB* genes allows sporulation, it is thought that this *abrB* gene plays a more positive role in the transition to the stationary phase (Scotcher et al., 2005).

When comparing *C. beijerinckii* Cbei_4885 *abrB* to that of the *abrB0310* gene from *C. acetobutylicum* there is a striking resemblance. Both genes share 87% nucleotide and 90% encoded protein sequence similarity (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Furthermore, transcriptome analysis of *C. beijerinckii* NCIMB 8052 revealed that expression of Cbei_4885 started just as the cells reached stationary phase and continued thereafter (Wang et al., 2012). Analysis of the Cbei_4885 upstream region undertaken as part of this PhD study revealed a putative 0A box with the sequence TGTCGAA was found upstream of Cbei_4885 indicating that it is controlled by Spo0A. Genomic analysis revealed that isolate FW1-1 contained a mutation in this 0A box

changing its sequence to TGTC**T**AA. Phenotypic analysis of this mutant (Chapter 3) revealed the same profile as the repressed *abrB0310* from *C. acetobutylicum* where solvent production and spore formation were significantly reduced and acids had accumulated. It would therefore be justified to assume that Cbei_4885 *abrB of C. beijerinckii* NCIMB 8052 acts as in a similar way as *C. acetobutylicum abrB0310* in aiding the transition to solvent formation and later on spore formation during the stationary phase.

The role of the hypothetical protein encoded by Cbei_4884 is however harder to establish. Protein sequence BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) revealed over 90% similarity to proteins encoded in many other members of the class *Clostridia* (*C. puniceum, C. saccharoperbutylacetonicum, C. butyricum, C. saccharobutylicum* and *C. botulinum* to name a few) and it seems to be conserved amongst a variety of clostridial species. Domain analysis of this protein (www.ebi.ac.uk/interpro) showed a potential Helix-Turn-Helix (HTH) DNA binding domain of transcription regulators from the MerR superfamily between residues 27-135 (Marchler-Bauer et al., 2017). This family of transcriptional regulators has been found to respond to a variety of external stimuli including metal ions, antibiotics and oxygen radicals (Brown et al., 2003; Chen and He, 2003; Hobman, 2007). Most of the MerR family of transcriptional regulators share a high similarity in the N- terminal DNA binding domain but little in the C- terminal domain which is responsible for stimuli binding and thus activation of DNA binding (Brown et al., 2003). Isolated strain FW3-2 contained SNVs and a deletion at the start of the Cbei_4884 gene and phenotypic analysis showed no solvent or spore production from this strain. Similarly, many other isolates that contained SNVs in the suspected promoter region such as FW5-1 and produced no solvents or spores. Analysing the sequence upstream of this gene revealed no apparent 0A boxes. As these regulators respond mainly to stress, perhaps it has a transcriptional regulatory

function when acids reach toxic levels causing the expression of the solvent and spore genes? Furthermore, as Cbei_4884 is in close proximity to the *abrB* gene previously described (Figure 4.2D), perhaps there is some regulatory expression shared between the two or that they are linked functionally with solvent and spore formation. The true function of this protein remains unknown and would require further phenotypic analysis to determine as well as identification of genes under its control, for instance though RNAseq analyses.

It has been suggested that a defect in a global regulator gene is part responsible for degeneration (Kashket and Cao, 1995; Woolley and Morris, 1990). Through the use of genomic analysis and ClosTron mutagenesis, this global regulator gene is most certainly Spo0A. It would seem that reducing the activity of Spo0A whether through direct mutation, reduced phosphorylation or preventing 0A box binding yields the traits of a degenerate. As Spo0A function is well known in solventogenic *Clostridia* (Dürre, 2014; Harris et al., 2002; Ravagnani et al., 2000) the ClosTron insertions into *spo0A* proved a good control as solvent and spore formation were abolished as expected and previously described (Harris et al., 2002; Heap et al., 2010; Ravagnani et al., 2000; Wilkinson and Young, 1994). Complementation of *spo0A* however was able to restore normal phenotypes to near P5 levels further proving its role. It was also interesting to observe the same colony morphology changes of both natural and artificial *spo0A* mutants forming the flat and white colonies. It has been suggested that dense centres of solventogenic *Clostridium* colonies contain large amounts of spores (Adler and Crow, 1987), something these mutants could not achieve.

Both the phenotypic and genomic data confirm that activity and the regulation of Spo0A is part of a far more extensive system. It would seem obvious that defects in this master regulator gene would be the cause of degeneration phenotypes however

the accumulation of degenerates is not solely based on direct *spo0A* mutation. The phenomenon also includes disruptions in several of the genes responsible for regulating Spo0A expression and activity, and potentially genes which are themselves regulated by Spo0A. It would seem that impairment of this network leads to fitness advantages under certain conditions which allows for degenerate mutants to be selected for in the medium.

For each of the hotspot regions, there appeared to be a link between colony morphology and genotype. Out of the 9 mutations found in Cbei_0017, 7 of those were RD strains (RD1-1, RD1-2, RD4, RD3-2, RD5-1, RD3-2 and RD4-2). The other two mutants were of the CIC morphotype (CIC6 and CIC4) however it would appear that mutants in this gene still retained relatively normal colony appearance. This is further reflected by the colony pictures obtained for the Cbei_0017::CT*ermB* mutant (Figure 4.7B) which looks nearly identical to a wild type colonies.

Out of the 21 unique isolates that had mutations in Cbei_3078, over half (12) were of the DCOG morphotypes. Isolates DCOG4, DCOG3-2, DCOG1-1, DCOG1-2, DCOG4-1, DCOG2-2, DCOG5-1, DCOG5-2, DCOG2-1, DCOG6-2, DCOG5-2 and DCOG6-1 all contained mutations in this gene. Furthermore, 2 out of the 3 isolates with mutations in the potentially linked TetR transcriptional regulator (Cbei_3077) were of the DCOG morphotype (DCOG4 and DCOG3-2) and the only mutant of the also linked Cbei_3074 histidine kinase was a DCOG mutant (DCOG4). This is further exemplified by the colony images seen for Cbei_3078::CT*ermB* which clearly display a dark centred and out growth appearance (Figure 4.7D). The colony morphology Cbei_3078::CT*ermB*_comp did not appear to fully restore colony morphology back to the wild type appearance but seemed less severe than Cbei_3078::CT*ermB*. This is

most likely due to the partial and not full restoration of phenotypes seen from the complement.

12 out of the 19 *spo0A* mutants were of the FW morphotype (FW4-2, FW2, FW3, FW6-1, FW4-1, FW1-2, FW7, FW4, FW5-2, FW3-1, FW5 and FW6). A clean deletion of the *C. beijerinckii* NCIMB 8052 *spo0A* using the CRISPR-Cas9 system produced colonies that appeared snowflake (Wang et al., 2015). Moreover, the colonies produced by the ClosTron insertion into *spo0A* also caused these snowflake colonies (Figure 4.7F). In *Bacillus, spo0A* mutants have been shown to develop flat colonies similar to what has been observed here and previously. This has been attributed to regulatory changes in the chemotaxis and biofilm systems of these organisms which Spo0A is known to contribute (Aguilar et al., 2007; Branda et al., 2001; Hamon and Lazazzera, 2002). Furthermore, *C. acetobutylicum spo0A* mutants showed increased expression of motility and chemotaxis genes (Alsaker et al., 2004; Tomas et al., 2003) which may cause these flat, diffuse colonies to arise due to increased cellular movement and a lack of sporulation. It was interesting to see a full restoration of wild type appearance for the *spo0A*::CT*ermB*_comp strain further supporting the role Spo0A plays on colony morphology. Moreover, the majority of mutations in Cbei_4884, the intergenic region between and Cbei_4885 were of the FW type (6 out of 9 unique isolates). As it was theorised that this *abrB* gene is linked to Spo0A this may well play a part in causing the FW morphotype.

The distribution of mutations for the isolates of the CIC morphotype was far less specific than the other types as many of these isolates had mutations in all the hotspots. The only region hit more by CIC morphotypes was the potential Cbei_3170 gene linked to Agr quorum sensing system as 2 out of 3 isolates for this gene were CIC. It would seem that the CIC morphotype represents a less defined link between

morphotype and genotype and may well embody a morphotype that is less strict to the mutations they gain. This may also be part due to the subjective nature of colony observations so perhaps other morphotypes were picked but were designated as CIC.

C. beijerinckii NCIMB 8052 was chosen due to its tendency to degenerate more rapidly than other strains (Kashket and Cao, 1995; Stephens et al., 1985; Woolley and Morris, 1990). Studies into prokaryotes mutation rates determined that the four dominant sources are either due to intrinsic DNA polymerase errors, endogenously induced DNA damage, DNA damage induced by exogenous agents and the activities of error-prone polymerases (Foster et al., 2015; Lee et al., 2012). Furthermore, loss of DNA repair mechanisms also show mutator phenotypes (Foster et al., 2015). It was interesting that the most common nucleotide changes found from this study were G -> T and C -> A transversions which equalled 49% and 38% of the total SNVs respectively. Studies using *E. coli* have showed mixed results with some reporting bias towards transversions (Akasaka et al., 1992) and others to transitions (Foster et al., 2015; Lee et al., 2012). Mutational studies using an *E. coli* Δ*MutL* (a DNA mismatch repair protein) strain found a significant bias towards transitions as there was a 200-fold increase compared to the wild type (Foster et al., 2015; Lee et al., 2012). Mismatch repair (MMR) systems are incredibly important in maintaining correct DNA however competition between MMR systems has been reported (Kim et al., 2003). Competition between MMR proteins sees either incorrect nucleotides either replaced with the correct bases or they can be maintained and incorporated into the genome. This occurs as each opposing system tries to change the DNA to the nucleotide it perceives as correct (Kim et al., 2003). This competition presumably serves to allow for genetic evolution when certain pressures require strains to adapt. Perhaps clostridial degeneration is a product of these systems to allow strains to adapt and compete in new environments. As it known that degeneration occurs without the

need for external mutagens (Kashket and Cao, 1995) it would be assumed that mutation rates in solventogenic *Clostridia* are due to either ineffective/error prone polymerases or improper DNA repair. Perhaps *C. beijerinckii* NCIMB 8052 has less efficient repair mechanisms than other *Clostridia* which causes random changes in the *spo0A* network which are subsequently selected for in the culture. Losing Spo0A activity clearly gives these degenerates an advantage under certain conditions as these mutants were isolated in large numbers. The concepts of fitness and its role in evolution will be discussed in the next chapter.

5. Understanding the evolutionary forces that shape strain degeneration

5.1. Introduction

The existence of a substantial proportion of mutants within a bacterial population indicates an evolutionary selection pressure that is causing these mutants to be enriched. In their natural habitats, bacteria exist in complex, multi-species communities all competing with each other for resources. In laboratory or fermentation conditions however, nutrients are often not limiting which allows for undesirable mutants to arise within a population that do not conform to the specific biological process they are designed to complete. The relatively new field of sociomicrobiology has emerged which uses concepts of social evolution theory to explain and understand bacterial community behaviour. Using these concepts may therefore prove useful when studying the emergence of degenerate mutants that have arisen in a previously homogeneous population. Social evolution theory usually splits bacterial communities into two populations, one being co-operators and the other cheaters (Cavaliere et al., 2017; West et al., 2006, 2007a). Cooperation is a behaviour that benefits both the self and other individuals in a population. This in itself it is often classified into two broad groups, one being cooperation that gives direct fitness to the performing co-operator and the other where cooperation is altruistic and fitness is gained indirectly via other members in the population (Lehmann and Keller, 2006; West et al., 2006, 2007a). Direct fitness for the cooperating individual is achieved through either a shared interest e.g. if one bacterium survives off the waste by-product of another or through repression of cheating characteristics that would be detrimental to a population (Frank, 2003; Sachs et al., 2004; West et al., 2006, 2007b). Indirect fitness is gained through altruistic behaviour whereby co-operators help close relatives reproduce to pass on their own

genetic material indirectly. This is known as kin selection (Hamilton, 1964; West et al., 2006, 2007a).

Social cheaters however are members of a population that do not conform to cooperative behaviour and will exploit other individuals for their own gain. Cheating is often observed when looking at the production of 'public goods' in a population (West et al., 2006). Public goods are resources produced by an individual that can be used by that individual and its neighbours (Cavaliere et al., 2017; West et al., 2006, 2007a). Production of these public goods is metabolically costly for the co-operators and makes them vulnerable to exploitation by cells (cheaters) that do not produce the public good but still utilise what is produced by others. Several examples of public goods are summarised by West et al. (2007) and include various extracellular enzymes, toxins and the iron sequestering molecules siderophores. It was interesting to note that in the case of siderophores, increased fitness for the cheaters was only observed in iron-rich environments and in not iron-limiting (Griffin et al., 2004) suggesting that different conditions can promote or reduce the emergence of cheaters.

Cheaters often display frequency-dependent fitness which sees their relative fitness change depending on their ratio to the co-operators. Cheaters will do much better when they are rare as there are more co-operators for them to exploit, however, as their numbers increase, the relative fitness of the cheaters will decrease as there is now less co-operators for exploitation (Ross-Gillespie et al., 2007). This decrease in fitness is due to the availability of public goods within the population as the existence of more cheaters to co-operators will thus mean less public goods produced. The productivity of the whole culture will therefore suffer when these cheaters increase (Ross-Gillespie et al., 2007).

Cooperation and conflict in bacterial communities has been studied most extensively in pathogenic species (Darch et al., 2012; Diggle et al., 2007; Griffin et al., 2004; West et al., 2006, 2007a) and has given a solid groundwork to investigate industrially relevant species. With the ever increasing use of bacteria for industrial bioprocesses, it is important to apply social evolution theory to these bacterial communities when optimising stable bioprocesses (Cavaliere et al., 2017). Although starting monocultures will begin homogenous, development of a heterogeneous population is almost inevitable as conditions in bioreactors are never uniform (Avery, 2006; Carlquist et al., 2012; Cavaliere et al., 2017; Enfors et al., 2001; Müller et al., 2010). Even in continuous processes in which conditions are maintained, certain micro-conditions may arise within the reactors which cause differing populations to emerge as each division of the cells will not produce entirely equal offspring (Müller et al., 2010). These changes in the population are often driven by an increased fitness certain strains gain which are subsequently capable of outcompeting cells that have not adapted to new environments in the bioreactor (Avery, 2006; Cavaliere et al., 2017; Müller et al., 2010). The arrival of populations that do not perform as expected is not limited to adaptive mutations but also varying gene expression that can allow for better growth without producing the desired product strains were designed to produce (Avery, 2006; Rugbjerg et al., 2018). Fitness therefore may not only be gained by social cheating but also loss of other properties which represent a burden under the given conditions which are not subject to social microbial interactions.

Although the causes of clostridial strain degeneration have been studied via phenotypic and limited genomic analysis, there has been no investigation into the potential social evolutionary forces driving this phenomenon. At first glance, the accumulation of various degenerate types seems to conform to the concepts of fitness and cheaters. Perhaps the degenerates that emerge are strains that have a gained an

advantage over the wild type? Presumably the FW types show the greatest fitness of all the morphotypes as once they emerge they are able to take over the population (Chapter 3)? As these mutants predominantly produce acids, could these toxic acids be a 'public burden' that the mutants rely on the wild type to reassimilate to prevent culture death? No definitive cause of clostridial strain degeneration has been identified thus far so applying new concepts and techniques may help to explain this phenomenon further.

5.2. Aims of this study

The aim of this study was to employ concepts of sociomicrobiology, such as fitness and cheating behaviour, to give a possible explanation to strain degeneration of solventogenic *Clostridia.* The predominant mutations found from the genetic analysis of Chapter 4 were of *spo0A* and the potential network that regulates this system. The high frequency of mutations accumulated in this regulatory system indicated that there must be an evolutionary driving force allowing for the enrichment and propagation of these mutants once they had emerged at the expense of the wild type. Therefore, it was of interest to explore the fitness of the isolated strains at various growth stages compared to the wild type, starting with one of the *spo0A* deficient degenerates. It was hoped to first understand if this *spo0A* deficient isolate could survive for extended time periods in monocultures or whether it required wild type co-culture for its survival. Following this, mixed cultures of chosen degenerates and wild type would be carried out to observe whether populations did better or worse in different starting ratios. This would serve as an indicator for frequency-dependent fitness and help establish whether degeneration is a social trait that gives

degenerated strains a fitness advantage due to the exploitation of wild type functions. The specific aims were as follows:

- To grow both the wild type and a fully degenerated isolate (FW7 was chosen) in monocultures to observe how well they sustain themselves at different stages of growth
- To identify conditions to help the mutant survive if it could not on its own as this may shed light on underlying physiological defects
- To mix the wild type and FW7 in various starting ratios to observe if one isolate out competes the other to collect evidence for or against cheating behaviour
- To mix different degenerate isolates together in various starting ratios to observe whether one is fitter than the other which might help explain the degenerate pattern observed for cultures during serial transfer (Chapter 1)

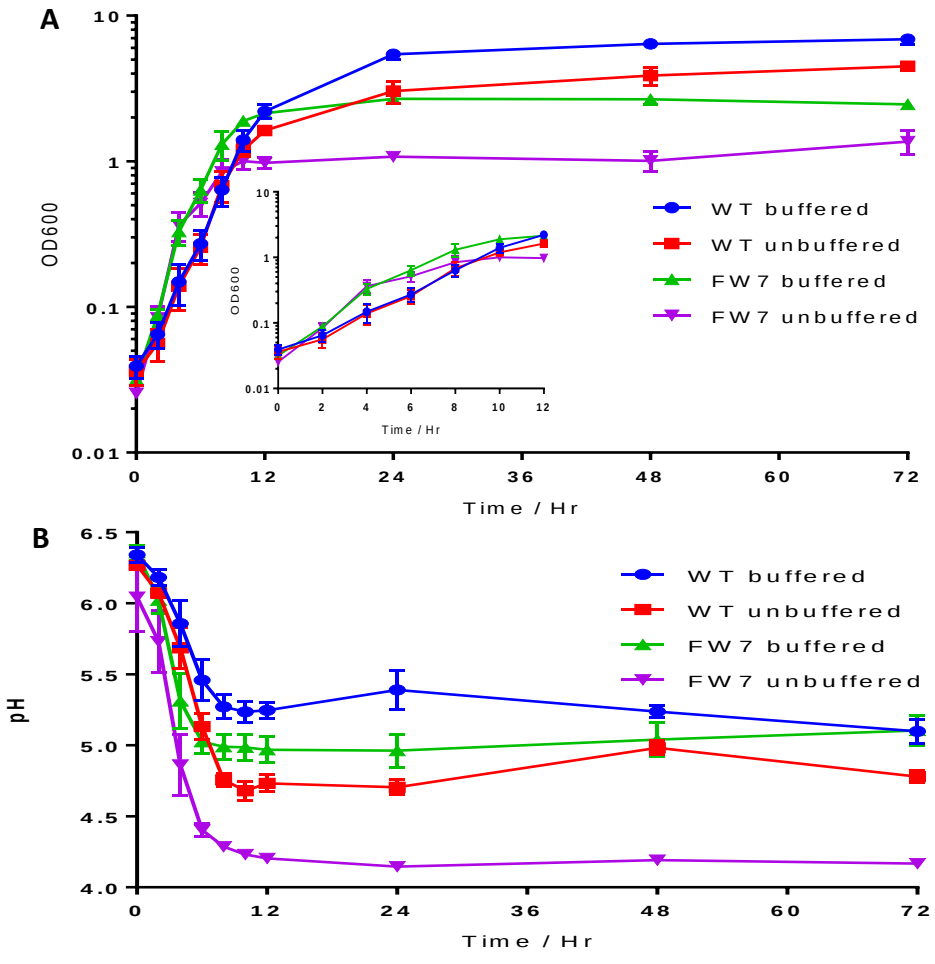
5.3. Results

5.3.1. Monoculture growth of wild type and FW7.

As discussed in Chapter 4, Spo0A is an incredibly important protein which regulates solventogenesis and sporulation (Al-Hinai et al., 2015; Dürre, 2014; Dürre and Hollergschwandner, 2004; Ravagnani et al., 2000). Insertional disruption of *spo0A* (Chapter 4) was able to remove the solvent and spore producing capabilities of the mutated strain. Chapter 4 also eluded to Spo0A and its regulatory network as a key area to be affected by mutations which accumulated in the population. To ascertain whether the loss of *spo0A* gives increased fitness a sufficiently defined *spo0A* mutant

was needed. From the phenotypic analysis of Chapter 3 several FW isolates produced phenotypic traits that conformed to this. Isolate FW7 was selected as it produced no solvents or spores, had a mutation in the crucial DNA binding motif of Spo0A (Ravagnani et al., 2000) and contained only one other SNV that was silent (Table 4.1).

To ascertain whether FW7 could survive on its own in batch cultures or whether it required the wild type for acidic pH protection, both were grown in monocultures at a starting OD_{600} of 0.02 in both 60 ml CBM-S 6% glucose buffered with and without calcium carbonate ($CaCO_3$). Medium without $CaCO_3$ was employed to observe the survival capacity of FW7 in medium with decreased buffering capacity. Optical density, pH and CFU/ml were measured throughout. This can be seen below in Figure 5.1.



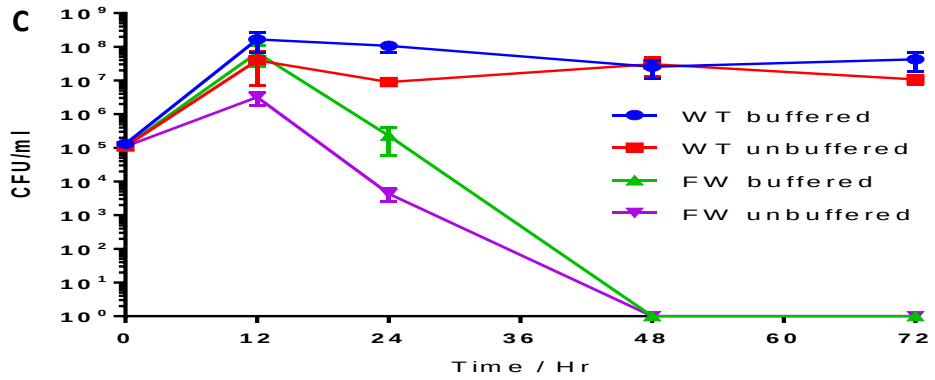


Figure 5.1 Monoculture growth of wild type and FW7 mutant grown in CBM-S 6% glucose with and without $CaC0_3$ buffering.

Growth measured as increase in optical density (OD_{600}) with the first 12 hours of growth shown at higher time resolution in the inset (A). pH (B) and CFU/ml (C) were also measured. No detectable CFU/ml was given a value of 1 (10^0) as 0 cannot be plotted on a log scale. Three independent experiments were carried out for each strain and condition, each consisting of three technical replicas. Data from the technical replicas were averaged for each independent experiment. The resulting averages were used to calculate the standard error of the mean (n = 3) as shown in the graphs.

There were clear growth differences between wild type and FW7 which were further enhanced in the absence of $CaCO_3$. Following inoculation, the FW7 mutant grew faster than the wild type in the first 6-8 hours. After this point growth of the FW7 cultures slowed down for both buffered and unbuffered conditions and entered the stationary phase after approximately 10 to 12 hours. As a result neither culture was able to reach the optical densities finally observed for the wild type. Final ODs not only differed between strains but also for buffered and unbuffered conditions. FW7 was only able to reach maximum ODs of 2.68 ± 0.2 and 1.07 ± 0.1 after 24 hours of growth in

buffered and unbuffered medium, respectively. The wild type however reached 5.42 ± 0.5 and 3.02 ± 0.5 under these conditions.

As to be expected, the pH dropped for all of the cultures in the first acidogenic phase of growth as acids accumulate in the media (Figure 5.1B). FW7 in the unbuffered medium saw the sharpest decline with the pH dropping to 4.14 ± 0.03 at 24 hours. This was far lower than for the other cultures even for the wild type in the unbuffered medium which only dropped to 4.70 ± 0.03. This suggests that the wild type was able to switch to solventogenesis when pH reached toxic levels unlike FW7.

The most striking observation was when comparing the CFU/ml of each of the cultures (Figure 5.1C). After the 12 hour time point, the CFU/ml in buffered and unbuffered FW7 cultures began to decline as cells presumably died. This decline continued and by 48 hours of growth there was no detectable CFU/ml in either the buffered or unbuffered FW7 culture. It was interesting that although the CFU/ml declined, optical density remained relatively constant after 24 hours indicating that optical density is not a good indicator of cell viability. For both conditions, wild type cultures CFU/ml declined marginally over the course of growth which was most likely due to spore formation. None of the wild type cultures saw the sharp decline of the FW7 cultures. It would thus seem that when grown in monoculture FW7 cannot sustain itself after entry into stationary phase even with $CaCO_3$ buffering the medium.

5.3.2. Mixed culture growth of wild type and FW7

Once the survival capacity of FW7 in monocultures had been established, analysis of growth in mixed cultures with the wild type was attempted. A pilot study was undertaken in which the wild type and FW7 were mixed in a 50:50 ratio in both

buffered and unbuffered medium (see appendix). It was found that the mutant out grew the wild type in the first 12 hours in the buffered medium but afterwards, the CFU/ml returned to a similar value for both wild type and FW7 monocultures (see appendix). The CFU/ml for both wild type and FW7 remained similar in unbuffered medium. This provided the hypothesis that the first 24 hours are important for these mutants to take over the population, as seen in the subculturing experiments of Chapter 3. It was therefore decided to modify the setup so that it mimicked the conditions of the previous subculture experiments (Chapter 3) as closely as possible.

As described in Chapter 3, subculturing was done using a 1:10 dilution every 24 hours. This regime only allowed for slightly more than three cell divisions (generations) to occur before cultures entered the stationary phase, but had led to an enrichment of degenerates in the previously performed experiments. Thus, mixed cultures would now be prepared using various volumetric ratios of 24 hour cultures of wild type and FW7. These ratios would be achieved by the addition of cultures to give a final volume of 1 ml which was added to 9 ml CBM-S broth 6% glucose (1:10 as in the previous subculturing experiments). Various starting ratios were carried out and CFU/ml measurements were taken at 0, 12 and 24 hours of growth, the 0 hour measurement thus representing the actual starting ratio. The relative fitness (w) of FW7 to wild type was calculated based on the equation $w = [x_2(1 - x_1)]/[x_1(1 - x_2)]$ from Diggle et al. (2007) and Ross-Gillespie et al. (2007) where x_1 is the initial proportion of mutant in the population and x_2 is the final proportion at a given time point. Values above and below 1 indicate increased and decreased fitness relative to the wild type, respectively. The mixed culture results for this experimental setup can be seen in Figure 5.2.

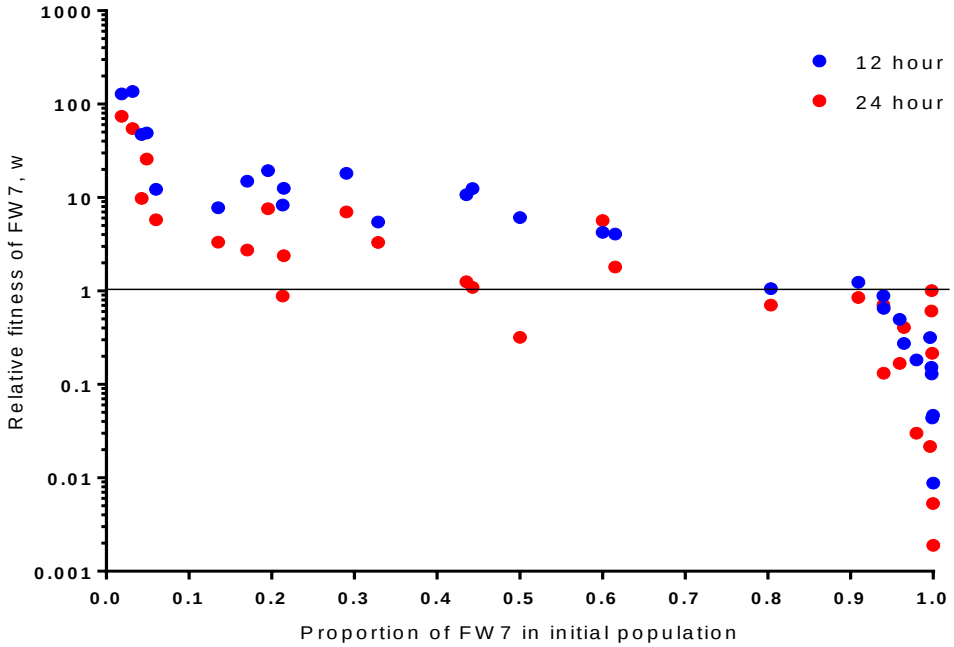


Figure 5.2 Relative fitness of FW7 against the wild type for various starting ratios.

Starting ratios were obtained by mixing 24 hour culture of wild type and FW7 in various volumetric ratios giving a final volume of 1 ml, which was added to 9 ml CBM-S broth 6 % glucose. Starting ratios were determined by plating on CBM-S 6% glucose agar for each individual experiment and plotted against the corresponding relative fitness. Relative fitness was calculated as stated above.

Values above the black line indicate a greater fitness than the wild type. At both 12 and 24 hours the relative fitness of FW7 when the initial starting ratio was low was much higher than that of the wild type. For example when the starting ratio of FW7 was around 1% the relative fitness was around 100. The increased fitness relative to the wild type was however not maintained when the proportion of FW7 increased as fitness values began to decline. When the initial proportion was 10-60% the relative fitness dropped to around 10 and from then on declined even further, reaching a value of 1 at a proportion of 80-90% suggesting that FW7 was only as fit as the wild

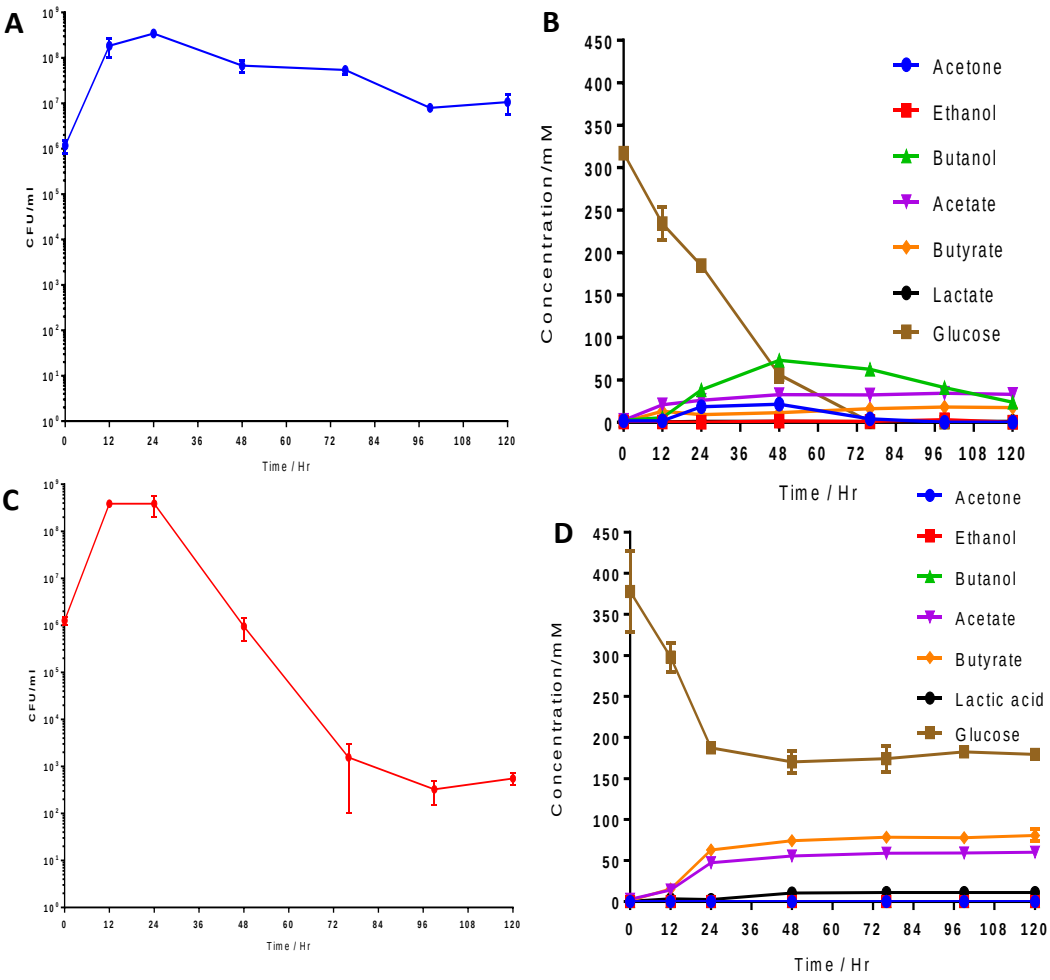
type under these conditions. When the initial proportions were above 90%, there was a further decline in relative fitness with values below 1 suggesting that the isolate was now less fit than the wild type. It was also found that the 12 hour time points gave fitness values higher than that of 24 hours demonstrating that the increased fitness is gained in the early stages of growth. These results are indicative of frequency-dependent fitness which sees cheaters able to do better when they are rare because there are more public goods/co-operators for them to exploit (Ross-Gillespie et al., 2007).

CFU/ml was also determined at 48 hours and as with the monoculture experiments, no detectable FW7 was found in the population. It would seem that FW7 is a social cheat that gains its fitness in the early stages of growth however this increased fitness cannot be sustained into the stationary phase. On the contrary, in contrast to the wild type, viability cannot be maintained after growth has ceased, even when FW7 is grown in mixed culture and low frequency.

5.3.3. Using pH controlled bioreactors to investigate FW7 survival in mono and mixed cultures

The FW7 mutant had been shown to die after 24 hours of growth in both mono- and mixed cultures. This was hypothesised to be a pH-dependent phenomenon, linked to the accumulation of undissociated acids, which the wild type was able to counteract by switching to solvent formation and, potentially, acid reutilisation. To test this hypothesis, the FW7 isolate was grown in a pH 6 controlled Multifor bioreactor (Infors, Switzerland) to determine whether this would maintain culture viability. This pH was chosen as it was considered non-toxic and would keep a vast proportion of the produced acids in their dissociated state. The effect of pH 6 control was tested for

both mono- and mixed culture fermenters. Bioreactors containing 300 ml CBM-S broth 6% glucose were ran in duplicate with either wild type, FW7 isolate or a mixed culture. 24 hour 30 ml pre-cultures were inoculated into 270 ml CBM-S broth 6% in a 1:10 dilution. One of the mixed culture duplicates was inoculated with a low volumetric ratio of mutant (1:30) and the other with a high ratio of mutant (30:1) to simulate the two ends of the fitness spectrum. CFU/ml, fermentation products, including lactic acid, and glucose concentration were analysed in samples removed from the bioreactors in regular intervals. Glucose content of the samples was analysed to determine if the lack of available carbon source was a cause of cell death. The results can be seen in Figure 5.3.



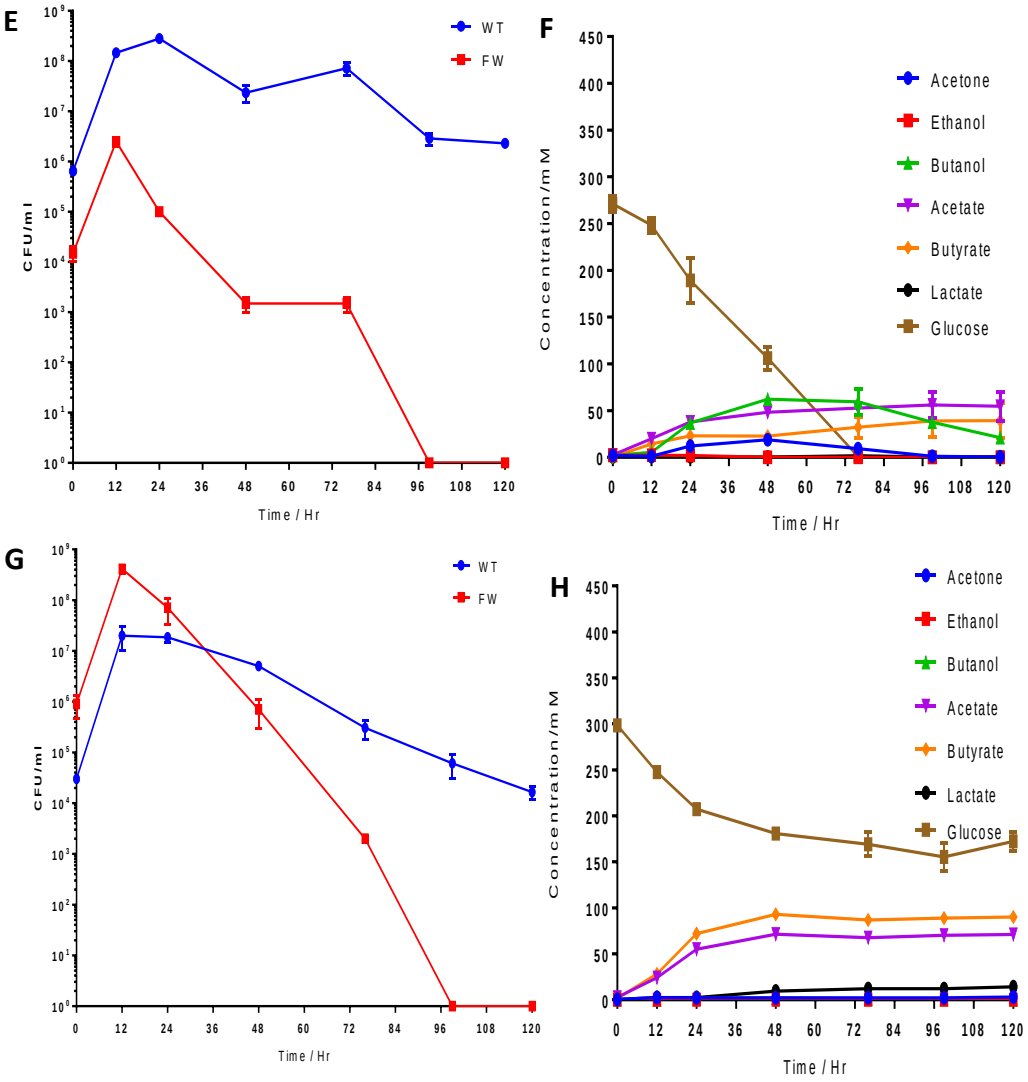


Figure 5.3 Wild type and FW7 grown in mono- and mixed cultures using pH 6 controlled bioreactor in batch fermentation mode.

CFU/ml (A) and fermentation products (B) for wild type monoculture, CFU/ml (C) and fermentation products (D) for FW7 monoculture, CFU/ml (E) and fermentation products (F) for mixed culture 1 and CFU/ml (G) and fermentation products (H) for mixed culture 2. Cultures were grown in 300 ml CBM-S broth 6% glucose and the pH maintained by automated titration with 3.4 M KOH. No detectable CFU/ml was given a value of 1 (10^0) as 0 cannot be plotted on a log scale. Error bars indicate the standard error of the mean; n = *2* for 2 separate bioreactors.

With pH 6 maintenance, the monocultures of FW7 were able to survive past 48 hours as CFUs were still detectable until the end of the fermentation (5 days). In the first 12 hours of growth FW7 was able to reach similarly high CFU/ml to the wild type suggesting that it is only the toxic low pH that would usually limit maximum FW7 growth in ordinary batch cultures. Although a low number of cells still survived until to the end of the fermentation, the CFU/ml of FW7 did decrease considerably throughout the course of the experiment indicating that it was not solely pH that was causing cell death. Unlike the wild type which depleted virtually all of the glucose, complete utilisation of glucose was not seen by FW7 as the concentration of glucose remained at around 175 mM after 24 hours. This suggested that the culture had stopped metabolising after 24 hours and that lack of glucose was not a factor in cell death.

It was interesting to see in the mixed culture fermenters that both pH 6 maintenance and the presence of the wild type did not prolong survival of FW7. Two duplicate mixed culture fermenters were inoculated with cultures consisting predominately of wild type or FW7, with starting ratios of 30:1 and 1:30 (WT:FW7), respectively. Both mixed culture duplicate fermenters saw the highest CFU/ml of FW7 at 12 hours, with now dramatic changes in the WT:FW7 ratios. The CFU/ml subsequently decreased for both strains until at the 4 day sampling point (99 hours) there was no detectable FW7 CFU/ml in either experiments. For these mixed cultures, fermentation product profiles seemed to mirror the metabolic behaviour of the dominating strain. The fermenters with a low wild type inoculum saw the wild type dominating only from 48 hours onwards (Figure 5.3G), due to a much more rapid decline in viable FW7 cells. However, even with decreasing numbers of FW7 (Figure 5.3G) there was no restoration of the fermenter productivity in terms of solvents by the now dominating wild type cells as the fermentation products contained mainly organic acids and the

wild type CFU/ml also decreased over time. For the fermenters with high wild type inoculum (Figure 5.3E) wild type CFU/ml were always much higher than those for FW7, explaining the overall fermentation product profile.

5.3.4. Confirming the role of Spo0A in fitness using the constructed *spo0A*::CT*ermB* mutant strain

To confirm that loss of Spo0A functionality provides a fitness benefit, one of the *spo0A* ClosTron mutants described in Chapter 4 was selected and grown together with the wild type in the same mixed culture setup as described previously for FW7 (5.3.2). Proportions of wild type and mutant were determined at 0, 12 and 24 h and used to calculate relative fitness. The results of this experiment can be seen below in Figure 5.4.

In terms of frequency-dependent relative fitness, a similar trend was seen for the *spo0A::*CT*ermB* strain when compared to that obtained previously for FW7 (Figure 5.2). The *spo0A::*CT*ermB* strain was fitter than the wild type throughout the increasing starting ratios with the relative fitness only being less than 1 when the starting population of *spo0A::*CT*ermB* was around 90%. This further confirms the considerable fitness benefit conferred by Spo0A functionality.

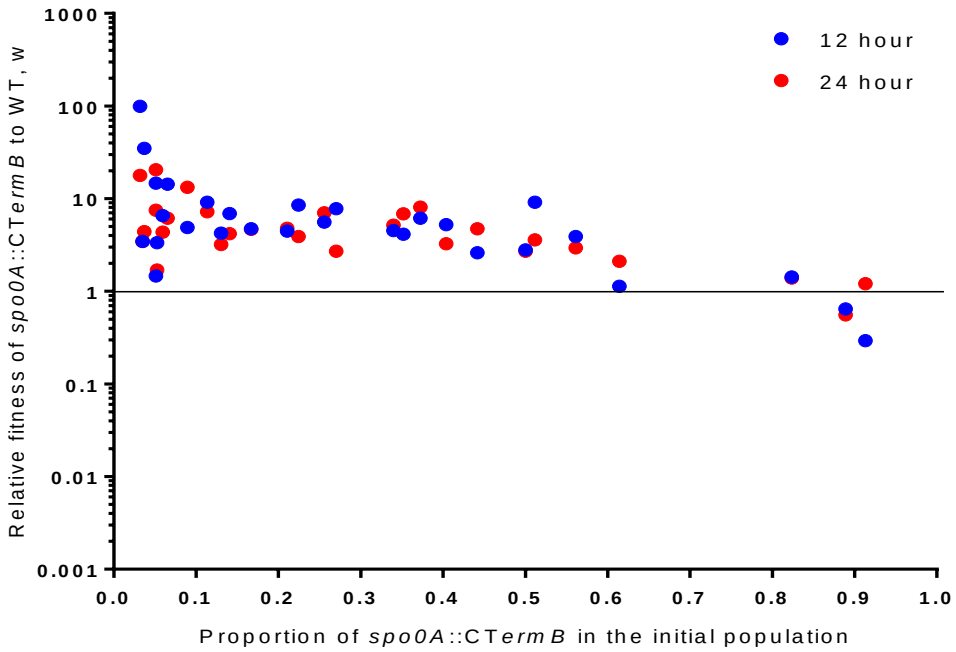


Figure 5.4 Relative fitness of *spo0A*::CT*ermB* against the wild type for various starting ratios.

Starting ratios were obtained by mixing 24 hour culture of wild type and *spo0A*::*CTermB* in various volumetric ratios giving a final volume of 1 ml, which was added to 9 ml CBM-S broth 6 % glucose. Starting ratios were determined by plating on CBM-S 6% glucose agar for each individual experiment and plotted against the corresponding relative fitness. Relative fitness was calculated as stated above.

5.3.5. Fitness benefit conferred by reduced Spo0A activity

The Spo0A protein produced by the FW7 isolate was unable to activate solventogenesis or sporulation due to the mutation in its DNA binding domain. It was interesting to investigate whether other mutations in this gene, potentially leading to only partial inactivation of Spo0A, would also give a fitness benefit to the mutation carrying strain. Isolate CIC4-1 was selected as it contained a SNV in *spo0A* at position

1981493 towards the start of the gene. From protein blast analysis this region is believed to be the signal receiver domain which would be phosphorylated by histidine kinases for Spo0A activation (Solá et al., 1999; Steiner et al., 2011). This isolate produced no detectable spores but did produce some butanol (Chapter 3) making it a good candidate to demonstrate partial Spo0A activity. As with FW7 and *spo0A*::CT*ermB* it was grown alongside the wild type using the same procedure as described before to determine relative fitness for different starting frequencies (5.3.2). It was initially hard to effectively and reliably distinguish CIC4-1 colonies from the wild type as they looked too similar. The use of granulose staining was therefore employed as can be seen in Figure 5.5A, B and C which allowed for colonies to be reliably distinguished. The results of the mixed culture analysis can be seen in Figure 5.5D.

Granulose staining proved effective in distinguishing CIC4-1 from the wild type as colonies only partially stained with the addition of iodine solution. Similar results for the previous mutants can also be observed for CIC4-1. The relative fitness of CIC4-1 was high when the starting ratio was low however this fitness decreased as the proportion of CIC4-1 increased. Although the same frequency-dependent fitness as FW7 was seen, CIC4-1 did not reach a relative fitness that was as high as that of FW7 in the very low initial proportions. In contrast to FW7 however, the 24 hour fitness for CIC4-1 were generally higher than that of the 12 hour suggesting that CIC4-1 maintains fitness for longer than FW7. Furthermore, CIC4-1 was able to extend a fitness higher than 1 for longer than FW7 when the initial proportions were increased as it was only when CIC4-1 occupied 90% of the culture did the values drop below 1. It is clear however that overall both isolates behaved similar suggesting that complete or partial loss of Spo0A gives the traits of a social cheat. It should also be noted that CFU/ml of the CIC4-1 decreased after the 24 hour time point as seen for FW7 (data

not shown) but was slightly less severe than seen from FW7 as at 72 hours there was a small amount of CIC4-1 remaining in some of the higher proportion starting ratios.

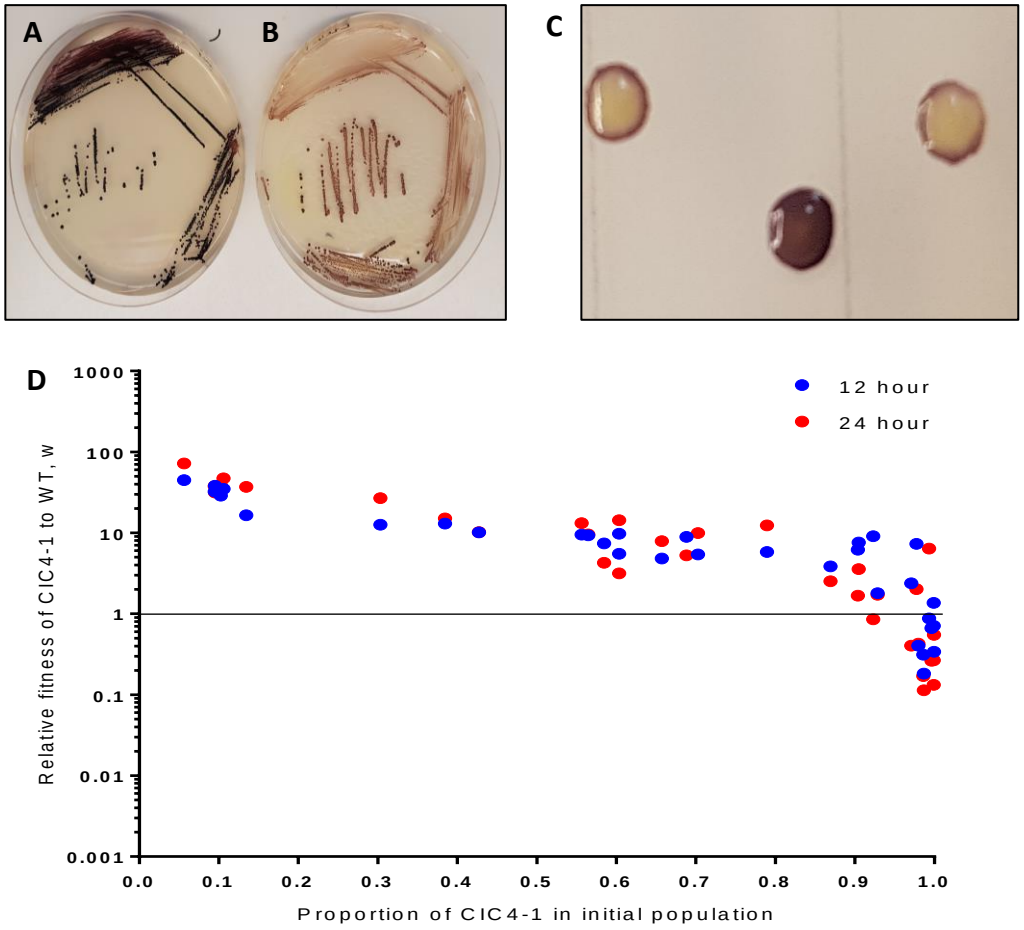


Figure 5.5 Relative fitness of CIC4-1 against the wild type for various starting ratios.

Granulose staining differentiated plate-grown wild type (A) and CIC4-1 (B) and was used to discern individual colonies after plating of mixed cultures (C). Relative fitness of CIC4-1 against the wild type was plotted for various starting ratios (D). Starting ratios were obtained by mixing 24 hour culture of wild type and *spo0A*::*CTermB* in various volumetric ratios giving a final volume of 1 ml, which was added to 9 ml CBM-S broth 6 % glucose. Starting ratios were determined by plating on CBM-S 6% glucose agar for each individual experiment and plotted against the corresponding relative fitness. Relative fitness was calculated as stated above.

5.3.6. Mixed culture analysis of FW7 vs CIC4-1

Both FW7 and CIC4-1 showed higher fitness over a large range of starting frequencies when grown alongside the wild type (Figure 5.2 and Figure 5.5) providing an obvious explanation as to why these strains emerged in the population. It was now interesting to investigate whether having a partial or fully inactivated Spo0A gave the highest fitness. Mixed culture analysis of FW7 versus CIC4-1 was carried out using the same protocol as before (5.3.2). The results can be seen in Figure 5.6.

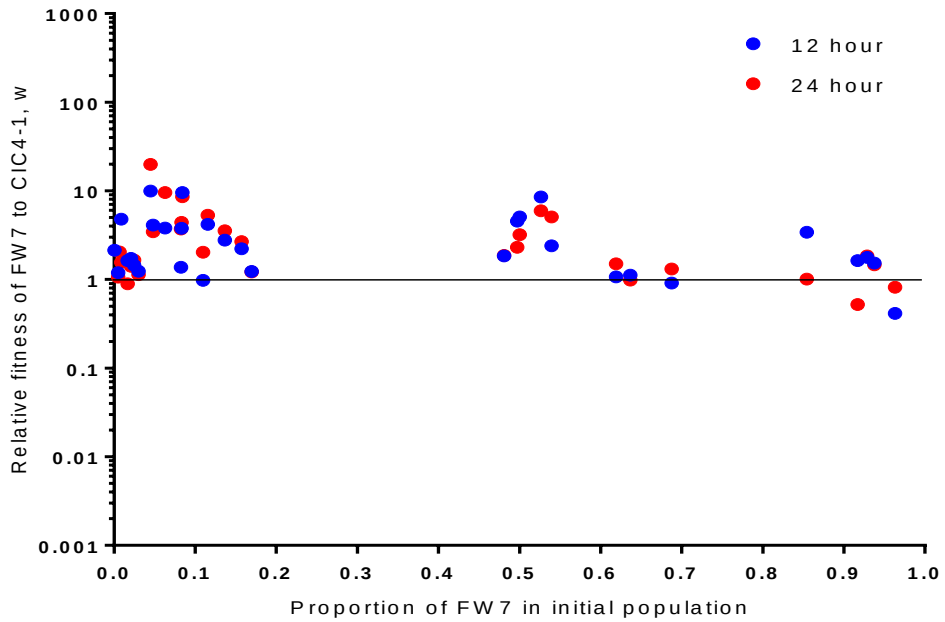


Figure 5.6 Relative fitness of FW7 against CIC4-1 for various starting ratios.

Starting ratios were obtained by mixing 24 hour culture of FW7 and CIC4-1 in various volumetric ratios giving a final volume of 1 ml, which was added to 9 ml CBM-S broth 6 % glucose. Starting ratios were determined by plating on CBM-S 6% glucose agar for each individual experiment and plotted against the corresponding relative fitness. Relative fitness was calculated as stated above.

When FW7 and CIC4-1 were grown alongside each other, the relative fitness of FW7 was generally higher than that of CIC4-1 for lower starting frequencies up to a ratio of 50:50. The increased fitness of FW7 over CIC4-1 at lower starting frequencies was however not as large as seen when grown against the wild type as overall the relative fitness values were between 1 and 10. This suggest that complete loss of Spo0A activity gives the greatest fitness advantage however this advantage is only marginal at higher frequencies when grown against a strain carrying partially active Spo0A.

5.3.7. Mixed culture of FW7 vs DCOG1-1

As the hybrid sensor histidine kinase/response regulator gene Cbei_3078 contained the highest number of mutations (Chapter 4) it was decided to establish the fitness of a mutant of this gene. Cbei_3078 is also believed to play a role in *spo0A* regulation or activation of the encoded protein (Chapter 4). DCOG1-1 was selected as the isolate only contained mutations in Cbei_3078. Furthermore as this isolate was a DCOG type, the second most abundant morphotype (after FW), it seem appropriate to understand this types relative fitness. This mutant was still able to produce solvents, albeit slightly less than the wild type at 6% less after 2 days (Figure 3.6A), and still produced spores however this was significantly less than the wild type at 200 % less (Figure 3.9B). Attempts were made to grow DCOG1-1 against the wild type however, consistent analysis of strain frequency was not possible as colonies looked too similar to be reliably distinguished. Granulose staining was attempted but again did not provide sufficient distinction of wild type and DCOG1-1 colonies. It was therefore decided to grow this strain together with FW7, already confirmed to be fitter than the wild type,

to observe the relationship between the two most abundant morphotypes. Results for the mixed cultures with FW7 can be seen in Figure 5.7.

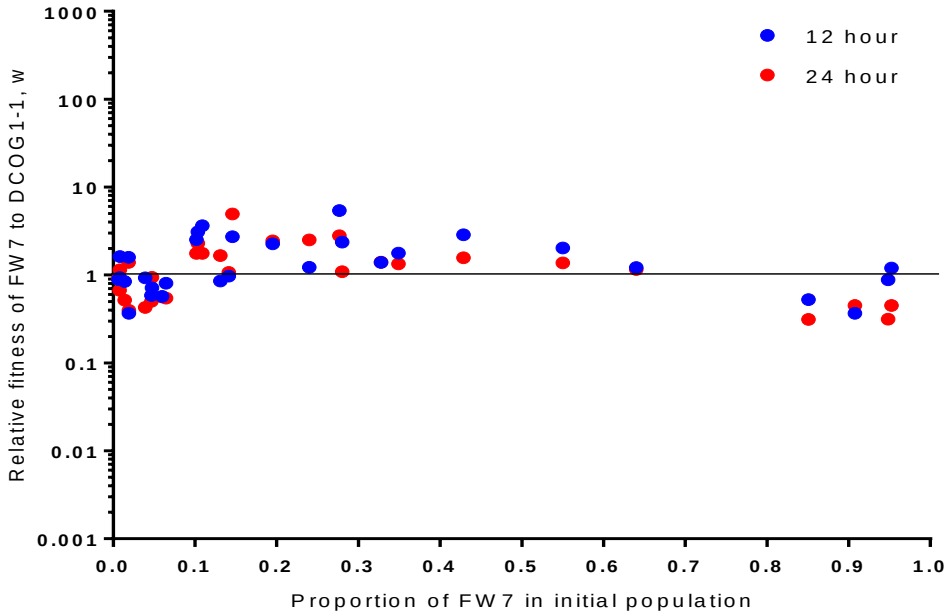


Figure 5.7 Relative fitness of FW7 against DCOG1-1 for various starting ratios.

Starting ratios were obtained by mixing 24 hour culture of FW7 and DCOG1-1 in various volumetric ratios giving a final volume of 1 ml, which was added to 9 ml CBM-S broth 6 % glucose. Starting ratios were determined by plating on CBM-S 6% glucose agar for each individual experiment and plotted against the corresponding relative fitness. Relative fitness was calculated as stated above.

A slightly different pattern emerged from this mixed culture experiment as at low initial frequencies FW7 appeared to have no fitness advantage unlike in other experiments. Values in which FW7 were fitter than DCOG1-1 were only seen when the initial proportion of FW7 was between 10-70% of the population indicating that a substantial proportion of FW7 is required before the isolate gains an advantage. Unlike the other mixed culture experiments, relative fitness values of FW7 to DCOG1-

1 were always less than 10. When FW7 in the initial proportion was increased past 70%, the relative fitness decreased as seen in the other experiments.

5.3.8. 12 and 72 hour serial transfers

As it had been found that the FW7 isolate gained the greatest fitness in the first 12 hours of growth and that after 48 hours there was little to no detectable CFU/ml of this type in non pH-controlled batch cultures, it was decided to carry out serial subculture experiments (as described in Chapter 3) but change the transfer time to 12 hours and 72 hours (3 days), respectively. This was to observe whether the culture degenerated more rapidly in 12 hour transfers, assuming a faster accumulation of FW types, and whether 72 hour transfer would limit the rate of accumulation of degenerates as presumably there would be little or no viable FW in the population. As before, this experiment was still carried out in 10 ml CBM-S 6% glucose and began with the addition of a single colony at subculture 0 to six separate tubes which was serially transferred 1/10 dilution for 10 subcultures. For each of the 10 subcultures, frequencies of colony types were determined 1 day after transfer though plating and subsequent CFU counts. Fermentation products were analysed 2 days post inoculation and heat resistant CFU/ml determined 5 days after inoculation. Whole culture samples were also taken for genomic extraction at each subculture. This was to provide information on the genetic changes throughout subculturing by deep sequencing, but due to time constraints was not analysed in this work. The results of both the 12 hour and 72 hour transfer can be seen below in Figure 5.8.

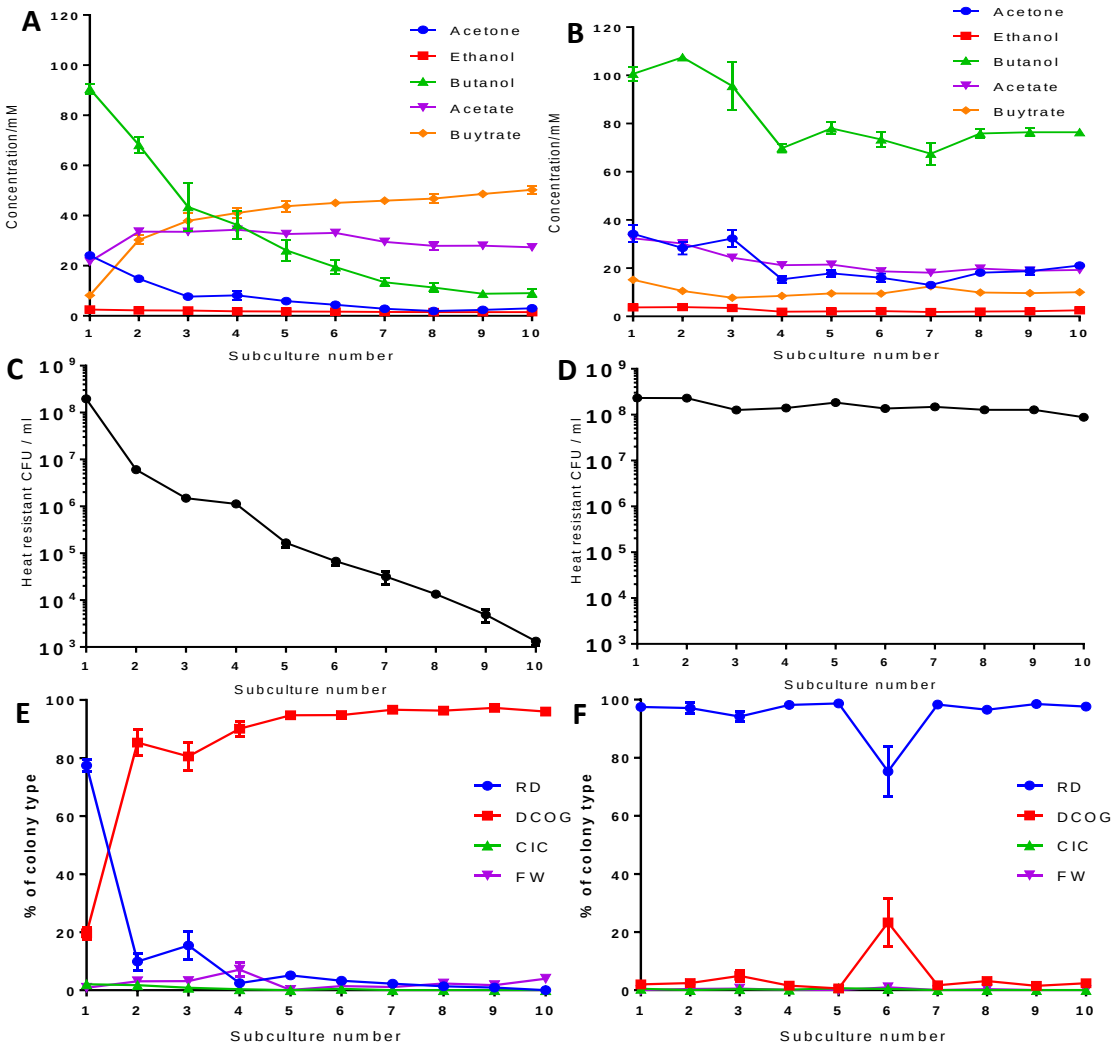


Figure 5.8 Dynamic changes of culture degeneration for 12 hour and 72 hour 1:10 serial transfers with respect to fermentation products (A+B), spore counts (C+D) and percentage of observed colony types (E+F).

Replica cultures were grown in 10 ml CBM-S 6% glucose and 1 ml removed for serial transfer into 9 ml of fresh medium after 12 hours and 72 hours, respectively. Individual cultures were incubated for 5 days in total and samples taken after 48 h for fermentation products, (A) 12 hour transfer and (B) 72 hour transfer; heat resistant CFU/ml after 5 days, (C) 12 hour transfer and (D) 72 hour transfer; percentage of colony types found from samples taken after 24 hours, (E) 12 hour transfer and (F) 72 hour transfer. Biological replicates represent standard error of the mean; n = 6.

A remarkable contrast was seen between the differing transfer times. As predicted, the 12 hour transfer saw the rapid loss of butanol production and the accumulation of acids at each successive subculture (Figure 5.8A). Butanol concentrations dropped from 90.5 ± 2.1 to 9.1 ± 1.6 mM representing a 10 fold decrease after 10 subcultures. Acetate concentrations grew slightly after the first transfer but remained mostly constant which has been seen previously (Chapter 3). Butyrate concentrations grew 513% from 8.2 ± 0.2 to 50.3 ± 1.6 mM by the end of subculturing which is very similar to what is seen from the 24 hour transfers (Chapter 3). By contrast, for the 72 hour transfer butanol concentrations began at 100.6 ± 3.1 mM and only dropped down to 69.8 ± 1.8 mM by the fourth subculture and from then on remained around this concentration (Figure 5.8B). Similarly acetone concentrations dropped from 34.3 ± 3.6 mM down to 15.4 ± 1.5 mM by the fourth subculture and then remained around this concentration. Acetate concentrations dropped from 32 mM to around 20 mM and butyrate concentrations remained around 10 mM.

This difference was further exemplified when observing heat resistant CFU/ml as spore numbers drastically dropped for the 12 hour transfer throughout the experiment but only marginally for the 72 hour transfer (Figure 5.8C+D). Both regimes started at around 10^8 heat resistant CFU/ml for the 12 hour transfer heat resistant CFU/ml dropping down to just above 10^3 after 10 transfers, representing a decrease by five orders of magnitude. For the 72 hour transfer only dropped to just below 10^8 after 10 transfers, representing a 2fold decrease, thus constituting a huge difference between the two transfer times.

The most interesting observations were the colony types that emerged from the transfers. The 12 hour transfer saw the complete takeover of the population by colonies that looked like the DCOG morphotype. This type occupied just 19.6 ± 2.2%

of the culture after 1 subculture eventually increasing to 96.0% ± 1.3% after 10 transfers. It should be mentioned that a potential intermediate type was seen alongside the classical DCOG colony morphology. This new type looked more light centred than dark and was included in the DCOG category but may skew actual DCOG numbers. This new type might represent an intermediate type between the DCOG and FW reflected by the loss of butanol and increased acids which was not seen by sole DCOG types. Stocks of these types were taken for future analysis. Only a low proportion of FW types were observed which was unexpected as it was believed that this type would take over even more rapidly based on the 24 transfer results. Similarly, few CIC were seen also.

The 72 hour transfer however saw a very different colony profile. The most dominant was the RD type which began at 97.5 ± 1.0% of the subculture 1 population and stayed around this percentage until the end. There was an increase of DCOG at subculture 6 where it reached 23.3 ± 8.4% of the population but this was not sustained. Interestingly at this subculture, two out of the six replica tubes only contained around 3% DCOG compared to the other four replicas which ranged from 20-50% which is why the error is large. The maintenance of the RD type (i.e. colonies that look most like the wild type) is clearly reflected by the relatively stable production of solvents and spores.

5.4. Discussion

Chapter 4 eluded to the role of *spo0A* in degeneration and how loss of protein functionality, and the networks that control it, unsurprisingly causes the phenotypes of degeneration. To investigate if losing Spo0A activity gives a quantifiable fitness

benefit, isolate FW7 was chosen as the archetypal *spo0A* mutant as it produced no solvents or spores and had acquired a mutation in the DNA binding domain of Spo0A (Ravagnani et al., 2000) (and, in contrast to other isolates, only one other SNV which was silent) which had rendered Spo0A inactive. It was first necessary to investigate whether this mutant could sustain itself on its own or whether it required the wild type to cheat from to maintain survival. It was found from the monoculture growth analysis that FW7 was unable to survive past 48 hours of growth even in the presence of calcium carbonate buffering. The loss of cell viability was most likely due to the accumulation of acids in the medium which would have caused the pH to reach toxic levels. Upon approaching the pKa of acetic and butyric acid (4.76 and 4.82, respectively (Fond et al., 1985)) when the pH drops towards these values, an increasing proportion of the previously produced acids would have been present in their undissociated form. High levels of undissociated acids in the medium may disrupt the proton gradient of across the cell membrane as in this form, they are able to cross membrane, dissociate to release a proton and thus have effectively brought additional protons into the cell (Jones and Woods, 1986). Disruption of this essential gradient can lead to cell death via several factors including loss of NDP/NTP ratios which can affect both DNA/RNA synthesis and inhibit metabolic functions such as loss of carbohydrate and ammonia uptake. Furthermore, elevated levels of acids can cause the CoA and phosphate pools to decrease (Gottwald and Gottsehalk, 1985; Herrero, 1983; Herrero et al., 1985; Jones and Woods, 1986).

Lack of cell viability after 48 hours in monoculture suggested that these mutants may require external support to survive. The concept of 'public goods' in a population was previously defined as resources produced by an individual in a population that can be utilised by self and neighbours (Cavaliere et al., 2017; West et al., 2006, 2007a). It was therefore hypothesised that organic acids produced in the acidogenic phase could be

considered a public good, or more likely a public burden, that FW7 exploited. In this case, degenerates could constantly produce acids to gain the most amount of energy generation (3.25 ATP/1 molecule of glucose for acid production compared to 2 ATP/1 molecule of glucose for solvent production in *C. acetobutylicum* (Jones and Woods, 1986)) and any wild type cells in the population would assimilate these acids to keep the pH at non-toxic levels. The investment of co-operating individuals in enzyme and activities required for solvent formation may constitute a public good compliant with provided definitions.

This hypothesis was tested by maintaining the pH at 6 in a pH controlled fermenter to observe whether FW7 viability could be maintained past the 48 hour point. It was found that although cell viability was prolonged until the end of the fermentation (5 days), much longer than previously seen, the mutant still died as CFU/ml slowly decreased after 24 hours of growth. By the end of the fermentation, there were only 0.0001% FW7 that had survived compared to maximum CFU/ml at 12 hours. More importantly, when grown with the wild type in mixed cultures kept at pH 6, after the 48 hour time point there were no detectable viable FW7 present. This suggested that the FW7 mutant was unable to exploit, or benefit from, the wild type cells present in the co-culture. It would appear the rapid acidification of the medium in pH uncontrolled fermentations during the exponential phase may be key in giving FW7 an advantage in the early stages of growth.

It has been found that the addition of acetate and butyrate to the fermentation medium causes increased solvent production in *C. beijerinckii* and *C. acetobutylicum* respectively (Bahl et al., 1982; Chen and Blaschek, 1999a). This is most likely due to the elevated levels of acids causing rapid induction of the solventogenic genes within these organisms. Perhaps the rapid production of acids by FW7 causes the solvent

genes of the wild type to be induced and thus enforcing a less efficient metabolism? Although a plausible reason for FW7 initial take over, wild type *C. beijerinckii* NCIMB 8052 has been shown to make solvents during the exponential phase (Figure 3.1) so although this may play a contributing factor, it is not the sole reason for FW7 take over. Lower growth rates of the wild type in the early stages may also reflect the burden associated with accumulation of the storage compound granulose (Reysenbach et al., 1986), which *spo0A* negative strains do not do, which may then be used as an internal carbon and energy source. Perhaps the lack of granulose accumulation also contributes to FW7 increased growth.

Although pH is most definitely important in cell survival, it is not the only cause of cell death for this mutant and other factors must therefore be considered. When observing glucose concentrations throughout the pH 6 controlled fermentations, FW7 effectively stops sugar consumption after 24 hours. Perhaps continued production of acids at some point leads to a metabolic slow down and finally standstill as solvent cannot be produced. This could therefore lead to an end of maintenance metabolism and finally cell death. The social aspect of FW7 exploitation may therefore only exist in the early stages of growth and not later on when for some reason metabolism ceases.

It has been found that the industrially relevant species *C. saccharobutylicum* P262 (formerly named *C. acetobutylicum* P262) produces a bacteriocin that only affects itself and the closely related species *C. felsineum* (Barber et al., 1979). The bacteriocin was found to be produced at the end of the exponential phase which saw the rapid lysis of the culture. Bacteriocins are usually produced in stressful environments to eliminate neighbouring competitors that are not resistant (Riley and Gordon, 1999). Perhaps both FW7 and the wild type produce the same bacteriocins but the wild type

is more resistant and thus survives when produced. When investigating whether *C. beijerinckii* NCIMB 8052 harbours its own bacteriocin, genome comparison with the known bacteriocin database (Bagel3, http://bagel.molgenrug.nl/) revealed 3 potential bacteriocin genes. Two of the genes identified, Cbei_1376 and Cbei_3744, are annotated as UviB-like proteins which from protein BLAST analysis were found to have Holin_BhlA domains. Holin proteins were found to be produced by phages to disrupt bacterial cell membranes for progeny release (Anthony et al., 2010). Similarly, BhlA bacteriocins from *Bacillus* have been found to exert the same cell membrane disruptions on host cells leading to cell death (Anthony et al., 2010; Aunpad and Panbangred, 2012).

A further potential bacteriocin was a gene labelled as a CHAP domain-containing protein (Cbei_3964) which shares 61% protein sequence similarity with the BCN5 bacteriocin of *C. perfringens* (Garnier and Cole, 1986). CHAP domains have been found to be involved in peptidoglycan hydrolysis which can contribute to cell lysis (Bateman and Rawlings, 2003; Layec et al., 2007). Interestingly, Cbei_3964 was located next to the Agr quorum sensing system which may play a role in regulating the expression of this bacteriocin when numbers in the population reach critical levels. It was also found the *C. beijerinckii* NCIMB 8052 contains a gene annotated as a bacteriocin ABC transporter permease (Cbei_3324) which may aid export of the previously mentioned bacteriocins. Attempts were made to test if wild type and FW7 produced a bacteriocin that lysed self-cells using spent culture supernatant however these preliminary experiments were abandoned as they proved inconclusive (data not shown).

It has also been suggested that cells may undergo autolysis to provide nutrients and DNA to surrounding cells (Webb et al., 2003; West et al., 2006). Perhaps this system of autolysis exists in *C. beijerinckii* NCIMB 8052 where degenerate cells grow faster

but eventually die off to provide nutrients for the wild type to scavenge. Large scale autolysis has been observed in *C. acetobutylicum* ATCC 824 which was hypothesised to provide nutrients to sporulating cells (Liu et al., 2015b). Similar results have been observed for the related strain *C. beijerinckii* NRRL B-598 however it was the pH break point that caused the begin of loss of cell viability (Branska et al., 2018). These are all questions that would need further investigation and might, in part, be answered with the use of gene expression profiles at different time points.

Although FW cells may not survive past 48 hours, what is clear is that the loss of Spo0A activity gives a fitness benefit to these cells in the first 24 hours of growth. Spo0A has always been seen as a late stage expression gene that regulates the late stage phenotypes of solventogenesis and sporulation (Dürre, 2014; Dürre and Hollergschwandner, 2004; Ravagnani et al., 2000). The data presented here suggests however it is involved in the early stages as well as mutants gain their fitness advantage early on during exponential phase even before solvent production would normally commence (Figure 5.1 and 3.1). The Figure 5.1A insert of the first 12 hours of growth suggested that the FW7 left the very short lag phase rapidly and initially had a higher growth rate in the first eight hours than the wild type. This may explain the rapid enrichment of the FW type during serial transfers, independent of any social interactions. However, there is evidence for social interactions and growth at the expense of the wild type. Both the data from FW7 and *spo0A*::CT*ermB* grown against the wild type showed relative fitness much higher when their initial proportions were low and only showed a decreased fitness when they dominated the initial proportions. These results are indicative of frequency-dependent fitness (and thus social cheating) which sees cheaters able to do better when they are rare because there are more public goods/co-operators for them to exploit (Ross-Gillespie et al., 2007). As this fitness was not maintained it seems that these degenerates grow by a

'get rich quick scheme' which sees rapid growth in the early stages which cannot be sustained and is eventually detrimental to the culture. Why the loss of Spo0A activity gives a fitness benefit in the early stages is unknown but perhaps without the metabolic burden of Spo0A controlled expression cells can grow faster. With evidence for 'low and high' threshold genes regulated by Spo0A (Fujita and Losick, 2005; Fujita et al., 2005), perhaps in the early stages of growth, low levels of Spo0A will activate certain low threshold genes and so without a fully active Spo0A, the burden of these genes would not be present.

It was also found that not only the complete loss of Spo0A activity can give a fitness benefit. Mixed culture analysis of CIC4-1 against the wild type showed similar trends to FW7 with much higher relative fitness than the wild type at most starting ratios. It was interesting to observe that many of the 24 hour CIC4-1 fitness values were often higher than the 12 hour which was not seen for FW7. Furthermore, when comparing the fitness values of FW7, *spo0A*::*CTermB* and CIC4-1 when each were grown alongside the wild type, it was interesting to observe that FW7 and *spo0A*::*CTermB* were only higher than the wild type before they occupied 60% of the initial population compared to CIC4-1 fitness values only dramatically decreased when CIC4-1 occupied above 90% of the initial population. Perhaps this prolonged fitness is due to this strains ability to still produce solvents and therefore offset any potential acid toxicity.

When CIC4-1 was grown against FW7 there was only a marginal increased fitness by FW7 as generally all fitness values were above 1 but below 10. This suggests a sequential system in which complete loss of Spo0A activity gives the highest fitness benefit followed by partial loss. This may explain the emergence of intermediate types (DCOG and CIC) which show only partial Spo0A activity as they contain mutations in

genes that regulate *spo0A* expression or activity of the encoded protein but are eventually outcompeted by the full *spo0A* mutants.

It was interesting to observe that when FW7 was grown together with DCOG1-1, the fitness benefit of FW7 was only seen after the initial population contained 10% or more of FW7. It would seem that for FW7 to take over the population there must be a minimum threshold of this mutant for it to successfully outcompete other types. This appears to be reflected in the 24 hour subculturing experiments of Chapter 3 in which the DCOG were the first types to arise and take over the population followed by the FW as their numbers accumulated to a point where they could outcompete. However, when the initial proportion of FW7 was high, this lead to a decrease in fitness. Explaining this in terms of whole culture dynamics of the subculturing, perhaps the accumulation of other mutants which outcompeted the DCOG type lead to a population in which the FW types were eventually the fittest.

It should be noted that interpretations provided here are based on the analyses of selected isolates but that work in Chapter 3 and 4 has shown that there is variation within each mutant category. In degenerating cultures presumably multiple and distinct versions of the different types will exist as seen by both the phenotypic and genetic variation within morphotypes. The situation is further complicated by the fact that all these variants will be competing with each other which is rather different to the two strain approach adopted here. However, the observations made from this simplified experimental setup are largely in line with the results obtain in the 24 hour transfer setup and give a plausible explanation for the dynamic changes in morphotypes throughout subculturing.

The profiles of the 12 and 72 hour subculturing regimes also showed interesting and unexpected results. As FW7 has been shown to lose cell viability following entry into

the stationary phase, it had been hypothesised that shortening the transfer times down to 12 hours should lead to a more rapid accumulation of FW types and thus faster degeneration. Indeed, rapid degeneration was observed, however the overall profile in terms of product and spore profiles closely resembled that of the 24 hour transfer experiment (Chapter 3). Furthermore, it was interesting that the FW types did not seem to grow well in this regime and the culture was dominated by cells that appeared to be belong to the DCOG category. However, as pointed out in 5.3.8, this could be due to the emergence of the light centre with outgrowths (LCOG) which may reflect either a new and distinct metabolic type or a DCOG variant that is metabolically more similar to the FW category. It is clear from the fermentation products that the accumulation of acidogenic populations occurs rapidly as seen before for the 24 hour experiment.

The 72 hour (3 days) transfer saw mainly the RD type remaining in the culture which was interesting as previous experiments (Chapter 3) showed that this type is normally the first to disappear. Isolates of this type from this experiment were taken as stocks but were never analysed due to time constraints. It would be interesting to see how closely the RD types in the 3 day transfer were to the wild type both phenotypically and genetically. However, based on the relatively minor changes in terms of fermentation products and endospore level (heat resistant CFU), wild type like characteristics are to be expected. Moreover analysis of the phenotypic and genotypic profiles for the new LCOG isolates would also be useful to determine the level of degeneration these isolates had undergone and also how related they were to other types.

Whole culture genomic samples were also taken at successive subcultures for both the 12 hour and 72 hour transfer regimes to see if there were dynamic changes of

gene mutations. Again this was not completed due to time constraints. It should be noted that differences between the first subculture of the 12 and 72 hour transfers is due to the absence of a 0 subculture data point which was used to inoculate subculture 1 either 12 or 72 hours after inoculation of a colony. This data point is not included as inoculation of a single colony would not give values representative of a whole culture that had grown for either 12 or 72 hours.

Future analysis of the other types against FW7 would also be useful to determine relative fitness however under the time constraints it was not possible to grow FW7 against a Cbei_0017, Cbei_4884 or the Cbei_4885 *abrB* mutants described in Chapter 4. Many of these types also looked like the FW type, especially Cbei_4884 and *abrB* mutants, so distinguishing them may be difficult. Furthermore, a more reliable method to distinguish colony types such as the DCOG would also be required to complete mixed culture analysis of other degenerate types against the wild type.

Overall it seems that degenerates have gained a fitness advantage over the wild type by accumulating random mutations in the Spo0A network. This allows for the rapid enrichment and propagation of these mutants as they outcompete the wild type in the early stages of growth, depending on the applied culturing regime. Overall mutant fitness appears to be influenced by several factors such as pH and growth rates. These appear to be based on the exploitation of competing strains whereas others are also observed in monoculture.

6. General discussion

6.1. Background

Solventogenic *Clostridia* offer an attractive alternative for the production of biofuels through their unique acetone-butanol-ethanol fermentative metabolism as biobutanol is seen as a far superior fuel than traditional bioethanol (Jones and Woods, 1986; Kök, 2015; Lee et al., 2008b; Tracy et al., 2012). An intrinsic problem with solventogenic species is their tendency to degenerate which occurs when solventogenic strains are repeatedly transferred, such seen in batch fermentations, but also when grown continuously. The phenomenon of strain degeneration sees these species lose the ability to produce solvents and form endospores making it a significant problem when attempting to commercialise these organisms (Hartmanis et al., 1986; Jones and Woods, 1986; Kashket and Cao, 1995; Kutzenok and Aschner, 1952; Stephens et al., 1985; Woolley and Morris, 1990). In addition to losing sporulation and solventogenesis, changes in colony morphology are also observed (Adler and Crow, 1987; Woolley and Morris, 1990).

Several mechanisms of degeneration have been proposed which include the uncontrolled production of acids that prevent sufficient induction of the solvent genes, involvements of a global regulatory gene and the ability of cells to differentiate to endospores (Kashket and Cao, 1995). These mechanisms do elude to plausible explanations but do not provide a definitive cause as to why strains degenerate. Without a definitive cause, the focus on solventogenic *Clostridia* soon shifted away from degeneration and towards strain improvements through metabolic engineering (Lütke-Eversloh and Bahl, 2011). This has still left many questions unanswered and the problem of strain degeneration still needs to be addressed.

The lack of a conclusive answer and prevention of the phenomenon was the main motivator behind this work. Furthermore, there has been little to no investigation into the genetic causes of degeneration. It was hoped that in the age of ''omics'', new techniques such as next generation sequencing could be applied to further elude to any genetic causes of degeneration. Similarly, the fledgling research area of sociomicrobiology (West et al., 2006) may also provide new information as to evolutionary forces that drive strains to degenerate. This study aimed to use a combination of genetic, ecological and evolutionary approaches to elude to the factors behind strain degeneration of solventogenic *Clostridia* using the species *Clostridium beijerinckii* NCIMB 8052. This strain was chosen due to its tendency to degenerate more rapidly than other strains and because it has served as a parental precursor strain for the hyper-butanol producer *C. beijerinckii* BA101 (Chen and Blaschek, 1999a; Kashket and Cao, 1995; Stephens et al., 1985; Woolley and Morris, 1990). Investigating degeneration may also help improve our understanding of the wild type organism as a whole, apart from being relevant to its application in a biotechnological process.

6.2. Key findings and future work

6.2.1. Development of a reliable subculturing protocol and the differing colony morphologies that emerged

To study degeneration, it was first necessary to implement a reliable protocol that would promote the emergence of degenerates quickly and easily. The subculturing protocol described in Chapter 3 was based on serial transfers, as this is known to promote degeneration, at 24 hour intervals. 24 hours was chosen as this has been

previously used in the literature (Finn and Nowrey, 1959; Hartmanis et al., 1986). The protocol developed in this study proved to be robust and reliable with the emergence of aberrant colony types with as little as 1 subculture. It was quickly found that serial transfers of *C. beijerinckii* NCIMB 8052 produced four distinct colony morphologies. Several examples of these morphologies can be seen in Chapter 3 and were divided into the round and dark (RD), dark centre with outgrowths (DCOG), caved in centre (CIC) and the flat and white (FW). A total of 71 degenerate colonies were isolated after five transfers through two isolations. This gave a total of 17 RD, 17 DCOG, 18 CIC and 19 FW that were used throughout this study.

6.2.2 Phenotypic analysis of the isolated degenerates

Phenotypic analysis of each of the types revealed an array of solvent, spore and granulose producing capabilities. It was found that with some exceptions, generally the RD, DCOG and CIC morphotypes would produce solvents at similar concentrations to that of the wild type after 48 hours of growth and all three types showed similar product profiles. It was interesting that acetone production often appeared slightly higher than the wild type; however, this was for the most part not statistically significantly different. Except for RD1, all of the RD, DCOG and CIC morphotypes produced significantly less spores than wild type even when their solvent producing capabilities seemed normal. The most striking phenotypes were seen from the FW morphotype. Most isolates of this type produced little to no solvents but had significantly increased concentrations of butyrate and acetate compared to the wild type. Furthermore, nearly all of the FW isolates produced no spores whatsoever. It

would seem that the FW represents strains that have either nearly or fully degenerated and pose the greatest problem in fermentations.

Previous studies into degeneration have also found vastly different colony phenotypes. Studies using *C. acetobutylicum* strains NRRL B643, ATCC 4259 and ATCC 824, found four colony types with the most severe unable to produce solvents or spores but included several intermediate types (Adler and Crow, 1987). Comparing the colony types observed in that study, one appeared with dark centres and many outgrowths (like the DCOG) and another with brown centres with smooth edges (like the RD). Two other types were described as one having a grey-brown centre and another with a light brown centre and indented edges. It was the latter colony type that produced no solvents and spores but it was interesting that they did not describe one that resembled the FW type from this study.

In previous *C. beijerinckii* NCIMB 8052 studies, three colony types were described with two being intermediate solvent and spore producers and a final colony type that was as unable to produce either (Kashket and Cao, 1995). The first colony type was described to have thin beige outgrowths (perhaps the DCOG), the second had a chestnut brown centre (perhaps the RD) and the third was described as fully degenerated (perhaps the FW). Interestingly from that study, cells from the later colony could not be cultured which seem unusual considering healthy stocks of the four types isolated in this study were made. Furthermore, four colony types were observed in this study compared to the three from previous literature exemplifying the far more in depth analysis completed here.

6.2.3 Subculturing revealed dynamic changes in the culture

To understand how colony type may influence the whole culture, a more detailed analysis of the whole culture whilst subculturing was undertaken. This study looked at the solvent, acid and spore producing capability of the culture whilst simultaneously observing colony morphology. It was found that at each progressive 1:10 subculture every 24 hours, the concentration of solvents decreased and the concentration of acids increased, a process which clearly sees the overall culture degenerating. A similar dynamic had previously been described for a strain of *C. butylicum* (Kutzenok and Aschner, 1952). In that study however, the rapid decline of butanol and the increase in butyrate was only seen after subculture 7. Furthermore, colony morphology was only observed with one image comparing wild type and a degenerate and that it was briefly described that the amount of clostridial forms and spores decreased along with solvents.

In this study, it was interesting to observe the varying colony types that emerged, and declined, throughout the successive subcultures. In these experiments, the sequence of events saw the culture first taken over by the DCOG type which was able to outcompete the initial population of the RD type within the first five subcultures. This take over by DCOG was however short lived as numbers of the FW type began to steadily increase from subculture two onwards until DCOG was finally outcompeted by FW which equalled <90% of the population at the end of the 24 hour subculturing. A small increases in the CIC type was seen half way through the subculturing but this was not sustained that they were outcompeted the FW also. Perhaps this type gained fitness when the DCOG type had declined but the large levels of FW meant that this fitness benefit was short lived.

The FW type appeared to outcompete all the other types implying that it had the greatest fitness advantage. Mixed culture analysis undertaken in Chapter 5 revealed, at least in the case of FW7, that the FW type was able to outcompete the wild type at low initial frequencies. Furthermore, this FW type was also fitter than a CIC type (CIC4-1) and a DCOG type (DCOG1-1) at various starting ratios. It was found however that a minimum amount of initial FW7 (10%+) was needed before it could outcompete DCOG1-1. This may explain the observed changes in colony type composition whilst subculturing as the DCOG type is the first to take over the population and is only outcompeted when sufficient FW type has accumulated in the population. It should be considered, however, that the situation is most likely far more complex as many variants of the DCOG and FW types exist with varying phenotypes and genotypes. Therefore, the initial DCOG that dominate the culture may be fitter than the first FW types to emerge but evolution of the FW types may create isolates that are better in the later subcultures compared to the start.

Out of the isolates tested in the mixed culture analysis, however, FW7 proved to have the greatest fitness. It would be interesting to compare other types against the wild type to observe if the fitness that FW7 presented against the wild type was mirrored by other types but this proved difficult due to colonies looking too similar. A tetrazolium blue plate assay has previously been employed to distinguish solvent and non-solvent producing colonies (Woolley and Morris, 1990). When sprayed with tetrazolium blue chloride mixture, colonies that produce solvents appear reddish-blue whilst non-solvent producing colonies appear pale pink/white due to a lower local pH from the acidogenic colonies. This assay could be employed to distinguish the colony types from each other to allow, perhaps in conjunction with granulose staining, fitness to be assessed for other types. Differences in colony colour between isolates would need to be tested beforehand to determine if they could be distinguished reliably.

It would also be useful for future analysis to compare other types against each other, for example DCOG1-1 vs CIC4-1, to understand culture dynamics more thoroughly. Furthermore, no mixed culture analysis of the hypothetical protein Cbei_4884 and Cbei_4885 *abrB* mutants against the wild type or other types was completed due to time constraints. These may prove to be valuable in further proving the role of Spo0A in fitness and the potential sequential order of Spo0A activity loss.

6.2.4 Genetic analysis of the isolated degenerates revealed four hot spot regions

Whole genome sequences of the 71 isolated degenerates were compared to the two parental wild types (of clone 3 lineage) that they were isolated from. Amongst the wide range of mutations found, genome sequence analysis revealed four distinct hot spot regions which contained more mutations than other regions in the genome. These regions included the genes for the master regulatory protein Spo0A, two histidine kinases, a hypothetical protein, the transcriptional regulator gene *abrB* as well as the intergenic region between the latter two. In solventogenic *Clostridia* Spo0A in an incredibly important protein as it controls the switch from acidogenesis to solventogenesis alongside spore formation (Al-Hinai et al., 2015; Alsaker et al., 2004; Dürre, 2014; Dürre and Hollergschwandner, 2004; Ravagnani et al., 2000). This switch is needed to prevent the accumulation of toxic levels of acids in the medium which would be detrimental to the pH gradient across the cell membrane and so they are reassimilated and converted to solvents (Jones and Woods, 1986). When comparing the genotype to phenotypes, many of the isolates that contained mutations in *spo0A* had a significantly reduced solvent and spore forming capability. One such mutant was FW7 which contained a mutation in the DNA binding site of Spo0A rending the

protein inactive. Interestingly, FW4 and FW5-2 had mutations within the specific DNA minding motif that did not render the protein completely inactive as both produced some solvents but did not produce spores. More interesting was that both had the exact same mutation even though they were isolated in different subculturing regimes. Several other isolates also existed with mutations outside of this specific domain causing a reduced activity (solvents were still produced) but not complete inactivation. Two in particular, FW2 and FW4-2, had mutations that caused an Aspartate (Asp) residue to change (in the N-terminal phosphoacceptor domain) which, as Asp residues are known to accept phosphoryl groups, may have caused a reduced ability for the protein to be phosphorylated and thus activated (Brown et al., 1994; Escobar et al., 2009; Grimsley et al., 1994; Lewis et al., 2000). It would be interesting to test if these Asp residues are important for Spo0A phosphorylation through phosphorylation assays to understand the Spo0A of *C. beijerinckii* NCIMB 8052 more thoroughly.

As Spo0A contains two domains, the N-terminal phosphoracceptor and C-terminal DNA binding (Brown et al., 1994; Escobar et al., 2009; Grimsley et al., 1994; Ravagnani et al., 2000), mutations in each of these domains will affect both of these functions causing varying levels of activity. Mutants have been isolated in *Bacillus* that have mutations in both of the domains of Spo0A which have resulted in reduced sporulation (Cervin and Spiegelman, 2000; Schmeisser et al., 2000). It would seem that both domains are important for Spo0A activity either through its activation or through the subsequent DNA binding. Analysis in Chapter 4 found mutations throughout the *spo0A* gene giving various phenotypes (Chapter 3). The mutational evidence found here confirms both the specificity of the binding site (Ravagnani et al., 2000), that Spo0A can exist with varying amounts of activity and still complete

activation of some gene transcription and perhaps this varying transcription is linked to controlling gene expression of various genes.

Previous work in *C. acetobutylicum* ATCC 824 revealed that Spo0A activity is regulated by direct phosphorylation by orphan histidine kinases instead of a phosphorelay system seen in *Bacillus* (Steiner et al., 2011). The orphan histidine kinases Cbei_0017 and Cbei_3078 may well be involved in the phosphorylation system that regulates Spo0A activity in *C. beijerinckii* NCIMB 8052 as several isolated mutants of these two genes produced much less spores than the wild type. ClosTron mutagenesis of these genes gave reduced sporulation phenotypes and complementation of these genes back onto the chromosome was able to restore function, although only partially for Cbei_3078. Although they had relatively low sequence identity to that of the orphan histidine kinases found by Steiner et al (2011) further investigations into the genes is needed to confirm their role. It would be interesting to repeat the protein autophosphorylation and the Spo0A phosphotransfer assays from Steiner et al (2011) using purified Cbei_0017 and Cbei_3078 protein. This would indicate whether these histidine kinases could autophosphorylate using ATP and transfer their phosphoryl group to Spo0A thus proving their role in Spo0A activation.

Analysis of the Cbei_4885 *abrB* gene indicated that it is likely to have the same role as *abrB0310* (CA_C0310) of *C .acetobutylicum* ATCC 824 due to the 0A box next to the Cbei_4885 gene and the sequence similarity. Scotcher et al. (2005) used antisense RNA to block expression of *abrB0310* which showed delayed onset of solventogenesis implying that *abrB0310* exists to aid the transition into the solvent phase. Furthermore, Xue et al. (2016), using the Targetron system to disrupt *abrB0310*, showed similar results as this mutant had produced reduced solvents concentrations to the wild type. If these two proteins (Cbei_4885 and AbrB0310) are functionally

homologous, the role of this gene would be to aid the transition from acidogenesis to solventogenesis through Spo0A regulation (Scotcher et al., 2005; Xue et al., 2016). Further investigations into Cbei_4885 would first require knocking the gene out to observe phenotypes. ClosTron mutagenesis was attempted however as the gene is small, the chance of off target insertions were high. Attempts were made as part of this study (Chapter 5) but no ClosTron mutants were isolated. With lack of a successful independent mutant, it was hoped to complement one of the spontaneous, degenerate Cbei_4885 mutants to demonstrate a link between the putative *abrB* gene and the observed phenotype. Attempts were made to transform a copy of Cbei_4885 including the upstream intact 0A box into FW1-1 and isolates were obtained that were resistance to the antibiotic conferred by the plasmid; however time constraints meant that phenotypic analysis was not undertaken. The role of this gene seems important as FW1-1 showed delayed solventogenesis but further investigations would be needed to confirm this. Furthermore, four other *abrB* genes are annotated in the *C. beijerinckii* NCIMB 8052 genome (Wang et al., 2012) which may also play a role in solventogenesis and sporulation as it has been shown that other AbrB exert different effects on Spo0A (Xue et al., 2016).

Similarly, attempts were made to use ClosTron mutagenesis on the hypothetical protein Cbei_4884 but this proved ineffective as no mutant was isolated. Furthermore, a plasmid based copy of Cbei_4884 was transformed into FW3-2 (this isolate had a deletion at the start of the gene) and isolates were obtained that were resistance to the antibiotic conferred by the plasmid. Again, no phenotypic analysis was undertaken due to time constraints. The role of this gene is currently unknown but mutants of this gene showed reduced solvents and spore production.

When linking colony type to genotype, there was a degree of similarity within morphotypes. Most of the Cbei_0017 mutants were of the RD morphotype (77%), most the Cbei_3078 mutants were DCOG (57%) and most of the direct *spo0A* mutants were FW (63%). The CIC type represented the least genotypically defined as mutants of this type were found in each of the hotspots. This may be reflected by the subjective nature of colony picking and so some of this type may have been classed incorrectly. Alternatively, it may reflect the complexity of the regulatory network and the nature of the defects introduced by the respective SNVs.

It should be noted that mutational rates for these hotspot regions were not determined as the underlying mutagenic cause or rate was not of importance to this study. This study set out to determine any genes that are involved in degeneration. However, it may be useful for future analysis to determine rates between regions and whether there are differences between strains. Many of the independently isolated degenerates had different mutations in different locations of the hotspot regions suggesting that increased mutations in these regions was not caused by physical instability in a particular short stretch of sequence. For instance, such fragile sites have been reported in eukaryotes and are a possible cause of cancer (Feng and Chakraborty, 2017; Ma et al., 2012) but do not seem to present in this organism. The existence of these four hotspots appeared to be a result of rapid enrichment due to a much increased fitness.

6.2.5 Spo0A is important in degeneration

The genetic analysis of the 71 degenerate isolates all point towards Spo0A and the network that regulates the expression and activity of this gene/protein. It has been

postulated that a global regulatory gene is responsible for degeneration (Kashket and Cao, 1995) and it would seem that Spo0A fits this assumption. This gene regulates solventogenesis and sporulation, two key physiological changes solventogenic *Clostridia* undergo (Dürre, 2014; Dürre and Hollergschwandner, 2004). Due to the regulatory nature of Spo0A, loss of this gene will give the phenotypes of a degenerate i.e. reduced solvents and spores which has been documented in the literature for both insertional and clean deletions (Alsaker et al., 2004; Tomas et al., 2003; Wang et al., 2015). What is most interesting is that mutants of *spo0A* and the genes that potentially control this network seem to be selected for in the population. Mutants of this network are able to propagate throughout the culture due to a presumed increased fitness. Some of the other mutations found during the analysis which have not been investigated may well be involved in the degeneration process; however, many most likely exist due to carry over through these propagators.

Several mixed culture experiments performed in Chapter 5 showed that loss of Spo0A activity gave a direct fitness benefit compared to the wild type. Similarly, even losing partial activity still gave a fitness benefit to the mutant. There appeared to be a hierarchy in terms of relative fitness level whereby complete loss of Spo0A (FW7) showed the greatest fitness as it produced fitness values higher than all isolates it was grown with. This is then presumably followed by CIC4-1, which had partial Spo0A loss with a mutation in the presumed phosphoacceptor domain, which produced higher fitness values when grown with the wild type further into the increasing initial proportions but did not do better when grown alongside FW7. Following these mutants, the next to have increased fitness would presumably be mutants with mutations in the histidine kinases that regulate Spo0A activity as mixed culture analysis of FW7 vs DCOG1-1 only showed a marginal increase in fitness for FW7. It was unfortunate that DCOG1-1 was unable to be grown against the wild type due to their

appearance being too similar on agar plates. Furthermore, it would be useful to grow DCOG1-1 against CIC4-1 to see if, presumably, CIC4-1 would be marginally fitter than DCOG1-1 however this was not possible to complete due to time constraints.

With the indication that Spo0A activity is important for culture stability, perhaps to prevent or reduce degeneration the selective pressure to maintain a fully functional gene needs to be increased. This could be achieved through the coupling of Spo0A to an essential gene so that loss of Spo0A would not be possible. This might be done by linking a 0A box to an essential gene so that expression of this gene would only be achieved via Spo0A activation. A problem with this approach, however, is that according to current understanding Spo0A activation only occurs at entry into stationary phase (e.g. *C. acetobutylicum*) or, the second half of exponential growth (e.g. *Clostridium beijerinckii*, as evidenced by the onset of solvent formation, Figure 3.1). However, the increased growth rate seen for FW7 mutants particularly during early exponential growth (Figure 5.1A) suggests roles for this regulator at even this early stage of growth. This is supported through recent transcriptome comparisons for FW7 and wild type, which revealed differential expression of genes even before solvent formation took place at OD 0.2 (Humphreys & Winzer, unpublished). Even though mutants with a reduced Spo0A activity may still emerge in the population of such an engineered strain, a slower growth due to reduced expression of the essential gene may prevent their enrichment.

The observed differences in the early gene expression for the *spo0A* mutant FW7 also suggest another approach to achieve a more stable strains. A faster initial growth rate suggests a lower burden due to a lack of costly functions that are not required for growth at this stage. If these unnecessary early functions can be identified, and are not required for solvent production or stress resistance later on, it may be possible to

eliminate those functions thereby increasing the relative fitness of the respective strains to levels seen for degenerates.

Furthermore, growth rate definitely appears important in preventing degeneration as transposon insert isolates that show decreased growth rates to the wild type are more resistant to degeneration (Evans et al., 1998; Kashket and Cao, 1993, 1995). This was observed from an isolated strain that had a transposon insertion in a peptide deformylase (PDF) gene. PDF is known to remove *N*-formyl from *N*-formylmethionine at the *N* termini of polypeptides, giving formate and deformylated polypeptides and are essential in *E. coli* (Evans et al., 1998). The insertion in their isolated strain was in the 5' region of the gene which the authors believed caused reduced, but not abolished, activity of the protein and thus gave a decrease in growth rate. Stability is therefore improved as the tendency of the wild type to produce excess acids and no solvents is reduced due to a vastly slower growth rate (Evans et al., 1998). However, without constant antibiotic selection for the insertion (an added cost for commercialisation), potential fast growing mutants may arise within the population negating the positive effects of the slower growing strains. Changing the growth rate of the whole culture may therefore be important in reducing the accumulation of fast, degenerate growers and should be explored for future degeneration prevention.

Prevention of degeneration may also be achieved through varying subculturing regimes. This study found that when subculturing was done at 12 and 24 hour transfers, the culture would degenerate sequentially and rapidly, contrasting with the 72 hour transfer regime in which solvents were largely maintained. Perhaps varying the subculturing regime will help to prevent degeneration during the preparation of inocula for large scale industrial fermentations. However, this would still leave the issue of degeneration in continuous chemostat cultures which is a phenomenon

known to occur (Finn and Nowrey, 1959; Stephens et al., 1985; Woolley and Morris, 1990). The use of continuous fermentation was never studied in this work but previous studies showed found five colony phenotypes all with varying solvent, granulose and spore producing capabilities (Woolley and Morris, 1990). From these studies, no strains were isolated that sporulated, produced granulose but did not produce solvents or ones that solely sporulated indicting the link between solvent and spore production. It would be interesting to observe which of the colony types described in this work would emerge when *C. beijerinckii* NCIMB 8052 is grown in a chemostat culture and how relatable they are to the ones isolated from subculturing. Moreover, as transfer times appear to change the rate of degeneration, would changing the dilution rate of a continuous fermentation have the same effect? These questions would need to be answered to tackle the degeneration problem for all aspects of industry.

Ultimately, with a further advances in our understanding of the organisms, solvent metabolism and improved gene editing tools, construction of strains that produce only alcohols, and also do not produce spores as sporulation will affect productivity, may become possible with the elimination of all other metabolic pathways. Without the option to resort to alternative fermentation metabolism, such strains may not be prone to degeneration as any mutations resulting in an increased fitness should not affect solvent metabolism as this would be the only means of ATP generation.

6.3 Concluding remarks

This study has contributed to a deeper understanding into the phenomenon of degeneration by eluding to the role of the master regulator gene *spo0A*. Genetic and

social experimentation revealed that a reduced activity or complete loss of this important gene gives a fitness benefit to the mutant. Mutants are therefore able to propagate throughout the population due their ability to outcompete the wild type. These mutants usually lack the ability to produce solvents and spores effectively making them undesirable in a commercial fermentation. Degeneration still poses a problem for industry but with a greater understanding of the genes involved, it is hoped that prevention can be achieved.

Bibliography

Adler, H.I., and Crow, W. (1987). A Technique for Predicting the Solvent-Producing Ability of Clostridium acetobutylicum. Appl. Environ. Microbiol. *53*, 2496–2499.

Aguilar, C., Vlamakis, H., Losick, R., and Kolter, R. (2007). Thinking about Bacillus subtilis as a multicellular organism. Curr. Opin. Microbiol. *10*, 638–643.

Akasaka, S., Takimoto, K., and Yamamoto, K. (1992). G:C-->T:A and G:C-->C:G transversions are the predominant spontaneous mutations in the Escherichia coli supF gene: an improved lacZ(am) E. coli host designed for assaying pZ189 supF mutational specificity. Mol. Gen. Genet. *235*, 173–178.

Al-Hinai, M.A., Fast, A.G., and Papoutsakis, E.T. (2012). Novel system for efficient isolation of Clostridium double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and DNA integration. Appl. Environ. Microbiol. *78*, 8112–8121.

Al-Hinai, M.A., Jones, S.W., and Papoutsakis, E.T. (2015). The Clostridium sporulation programs: diversity and preservation of endospore differentiation. Microbiol. Mol. Biol. Rev. *79*, 19–37.

Alsaker, K. V, Spitzer, T.R., and Papoutsakis, E.T. (2004). Transcriptional analysis of spo0A overexpression in Clostridium acetobutylicum and its effect on the cell's response to butanol stress. J. Bacteriol. *186*, 1959–1971.

Alvira, P., Tomás-Pejó, E., Ballesteros, M., and Negro, M.J. (2010). Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. Bioresour. Technol. *101*, 4851–4861.

Andersch, W., Bahl, H., and Gottschalk, G. (1983). Level of enzymes involved in

acetate, butyrate, acetone and butanol formation by Clostridium acetobutylicum. Eur. J. Appl. Microbiol. Biotechnol. *18*, 327–332.

Andersen, S.B., Marvig, R.L., Molin, S., Krogh Johansen, H., and Griffin, A.S. (2015). Long-term social dynamics drive loss of function in pathogenic bacteria. Proc. Natl. Acad. Sci. U. S. A. *112*, 10756–10761.

Anthony, T., Gunasekaran, ·, Chellappa, S., Rajesh, T., and Gunasekaran, P. (2010). Functional analysis of a putative holin-like peptide-coding gene in the genome of Bacillus licheniformis AnBa9. Arch Microbiol *192*, 51–56.

Artzi, L., Bayer, E.A., and Moraïs, S. (2017). Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nat. Rev. Microbiol. *15*, 83–95.

Aunpad, R., and Panbangred, W. (2012). Evidence for Two Putative Holin-Like Peptides Encoding Genes of Bacillus pumilus Strain WAPB4. Curr. Microbiol. *64*, 343–348.

Avery, S. V. (2006). Microbial cell individuality and the underlying sources of heterogeneity. Nat. Rev. Microbiol. *4*, 577–587.

Bahl, H., Andersch, W., Braun, K., and Gottschalk, G. (1982). Microbiology and Effect of pH and Butyrate Concentration on the Production of Acetone and Butanol by Clostridium acetobutylicum Grown in Continuous Culture. Eur. J. Appl. Microbiol. Biotechnol. *14*, 17–20.

Baldus, J.M., Green, B.D., Youngman, P., and Moran, C.P. (1994). Phosphorylation of Bacillus subtilis Transcription Factor SpoOA Stimulates Transcription from the spoIIG Promoter by Enhancing Binding to Weak OA Boxes. J. Bacteriol. 296–306.

Barbeau, J.Y., Marchal, R., and Vandecasteele, J.P. (1988). Conditions promoting stability of solventogenesis or culture degeneration in continuous fermentations of

Clostridium acetobutylicum. Appl. Microbiol. Biotechnol. *29*, 447–455.

Barber, J.M., Robb, F.T., Webster, J.R., and Woods, D.R. (1979). Bacteriocin production by Clostridium acetobutylicum in an industrial fermentation process. Appl. Environ. Microbiol. *37*, 433–437.

Bateman, A., and Rawlings, N.D. (2003). The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. Trends Biochem. Sci. *28*, 234–237.

Bayer, E.A., Belaich, J.-P., Shoham, Y., and Lamed, R. (2004). The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu. Rev. Microbiol. *58*, 521–554.

Behjati, S., and Tarpey, P.S. (2013). What is next generation sequencing? Arch. Dis. Child. Educ. Pract. Ed. *98*, 236–238.

Bertelli, C., and Greub, G. (2013). Rapid bacterial genome sequencing: methods and applications in clinical microbiology. Clin. Microbiol. Infect. *19*, 803–813.

Biegel, E., Schmidt, S., González, J.M., and Müller, V. (2011). Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. Cell. Mol. Life Sci. *68*, 613–634.

Boeke, J.D., LaCroute, F., and Fink, G.R. (1984). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol. Gen. Genet. *197*, 345–346.

Bowles, L.K., and Ellefson, W.L. (1985). Effects of butanol on Clostridium acetobutylicum. Appl. Environ. Microbiol. *50*, 1165–1170.

BP (2016). BP Statistical Review of World Energy 2016.

Branda, S.S., González-Pastor, J.E., Ben-Yehuda, S., Losick, R., and Kolter, R. (2001). Fruiting body formation by Bacillus subtilis. Proc. Natl. Acad. Sci. U. S. A. *98*, 11621–11626.

Branska, B., Pechacova, Z., Kolek, J., Vasylkivska, M., and Patakova, P. (2018). Flow cytometry analysis of Clostridium beijerinckii NRRL B-598 populations exhibiting different phenotypes induced by changes in cultivation conditions. Biotechnol. Biofuels *11*, 99.

Brown, D.P., Ganova-Raeva, L., Green, B.D., Wilkinson, S.R., Young, M., and Youngman, P. (1994). Characterization of spo0A homologues in diverse Bacillus and Clostridium species identifies a probable DNA-binding domain. Mol. Microbiol. *14*, 411–426.

Brown, N.L., Stoyanov, J. V, Kidd, S.P., and Hobman, J.L. (2003). The MerR family of transcriptional regulators. FEMS Microbiol. Rev. *27*, 145–163.

Buckel, W., and Thauer, R.K. (2013). Energy conservation via electron bifurcating ferredoxin reduction and proton/Na+ translocating ferredoxin oxidation. Biochim. Biophys. Acta - Bioenerg. *1827*, 94–113.

Carlquist, M., Fernandes, R., Helmark, S., Heins, A.-L., Lundin, L., Sørensen, S.J., Gernaey, K. V, and Lantz, A. (2012). Physiological heterogeneities in microbial populations and implications for physical stress tolerance. Microb. Cell Fact. *11*, 94.

Cartman, S.T., and Minton, N.P. (2010). A mariner-based transposon system for in vivo random mutagenesis of Clostridium difficile. Appl. Environ. Microbiol. *76*, 1103–1109.

Cartman, S.T., Kelly, M.L., Heeg, D., Heap, J.T., and Minton, N.P. (2012). Precise manipulation of the Clostridium difficile chromosome reveals a lack of association

between the tcdC genotype and toxin production. Appl. Environ. Microbiol. *78*, 4683–4690.

Cavaliere, M., Feng, S., Soyer, O.S., and Jim enez, J.I. (2017). Minireview Cooperation in microbial communities and their biotechnological applications. Environ. Microbiol. *19*, 2949–2963.

Cervin, M.A., and Spiegelman, G.B. (2000). A Role for Asp 75 in Domain Interactions in the Bacillus subtilis Response Regulator Spo0A. J. Biol. Chem. *275*, 22025–22030.

Chaturvedi, V., and Verma, P. (2013). An overview of key pretreatment processes employed for bioconversion of lignocellulosic biomass into biofuels and value added products. 3 Biotech *3*, 415–431.

Chen, C.K., and Blaschek, H.P. (1999a). Acetate enhances solvent production and prevents degeneration in Clostridium beijerinckii BA101. Appl. Microbiol. Biotechnol. *52*, 170–173.

Chen, C.K., and Blaschek, H.P. (1999b). Effect of acetate on molecular and physiological aspects of Clostridium beijerinckii NCIMB 8052 solvent production and strain degeneration. Appl. Environ. Microbiol. *65*, 499–505.

Chen, P., and He, C. (2003). A General Strategy to Convert the MerR Family Proteins into Highly Sensitive and Selective Fluorescent Biosensors for Metal Ions. JACS Commun. *126*, 728–729.

Chylinski, K., Le Rhun, A., and Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biol. *10*, 726–737.

Cook, T.M., Protheroe, R.T., and Handel, J.M. (2001). Tetanus: a review of the literature. Br. J. Anaesth. *87*, 477–487.

Cooksley, C.M., Zhang, Y., Wang, H., Redl, S., Winzer, K., and Minton, N.P. (2012). Targeted mutagenesis of the Clostridium acetobutylicum acetone–butanol–ethanol fermentation pathway. Metab. Eng. *14*, 630–641.

Cornillot, E., Nair, R. V., Papoutsakis, E.T., and Soucaille, P. (1997). The genes for butanol and acetone formation in Clostridium acetobutylicum ATCC 824 reside on a large plasmid whose loss leads to degeneration of the strain. J. Bacteriol. *179*, 5442–5447.

Cousineau, B., Smith, D., Lawrence-Cavanagh, S., Mueller, J.E., Yang, J., Mills, D., Manias, D., Dunny, G., Lambowitz, A.M., and Belfort, M. (1998). Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. Cell *94*, 451–462.

Crabbendam, P.M., Neijssel, O.M., and Tempest, D.W. (1985). Metabolic and energetic aspects of the growth of Clostridium butyricum on glucose in chemostat culture. Arch. Microbiol. *142*, 375–382.

Darch, S.E., West, S.A., Winzer, K., and Diggle, S.P. (2012). Density-dependent fitness benefits in quorum-sensing bacterial populations. Proc. Natl. Acad. Sci. *109*, 8259–8263.

Debabov, V.G. (2015). Modern approaches to the creation of industrial microorganism strains. Russ. J. Genet. *51*, 365–376.

Demirbas, A. (2008). Biofuels sources, biofuel policy, biofuel economy and global biofuel projections. Energy Convers. Manag. *49*, 2106–2116.

Dentener, F., Derwent, R., Dlugokencky, E., Holland, E., Isaksen, I., Katima, J., Kirchhoff, V., Matson, P., and P. Midgley, M.W. (2007). Atmospheric Chemistry and Greenhouse Gases. Agriculture *33*, 1–44.

Denver, D.R., Dolan, P.C., Wilhelm, L.J., Sung, W., Lucas-Lledo, J.I., Howe, D.K., Lewis, S.C., Okamoto, K., Thomas, W.K., Lynch, M., et al. (2009). A genome-wide view of Caenorhabditis elegans base-substitution mutation processes. Proc. Natl. Acad. Sci. *106*, 16310–16314.

Diggle, S.P., Griffin, A.S., Campbell, G.S., and West, S.A. (2007). Cooperation and conflict in quorum-sensing bacterial populations. Nature *450*, 411–414.

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. Trends Genet. *30*, 418–426.

Dillon, M.M., Sung, W., Lynch, M., and Cooper, V.S. (2015). The Rate and Molecular Spectrum of Spontaneous Mutations in the GC-Rich Multichromosome Genome of Burkholderia cenocepacia. Genetics *200*, 935–946.

Driks, A., Setlow, P., and Setlow, P. (2000). Morphogenesis and Properties of the Bacterial Spore. In Prokaryotic Development, (American Society of Microbiology), pp. 191–218.

Dürre, P. (2014). Physiology and Sporulation in Clostridium. Microbiol. Spectr. *2*.

Dürre, P. (Peter) (2005). Handbook on clostridia (Taylor & Francis).

Dürre, P., and Hollergschwandner, C. (2004). Initiation of endospore formation in Clostridium acetobutylicum. Anaerobe *10*, 69–74.

Edwards, D.J., and Holt, K.E. (2013). Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. Microb. Inform. Exp. *3*, 2.

Ehsaan, M., Kuit, W., Zhang, Y., Cartman, S.T., Heap, J.T., Winzer, K., Minton, N.P., Stevens, D., Aldape, M., Bryant, A., et al. (2016). Mutant generation by allelic exchange and genome resequencing of the biobutanol organism Clostridium

acetobutylicum ATCC 824. Biotechnol. Biofuels *9*, 4.

Enfors, S.-O., Jahic, M., Rozkov, A., Xu, B., Hecker, M., Jürgen, B., Krüger, E., Schweder, T., Hamer, G., O'Beirne, D., et al. (2001). Physiological responses to mixing in large scale bioreactors. J. Biotechnol. *85*, 175–185.

Escobar, A.J., Gutiérrez, and Castano, D.M. (2009). Evolutionary Analysis for the Functional Divergence of the Spo0A Protein: The Key Sporulation Control Element. In Silico Biol. *9*, 149–162.

Evans, V.J., Liyanage, H., Ravagnani, A., Young, M., and Kashket, E.R. (1998). Truncation of peptide deformylase reduces the growth rate and stabilizes solvent production in Clostridium beijerinckii NCIMB 8052. Appl. Environ. Microbiol. *64*, 1780–1785.

Ezeji, T.C., and Li, Y. (2010). Advanced Product Recovery Technologies. In Biomass to Biofuels, (Oxford, UK: Blackwell Publishing Ltd.), pp. 331–345.

Ezeji, T., Milne, C., Price, N.D., and Blaschek, H.P. (2010). Achievements and perspectives to overcome the poor solvent resistance in acetone and butanol-producing microorganisms. Appl Microbiol Biotechnol *85*, 1697–1712.

Ezeji, T.C., Karcher, P.M., Qureshi, N., and Blaschek, H.P. (2005). Improving performance of a gas stripping-based recovery system to remove butanol from Clostridium beijerinckii fermentation. Bioprocess Biosyst. Eng. *27*, 207–214.

Fargione, J., Hill, J., Tilman, D., Polasky, S., and Hawthorne, P. (2008). Land clearing and the biofuel carbon debt. Science *319*, 1235–1238.

Feng, W., and Chakraborty, A. (2017). Fragility Extraordinaire: Unsolved Mysteries of Chromosome Fragile Sites. Adv. Exp. Med. Biol. *1042*, 489–526.

Finn, R.K., and Nowrey, J.E. (1959). A note on the stability of clostridia when held in continuous culture. Appl. Microbiol. *7*, 29–32.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., et al. (1995). Whole-Genome Random Sequencing and Assembly of Haemophilus influenzae Rd. Science (80-.). *269*, 496–512.

Fond, O., Matta-Ammouri, G., Petitdemange, H., and Engasser, J.M. (1985). The role of acids on the production of acetone and butanol by Clostridium acetobutylicum.

Foster, P.L., Lee, H., Popodi, E., Townes, J.P., and Tang, H. (2015). Determinants of spontaneous mutation in the bacterium Escherichia coli as revealed by whole-genome sequencing. PNAS.

Frank, S.A. (2003). Perspective: repression of competition and the evolution of cooperation. Evolution *57*, 693–705.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., et al. (1995). The minimal gene complement of Mycoplasma genitalium. Science *270*, 397–403.

Fujita, M., and Losick, R. (2005). Evidence that entry into sporulation in Bacillus subtilis is governed by a gradual increase in the level and activity of the master regulator Spo0A. Genes Dev. *19*, 2236–2244.

Fujita, M., González-Pastor, J.E., and Losick, R. (2005). High- and low-threshold genes in the Spo0A regulon of Bacillus subtilis. J. Bacteriol. *187*, 1357–1368.

Gapes, J.R., Nimcevic, D., and Friedl, A. (1996). Long-Term Continuous Cultivation of Clostridium beijerinckii in a Two-Stage Chemostat with On-Line Solvent Removal. Appl. Environ. Microbiol. *62*, 3210–3219.

Garnier, T., and Cole, S.T. (1986). Characterization of a Bacteriocinogenic Plasmid from Clostridium perfringens and Molecular Genetic Analysis of the Bacteriocin-Encoding Gene. J. Bacteriol. *168*, 1189–1196.

Geider, R.J., Delucia, E.H., Falkowski, P.G., Finzi, A.C., Grime, J.P., Grace, J., Kana, T.M., La Roche, J., Long, S.P., Osborne, B.A., et al. (2001). Primary productivity of planet earth: biological determinants and physical constraints in terrestrial and aquatic habitats. Glob. Chang. Biol. *7*, 849–882.

Ghim, S.-Y., Nielsen, P., and Neuhard, J. (1994). Molecular characterization of pyrimidine biosynthesis genes from the thermophile Bacillus caldolyticus. Microbiology *140*, 479–491.

Gomez, L.D., Steele-King, C.G., and McQueen-Mason, S.J. (2008). Sustainable liquid biofuels from biomass: the writing's on the walls. New Phytol. *178*, 473–485.

Gottschal, J.C., and Morris, J.G. (1982). Continuous production of acetone and butanol by Clostridium acetobuytlicum growing in a turbidostat culture. Biotechnol. Lett. *4*, 477–482.

Gottwald, M., and Gottsehalk, G. (1985). The internal pH of Clostridium acetobutylicum and its effect on the shift from acid to solvent formation. Arch Microbiol *143*, 42–46.

Gravius, B. (1993). Genetic Instability and Strain Degeneration in Streptomyces-Rimosus. Appl. Environ. Microbiol. *59*, 2220–2228.

Green, E.M. (2011). Fermentative production of butanol—the industrial perspective. Curr. Opin. Biotechnol. *22*, 337–343.

Griffin, A.S., West, S.A., and Buckling, A. (2004). Cooperation and competition in pathogenic bacteria. Nature *430*, 1024–1027.

Grimbert, L. (1893). Fermentation anaerobie produite par le Bacillus orthobutylicus. Ses variations sous certaines influences biologiques. Ann. Inst. Pasteur *7*, 353–402.

Grimmler, C., Janssen, H., Krausse, D., Fischer, R.-J., Bahl, H., Dürre, P., Liebl, W., and Ehrenreich, A. (2011). Genome-wide gene expression analysis of the switch between acidogenesis and solventogenesis in continuous cultures of Clostridium acetobutylicum. J. Mol. Microbiol. Biotechnol. *20*, 1–15.

Grimsley, J.K., Tjalkens, R.B., Strauch, M.A., Bird, T.H., Spiegelman, G.B., Hostomsky, Z., Whiteley, J.M., and Hochn, J.A. (1994). Subunit Composition and Domain Structure of the SpoOA Sporulation Transcription Factor of Bacillus subtilis*. J. Biol. Chem. *269*, 16977–16982.

Groisman, E.A. (2016). Feedback Control of Two-Component Regulatory Systems. Annu. Rev. Microbiol *70*, 103–124.

Gu, Y., Jiang, Y., Wu, H., Liu, X., Li, Z., Li, J., Xiao, H., Shen, Z., Dong, H., Yang, Y., et al. (2011). Economical challenges to microbial producers of butanol: Feedstock, butanol ratio and titer. Biotechnol. J. *6*, 1348–1357.

Hall, N. (2007). Advanced sequencing technologies and their wider impact in microbiology. J. Exp. Biol. *210*, 1518–1525.

Hamilton, W.D. (1964). The Genetical Evolution of Social Behaviour. I. Biol *7*, 1–16.

Hamon, M.A., and Lazazzera, B.A. (2002). The sporulation transcription factor Spo0A is required for biofilm development in Bacillus subtilis. Mol. Microbiol. *42*, 1199–1209.

Harris, L.M., Desai, R.P., Welker, N.E., and Papoutsakis, E.T. (2000). Characterization of recombinant strains of the Clostridium acetobutylicum butyrate kinase inactivation mutant: need for new phenomenological models for solventogenesis

and butanol inhibition? Biotechnol. Bioeng. *67*, 1–11.

Harris, L.M., Welker, N.E., and Papoutsakis, E.T. (2002). Northern, morphological, and fermentation analysis of spo0A inactivation and overexpression in Clostridium acetobutylicum ATCC 824. J. Bacteriol. *184*, 3586–3597.

Hartmanis, M.N., Åhlman, H., and Gatenbeck, S. (1986). Stability of solvent formation in Clostridium acetobutylicum during repeated subculturing. Appl. Microbiol. Biotechnol. *23*, 369–371.

Hasman, H., Saputra, D., Sicheritz-Ponten, T., Lund, O., Svendsen, C.A., Frimodt-Møller, N., and Aarestrup, F.M. (2014). Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. J. Clin. Microbiol. *52*, 139–146.

Hayashida, S., and Yoshino, S. (1989). Degeneration of Solventogenic Clostridium Caused by a Defect in NADH Generation. Agric. Biol. Chem. 427–435.

Heap, J.T., Pennington, O.J., Cartman, S.T., Carter, G.P., and Minton, N.P. (2007). The ClosTron: a universal gene knock-out system for the genus Clostridium. J. Microbiol. Methods *70*, 452–464.

Heap, J.T., Pennington, O.J., Cartman, S.T., and Minton, N.P. (2009). A modular system for Clostridium shuttle plasmids. J. Microbiol. Methods *78*, 79–85.

Heap, J.T., Kuehne, S.A., Ehsaan, M., Cartman, S.T., Cooksley, C.M., Scott, J.C., and Minton, N.P. (2010). The ClosTron: Mutagenesis in Clostridium refined and streamlined. J. Microbiol. Methods *80*, 49–55.

Heap, J.T., Ehsaan, M., Cooksley, C.M., Ng, Y.-K., Cartman, S.T., Winzer, K., and Minton, N.P. (2012). Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. Nucleic Acids Res. *40*, e59.

Heap, J.T., Theys, J., Ehsaan, M., Kubiak, A.M., Dubois, L., Paesmans, K., Van Mellaert, L., Knox, R., Kuehne, S.A., Lambin, P., et al. (2014). Spores of Clostridium engineered for clinical efficacy and safety cause regression and cure of tumors in vivo. Oncotarget *5*, 1761–1769.

Hermann, M., Fayolle, F., Marchal, R., Podvin, L., Sebald, M., and Vandecasteele, J.P. (1985). Isolation and characterization of butanol-resistant mutants of Clostridium acetobutylicum. Appl. Environ. Microbiol. *50*, 1238–1243.

Herrero, A.A. (1983). End-product inhibition in anaerobic fermentations. Trends Biotechnol. *1*, 49–53.

Herrero, A.A., Gomez, R.F., Snedecor, B., Tolman, C.J., and Roberts, M.F. (1985). Growth inhibition of Clostridium thermocellum by carboxylic acids: A mechanism based on uncoupling by weak acids. Appl Microbiol Biotechnol *22*, 53–62.

Herrmann, G., Jayamani, E., Mai, G., and Buckel, W. (2008). Energy Conservation via Electron-Transferring Flavoprotein in Anaerobic Bacteria. J. Bacteriol. *190*, 784–791.

Hira, A., and Guilherme De Oliveira, L. (2009). No substitute for oil? How Brazil developed its ethanol industry.

Hobman, J.L. (2007). MerR family transcription activators: similar designs, different specificities. Mol. Microbiol. *63*, 1275–1278.

Hoegh-Guldberg, O., Mumby, P.J., Hooten, A.J., Steneck, R.S., Greenfield, P., Gomez, E., Harvell, C.D., Sale, P.F., Edwards, A.J., Caldeira, K., et al. (2007). Coral Reefs Under Rapid Climate Change and Ocean Acidification. Science (80-.). *318*, 1737–1742.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Leading Edge Review Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell *157*, 1262–1278.

Hughes, L., and Rudolph, J. (2011). Future world oil production: growth, plateau, or peak? Curr. Opin. Environ. Sustain. *3*, 225–234.

Humphreys, J.R., Daniel, R., and Poehlein, A. (2017). Genome Sequence of the Homoacetogenic, Gram-Negative, Endospore-Forming Bacterium Sporomusa acidovorans DSM 3132. Genome Announc. *5*.

Hunter-Cevera, J.C., and Belt, A. (1996). Maintaining cultures for biotechnology and industry (Academic Press).

James Hansen, Makiko Sato, Reto Ruedy, Ken Lo, David W. Lea, and and Martin Medina-Elizade (2000). Global temperature change. Proc. Natl. Acad. Sci. U. S. A. *97*, 1355–1358.

Jang, Y.-S., Young Lee, J., Lee, J., Hwan Park, J., Ae Im, J., Eom, M.-H., Lee, J., Lee, S.-H., Song, H., Cho, J.-H., et al. (2012). Enhanced Butanol Production Obtained by Reinforcing the Direct Butanol-Forming Route in Clostridium acetobutylicum A cetone-butanol-ethanol (ABE) fermentation by clostridia. MBio 14–12.

Jenal, U., and Malone, J. (2006). Mechanisms of Cyclic-di-GMP Signaling in Bacteria. Annu. Rev. Genet. *40*, 385–407.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. *31*, 233–239.

Jiang, Y., Liu, J., Jiang, W., Yang, Y., and Yang, S. (2015). Current status and prospects of industrial bio-production of n-butanol in China. Biotechnol. Adv. *33*, 1493–1501.

Jiao, S., Zhang, Y., Wan, C., Lv, J., Du, R., Zhang, R., and Han, B. (2016). Transcriptional analysis of degenerate strain Clostridium beijerinckii DG-8052 reveals a pleiotropic response to CaCO3-associated recovery of solvent production. Sci. Rep.

6, 38818.

Jin, C., Yao, M., Liu, H., Lee, C.F.F., and Ji, J. (2011). Progress in the production and application of n-butanol as a biofuel. Renew. Sustain. Energy Rev. *15*, 4080–4106.

Jones, D.T., and Woods, D.R. (1986). Acetone-butanol fermentation revisited. Microbiol. Rev. *50*, 484–524.

Jones, D.T., van der Westhuizen, A., Long, S., Allcock, E.R., Reid, S.J., and Woods, D.R. (1982a). Solvent Production and Morphological Changes in Clostridium acetobutylicum. Appl. Environ. Microbiol. *43*, 1434–1439.

Jones, D.T., Van Der Westhuizen, a., and Long, S. (1982b). Solvent production and morphological changes in Clostridium acetobutylicum. Appl. Environ. Microbiol. *43*, 1434–1439.

Jönsson, L.J., Alriksson, B., and Nilvebrant, N.-O. (2013). Bioconversion of lignocellulose: inhibitors and detoxification. Biotechnol. Biofuels *6*, 16.

Joseph, R.C., Kim, N.M., and Sandoval, N.R. (2018). Recent Developments of the Synthetic Biology Toolkit for Clostridium. Front. Microbiol. *9*, 154.

Karberg, M., Guo, H., Zhong, J., Coon, R., Perutka, J., and Lambowitz, A.M. (2001). Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. Nat. Biotechnol. *19*, 1162–1167.

Karl, T.R., and Trenberth, K.E. (2003). Modern Global Climate Change. Science (80-.). *302*, 1719–1723.

Kashket, E.R., and Cao, Z.Y. (1993). Isolation of a degeneration-resistant mutant of Clostridium acetobutylicum NCIMB 8052. Appl. Environ. Microbiol. *59*, 4198–4202.

Kashket, E.R., and Cao, Z.Y. (1995). Clostridial strain degeneration. In FEMS

Microbiology Reviews, pp. 307–315.

Kim, M., Huang, T., and Miller, J.H. (2003). Competition between MutY and Mismatch Repair at A C Mispairs In Vivo. J. Bacteriol. *185*, 4626–4629.

Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B.A., and Blanch, H.W. (2012). The challenge of enzyme cost in the production of lignocellulosic biofuels. Biotechnol. Bioeng. *109*, 1083–1087.

Klemm, E., and Dougan, G. (2016). Advances in Understanding Bacterial Pathogenesis Gained from Whole-Genome Sequencing and Phylogenetics. Cell Host Microbe *19*, 599–610.

Kök, M.S. (2015). An integrated approach: advances in the use of *Clostridium* for biofuel. Biotechnol. Genet. Eng. Rev. *31*, 69–81.

Köpke, M., Noack, S., Dürre, P., Michael, K., Steffi, N., and Peter, D. (2011). The Past, Present, and Future of Biofuels – Biobutanol as Promising Alternative. Biofuel Prod. Dev. Prospect. 451–486.

Kosaka, T., Nakayama, S., Nakaya, K., Yoshino, S., and Furukawa, K. (2007). Characterization of the sol operon in butanol-hyperproducing Clostridium saccharoperbutylacetonicum strain N1-4 and its degeneration mechanism. Biosci. Biotechnol. Biochem. *71*, 58–68.

Kotte, A.-K., Severn, O., Bean, Z., Schwarz, K., Minton, N.P., and Winzer, and K. (2017). RNPP-type quorum sensing regulates solvent formation and sporulation in Clostridium acetobutylicum. BioRxiv *44*.

Kuehne, S.A., Heap, J.T., Cooksley, C.M., Cartman, S.T., and Minton, N.P. (2012). ClosTron-mediated engineering of clostridium. Methods Mol. Biol. *765*, 389–407.

Kutzenok, a., and Aschner, M. (1952). Degenerative processes in a strain of Clostridium butylicum. J. Bacteriol. *64*, 829–836.

Lambowitz, A.M., and Zimmerly, S. (2011). Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb. Perspect. Biol. *3*, a003616.

Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M.R., Ahn, T.-H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., et al. (2015). Insights from 20 years of bacterial genome sequencing. Funct. Integr. Genomics *15*, 141–161.

Laub, M.T. (2011). The Role of Two-Component Signal Transduction Systems in Bacterial Stress Responses. Bact. Stress Responses 45–58.

Laursen, R.A., and Westheimer, F.H. (1966). The Active Site of Acetoacetate Decarboxylase 1. J. Am. Chem. Soc. *88*, 3426–3430.

Lawson, P.A., Llop-Perez, P., Hutson, R.A., Hippe, H., and Collins, M.D. (1993). Towards a phylogeny of the clostridia based on 16S rRNA sequences. FEMS Microbiol. Lett. *113*, 87–92.

Layec, S., Decaris, B., and Leblond-Bourget, N. (2007). Characterization of proteins belonging to the CHAP-related superfamily within the Firmicutes. J. Mol. Microbiol. Biotechnol. *14*, 31–40.

Lee, H., Popodi, E., Tang, H., and Foster, P.L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. Proc. Natl. Acad. Sci. *109*, E2774–E2783.

Lee, S.M., Cho, M.O., Park, C.H., Chung, Y.C., Kim, J.H., Sang, B.I., and Um, Y. (2008a). Continuous butanol production using suspended and immobilized Clostridium beijerinckii NCIMB 8052 with supplementary butyrate. Energy and Fuels *22*, 3459–3464.

Lee, S.M., Cho, M.O., Um, Y., and Sang, B.I. (2010). Development of real-time PCR primer and probe sets for detecting degenerated and non-degenerated forms of the butanol-producing bacterium clostridium acetobutylicum ATCC 824. Appl. Biochem. Biotechnol. *161*, 75–83.

Lee, S.Y., Bennett, G.N., and Papoutsakis, E.T. (1992a). Construction ofEscherichia coli-Clostridium acetobutylicum shuttle vectors and transformation ofClostridium acetobutylicum strains. Biotechnol. Lett. *14*, 427–432.

Lee, S.Y., Mermelstein, L.D., Bennett, G.N., and Papoutsakis, E.T. (1992b). Vector construction, transformation, and gene amplification in Clostridium acetobutylicum ATCC 824. Ann. N. Y. Acad. Sci. *665*, 39–51.

Lee, S.Y., Park, J.H., Jang, S.H., Nielsen, L.K., Kim, J., and Jung, K.S. (2008b). Fermentative butanol production by clostridia. Biotechnol. Bioeng. *101*, 209–228.

Lehmann, L., and Keller, L. (2006). The evolution of cooperation and altruism – a general framework and a classification of models. J. Evol. Biol. *19*, 1365–1376.

Lewis, R.J., Krzywda, S., Brannigan, J.A., Turkenburg, J.P., Muchová, K., Dodson, E.J., Barák, I., and Wilkinson, A.J. (2000). The trans-activation domain of the sporulation response regulator Spo0A revealed by X-ray crystallography. Mol. Microbiol. *38*, 198–212.

Li, Q., Chen, J., Minton, N.P., Zhang, Y., Wen, Z., Liu, J., Yang, H., Zeng, Z., Ren, X., Yang, J., et al. (2016). CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. Biotechnol. J. *11*, 961–972.

Liew, F., Henstra, A.M., Köpke, M., Winzer, K., Simpson, S.D., and Minton, N.P. (2017). Metabolic engineering of Clostridium autoethanogenum for selective alcohol

production. Metab. Eng. *40*, 104–114.

Lin, Y.L., and Blaschek, H.P. (1983). Butanol Production by a Butanol-Tolerant Strain of Clostridium acetobutylicum in Extruded Corn Broth. Appl. Environ. Microbiol. *45*, 966–973.

Liu, H., Bouillaut, L., Sonenshein, A.L., and Melville, S.B. (2013). Use of a mariner-based transposon mutagenesis system to isolate Clostridium perfringens mutants deficient in gliding motility. J. Bacteriol. *195*, 629–636.

Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., Lahrmann, U., Zhao, Q., Zheng, Y., Zhao, Y., et al. (2015a). IBS: an illustrator for the presentation and visualization of biological sequences: Fig. 1. Bioinformatics *31*, 3359–3361.

Liu, Z., Qiao, K., Tian, L., Zhang, Q., Liu, Z.-Y., and Li, F.-L. (2015b). Spontaneous large-scale autolysis in Clostridium acetobutylicum contributes to generation of more spores. Front. Microbiol. *6*, 950.

Loman, N.J., Constantinidou, C., Chan, J.Z.M., Halachev, M., Sergeant, M., Penn, C.W., Robinson, E.R., and Pallen, M.J. (2012). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat. Rev. Microbiol. *10*, 599–606.

Long, S., Jones, D.T., and Woods, D.R. (1984). The relationship between sporulation and solvent production in Clostridium acetobuytlicum P262. Biotechnol. Lett. *6*, 529–534.

Longley, D.B., Harkin, D.P., and Johnston, P.G. (2003). 5-Fluorouracil: mechanisms of action and clinical strategies. Nat. Rev. Cancer *3*, 330–338.

López-Contreras, A.M., Gabor, K., Martens, A.A., Renckens, B.A.M., Claassen, P.A.M., Van Der Oost, J., and De Vos, W.M. (2004). Substrate-induced production and

secretion of cellulases by Clostridium acetobutylicum. Appl. Environ. Microbiol. *70*, 5238–5243.

Losick, R., and Stragier, P. (1992). Crisscross regulation of cell-type-specific gene expression during development in B. subtilis. Nature *355*, 601–604.

Lu, C., Dong, J., and Yang, S.-T. (2013). Butanol production from wood pulping hydrolysate in an integrated fermentation–gas stripping process. Bioresour. Technol. *143*, 467–475.

Ludwig, W., Schleifer, K.-H., and Whitman, W.B. (2009). Revised road map to the phylum Firmicutes. In Systematic Bacteriology, (New York, NY: Springer New York), pp. 1–13.

Lütke-Eversloh, T. (2014). Application of new metabolic engineering tools for Clostridium acetobutylicum. Appl. Microbiol. Biotechnol. *98*, 5823–5837.

Lütke-Eversloh, T., and Bahl, H. (2011). Metabolic engineering of Clostridium acetobutylicum: recent advances to improve butanol production. Curr. Opin. Biotechnol. *22*, 634–647.

Lv, J., Jiao, S., Du, R., Zhang, R., Zhang, Y., and Han, B. (2016). Proteomic analysis to elucidate degeneration of Clostridium beijerinckii NCIMB 8052 and role of Ca2+ in strain recovery from degeneration. J. Ind. Microbiol. Biotechnol. *43*, 741–750.

Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C.R., Dopman, E.B., Dickinson, W.J., Okamoto, K., Kulkarni, S., Hartl, D.L., et al. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc. Natl. Acad. Sci. *105*, 9272–9277.

Ma, K., Qiu, L., Mrasek, K., Zhang, J., Liehr, T., Quintana, L.G., and Li, Z. (2012). Common fragile sites: genomic hotspots of DNA damage and carcinogenesis. Int. J. Mol. Sci. *13*, 11974–11999.

Mabee, W.E., Gregg, D.J., and Saddler, J.N. (2005). Assessing the Emerging Biorefinery Sector in Canada. *121*, 765–778.

Mabee, W.E., McFarlane, P.N., and Saddler, J.N. (2011). Biomass availability for lignocellulosic ethanol production. Biomass and Bioenergy *35*, 4519–4529.

Maddox, I.S., Steiner, E., Hirsch, S., Wessner, S., Gutierrez, N.A., Gapes, J.R., and Schuster, K.C. (2000). The cause of “acid-crash” and “acidogenic fermentations” during the batch acetone-butanol-ethanol (ABE-) fermentation process. J. Mol. Microbiol. Biotechnol. *2*, 95–100.

Madhusudan, Zapf, J., Whiteley, J.M., Hoch, J.A., Xuong, N.H., and Varughese, K.I. (1996). Crystal structure of a phosphatase-resistant mutant of sporulation response regulator Spo0F from Bacillus subtilis. Structure *4*, 679–690.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. *45*, D200–D203.

Mardis, E.R. (2013). Next-Generation Sequencing Platforms. Annu. Rev. Anal. Chem. *6*, 287–303.

Mardis, E.R. (2017). DNA sequencing technologies: 2006–2016. Nat. Protoc. *12*, 213–218.

Maxon, W.D. (1955). Continuous fermentation; a discussion of its principles and applications. Appl. Microbiol. *3*, 110–120.

MDowell, P., Affas, Z., Reynolds, C., Holden, M.T.G., Wood, S.J., Saint, S., Cockayne, A., Hill, P.J., Dodd, C.E.R., Bycroft, B.W., et al. (2001). Structure, activity and evolution of the group I thiolactone peptide quorum-sensing system of

Staphylococcus aureus. Mol. Microbiol. *41*, 503–512.

Mermelstein, L.D., and Papoutsakis, E.T. (1993). In vivo methylation in Escherichia coli by the Bacillus subtilis phage phi 3T I methyltransferase to protect plasmids from restriction upon transformation of Clostridium acetobutylicum ATCC 824. Appl. Environ. Microbiol. *59*, 1077–1081.

Mermelstein, L.D., Welker, N.E., Bennett, G.N., and Papoutsakis, E.T. (1992). Expression of cloned homologous fermentative genes in Clostridium acetobutylicum ATCC 824. Biotechnology. (N. Y). *10*, 190–195.

Metzker, M.L. (2005). Emerging technologies in DNA sequencing. Genome Res. *15*, 1767–1776.

Meyer, J.-M., Stintzi, A., De Vos, D., Cornelis, P., Taraz, K., and Budzikiewicz, H. (1997). Use of siderophores to type pseudomonads: the three Pseudornonas aeruginosa pyoverdine systems. Microbiology *143*, 35–43.

Mitchell, D. (2008). A Note on Rising Food Prices. World *World Bank*, 21.

Mohr, A., and Raman, S. (2013). Lessons from first generation biofuels and implications for the sustainability appraisal of second generation biofuels. Energy Policy *63*, 114–122.

Mohr, S.C., Sokolov, N. V, He, C.M., and Setlow, P. (1991). Binding of small acid-soluble spore proteins from Bacillus subtilis changes the conformation of DNA from B to A. Proc. Natl. Acad. Sci. U. S. A. *88*, 77–81.

Müller, S., Harms, H., and Bley, T. (2010). Origin and analysis of microbial population heterogeneity in bioprocesses. Curr. Opin. Biotechnol. *21*, 100–113.

Münchau, A., and Bhatia, K.P. (2000). Uses of botulinum toxin injection in medicine

today. BMJ *320*, 161–165.

Nagaraju, S., Davies, N.K., Walker, D.J.F., Köpke, M., and Simpson, S.D. (2016). Genome editing of Clostridium autoethanogenum using CRISPR/Cas9. Biotechnol. Biofuels *9*, 219.

Naik, S.N., Goud, V. V., Rout, P.K., and Dalai, A.K. (2010). Production of first and second generation biofuels: A comprehensive review. Renew. Sustain. Energy Rev. *14*, 578–597.

Ng, Y.K., Ehsaan, M., Philip, S., Collery, M.M., Janoir, C., Collignon, A., Cartman, S.T., and Minton, N.P. (2013). Expanding the Repertoire of Gene Tools for Precise Manipulation of the Clostridium difficile Genome: Allelic Exchange Using pyrE Alleles. PLoS One *8*, e56051.

Ni, Y., and Sun, Z. (2009). Recent progress on industrial fermentative production of acetone–butanol–ethanol by Clostridium acetobutylicum in China. Appl. Microbiol. Biotechnol. *83*, 415–423.

Nölling, J., Breton, G., Omelchenko, M. V, Makarova, K.S., Zeng, Q., Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J., et al. (2001). Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum. J. Bacteriol. *183*, 4823–4838.

Oh, J.-H., and van Pijkeren, J.-P. (2014). CRISPR–Cas9-assisted recombineering in Lactobacillus reuteri. Nucleic Acids Res. *42*, e131–e131.

Ossowski, S., Schneeberger, K., Lucas-Lledo, J.I., Warthmann, N., Clark, R.M., Shaw, R.G., Weigel, D., and Lynch, M. (2010). The Rate and Molecular Spectrum of Spontaneous Mutations in Arabidopsis thaliana. Science (80-.). *327*, 92–94.

Oultram, J.D., Loughlin, M., Swinfield, T.-J., Brehm, J.K., Thompson, D.E., and Minton,

N.P. (1988). Introduction of plasmids into whole cells of Clostridium acetobutylicum by electroporation. FEMS Microbiol. Lett. *56*, 83–88.

Ourisson, G., Albrecht, P., and Rohmer, M. (1984). The Microbial Origin of Fossil Fuels. Sci. Am. *251*, 44–51.

Palmer, R.J., Kazmerzak, K., Hansen, M.C., and Kolenbrander, P.E. (2001). Mutualism versus Independence: Strategies of Mixed-Species Oral Biofilms In Vitro Using Saliva as the Sole Nutrient Source. Infect. Immun. *69*, 5794–5804.

Papoutsakis, E.T. (2008). Engineering solventogenic clostridia. Curr. Opin. Biotechnol. *19*, 420–429.

Paredes, C.J., Alsaker, K. V., and Papoutsakis, E.T. (2005). A comparative genomic view of clostridial sporulation and physiology. Nat. Rev. Microbiol. *3*, 969–978.

Parkinson, J.S., and Kofoid, E.C. (1992). Communication Modules In Bacterial Signalling Proteins.

Pauly, M., and Keegstra, K. (2008). Cell-wall carbohydrates and their modification as a resource for biofuels. Plant J. *54*, 559–568.

Perego, M. (1998). Kinase–phosphatase competition regulates Bacillus subtilis development. Trends Microbiol. *6*, 366–370.

Perego, M., Spiegelman, G.B., and Hoch, J.A. (1988). Structure of the gene for the transition state regulator, abrB: regulator synthesis is controlled by the spo0A sporulation gene in Bacillus subtilis. Mol. Microbiol. *2*, 689–699.

Pfromm, P.H., Amanor-Boadu, V., Nelson, R., Vadlani, P., and Madl, R. (2010). Bio-butanol vs. bio-ethanol: A technical and economic assessment for corn and switchgrass fermented by yeast or Clostridium acetobutylicum. Biomass and

Bioenergy *34*, 515–524.

Piggot, P.J., and Hilbert, D.W. (2004). Sporulation of Bacillus subtilis. Curr. Opin. Microbiol. *7*, 579–586.

Piggot, P.J., and Losick, R. (2002). Sporulation Genes and Intercompartmental Regulation. In Bacillus Subtilis and Its Closest Relatives, (American Society of Microbiology), pp. 483–517.

Poehlein, A., Solano, J.D.M., Flitsch, S.K., Krabben, P., Winzer, K., Reid, S.J., Jones, D.T., Green, E., Minton, N.P., Daniel, R., et al. (2017). Microbial solvent formation revisited by comparative genome analysis. Biotechnol. Biofuels *10*, 58.

Poxton, I.R., McCoubrey, J., and Blair, G. (2001). The pathogenicity of Clostridium difficile. Clin. Microbiol. Infect. *7*, 421–427.

Puppán, D. (2002). Environmental evaluation of biofuels. Period. Polytech. Soc. Manag. Sci. *10*, 95–116.

Pyne, M.E., Bruder, M., Moo-Young, M., Chung, D.A., and Chou, C.P. (2014). Technical guide for genetic advancement of underdeveloped and intractable Clostridium. Biotechnol. Adv. *32*, 623–641.

Pyne, M.E., Bruder, M.R., Moo-Young, M., Chung, D.A., and Chou, C.P. (2016). Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. Sci. Rep. *6*, 25666.

Qureshi, N., and Blaschek, H.P. (2001). Recent advances in ABE fermentation: hyper-butanol producing Clostridium beijerinckii BA101. J. Ind. Microbiol. Biotechnol. *27*, 287–291.

Qureshi, N., and Maddox, I.S. (1995). Continuous Production of Acetone-Butanol-

Ethanol Using Immobilized Cells of Clostridium acetobutylicum and Integration with Product Removal by Liquid-Liquid Extraction.

R. W. O'Brien & J. G. Morris (1971). Oxygen and the Growth and Metabolism of Clostidium acetobutylicum. *68*, 307–318.

Ragsdale, S.W., and Pierce, E. (2008). Acetogenesis and the Wood–Ljungdahl pathway of CO2 fixation. Biochim. Biophys. Acta - Proteins Proteomics *1784*, 1873–1898.

Ravagnani, A., Jennert, K.C.B., Steiner, E., Grunberg, R., Jefferies, J.R., Wilkinson, S.R., Young, D.I., Tidswell, E.C., Brown, D.P., Youngman, P., et al. (2000). Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. Mol. Microbiol. *37*, 1172–1185.

Reuter, J.A., Spacek, D. V., and Snyder, M.P. (2015). High-Throughput Sequencing Technologies. Mol. Cell *58*, 586–597.

Reysenbach, A.L., Ravenscroft, N., Long, S., Jones, D.T., and Woods, D.R. (1986). Characterization, Biosynthesis, and Regulation of Granulose in Clostridium acetobutylicum. Appl. Environ. Microbiol. *52*, 185–190.

Riley, M.A., and Gordon, D. (1999). The ecological role of bacteriocins in bacterial competition. Trends Microbiol. *3*, 129–133.

Rood, J.I., and Cole, S.T. (1991). Molecular genetics and pathogenesis of Clostridium perfringens. Microbiol. Rev. *55*, 621–648.

Ross-Gillespie, A., Gardner, A., West, S., and Griffin, A. (2007). Frequency Dependence and Cooperation: Theory and a Test with Bacteria. Am. Nat. *170*, 331.

Rugbjerg, P., Myling-Petersen, N., Porse, A., Sarup-Lytzen, K., and Sommer, M.O.A.

(2018). Diverse genetic error modes constrain large-scale bio-based production. Nat. Commun. *9*.

Sabathé, F., Croux, C., Cornillot, E., and Soucaille, P. (2002). amyP, a reporter gene to study strain degeneration in Clostridium acetobutylicum ATCC 824. FEMS Microbiol. Lett. *210*, 93–98.

Sachs, J.L., Mueller, U.G., Wilcox, T.P., and Bull, J.J. (2004). The evolution of cooperation. Q. Rev. Biol. *79*, 135–160.

Saini, J.K., Saini, R., and Tewari, L. (2015). Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. 3 Biotech *5*, 337–353.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. *74*, 5463–5467.

Sauer, U. (1995). Differential induction of genes related to solvent formation during the shift from acidogenesis to solventogenesis in continuous culture of Clostridium acetobutylicum. FEMS Microbiol. Lett. *125*, 115–120.

Sauer, U., Treuner, A., Buchholz, M., Santangelo, J.D., and Dürre, P. (1994). Sporulation and primary sigma factor homologous genes in Clostridium acetobutylicum. J. Bacteriol. *176*, 6572–6582.

Schmeisser, F., Brannigan, J.A., Lewis, R.J., Wilkinson, A.J., Youngman, P., and Barák, I. (2000). A new mutation in spo0A with intragenic suppressors in the effector domain. FEMS Microbiol. Lett. *185*, 123–128.

Schmidhuber, J. (2007). Biofuels : An emerging threat to Europe ' s Food Security ? Biofuels : An emerging threat to Europe ' s Food Security ? Impact of an increased biomass use on.

Schuster, K.C., Goodacre, R., Gapes, J.R., and Young, M. (2001). Degeneration of solventogenic Clostridium strains monitored by Fourier transform infrared spectroscopy of bacterial cells. J. Ind. Microbiol. Biotechnol. *27*, 314–321.

Scotcher, M.C., Rudolph, F.B., and Bennett, G.N. (2005). Expression of abrB310 and SinR, and effects of decreased abrB310 expression on the transition from acidogenesis to solventogenesis, in Clostridium acetobutylicum ATCC 824. Appl. Environ. Microbiol. *71*, 1987–1995.

Searchinger, T., Heimlich, R., Houghton, R.A., Dong, F., Elobeid, A., Fabiosa, J., Tokgoz, S., Hayes, D., and Yu, T.-H. (2008). Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land-use change. Science *319*, 1238–1240.

Setlow, P. (2006). Spores of Bacillus subtilis: Their resistance to and killing by radiation, heat and chemicals. In Journal of Applied Microbiology, pp. 514–525.

Shafiee, S., and Topal, E. (2009). When will fossil fuel reserves be diminished? Energy Policy *37*, 181–189.

Shao, L., Hu, S., Yang, Y., Gu, Y., Chen, J., Yang, Y., Jiang, W., and Yang, S. (2007). Targeted gene disruption by use of a group II intron (targetron) vector in Clostridium acetobutylicum. Cell Res. *1791*.

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. Nat. Biotechnol. *26*, 1135–1145.

Shi, Z., and Blaschek, H.P. (2008). Transcriptional analysis of Clostridium beijerinckii NCIMB 8052 and the hyper-butanol-producing mutant BA101 during the shift from acidogenesis to solventogenesis. Appl. Environ. Microbiol. *74*, 7709–7714.

Sims, R.E.H., Mabee, W., Saddler, J.N., and Taylor, M. (2010). An overview of second

generation biofuel technologies. Bioresour. Technol. *101*, 1570–1580.

Solá, M., Gomis-Rüth, F.X., Serrano, L., González, A., and Coll, M. (1999). Three-dimensional crystal structure of the transcription factor PhoB receiver domain. J. Mol. Biol. *285*, 675–687.

Solomon, S., Plattner, G.-K., Knutti, R., and Friedlingstein, P. (2009). Irreversible climate change due to carbon dioxide emissions. Proc. Natl. Acad. Sci. U. S. A. *106*, 1704–1709.

Sorek, R., Lawrence, C.M., and Wiedenheft, B. (2013). CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea. Annu. Rev. Biochem. *82*, 237–266.

Stanbury, P.F., Whitaker, A., Hall, S.J., Stanbury, P.F., Whitaker, A., and Hall, S.J. (2017). Culture preservation and inoculum development. Princ. Ferment. Technol. 335–399.

Steiner, E., Dago, A.E., Young, D.I., Heap, J.T., Minton, N.P., Hoch, J.A., and Young, M. (2011). Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in Clostridium acetobutylicum. Mol. Microbiol. *80*, 641–654.

Steiner, E., Scott, J., Minton, N.P., and Winzer, K. (2012). An agr quorum sensing system that regulates granulose formation and sporulation in Clostridium acetobutylicum. Appl. Environ. Microbiol. *78*, 1113–1122.

Stephens, G.M., Holt, R.A., Gottschall, J.C., and Morris, J.G. (1985). Studies on the stability of solvent production by Clostridium acetobutylicum in continuous culture. J. Appl. Bacteriol. *58*, 597–605.

Stock, J.B., Ninfa, A.J., and Stock, A.M. (1989). Protein phosphorylation and regulation of adaptive responses in bacteria. Microbiol. Rev. *53*, 450–490.

Strauch, M.A., and Hoch, J.A. (1993). Signal transduction in Bacillus subtilis sporulation. Curr. Opin. Genet. Dev. *3*, 203–212.

Strauch, M., Webbt, V., Spiegelmant, G., and Hoch, J.A. (1990). The SpoOA protein of Bacillus subtilis is a repressor of the abrB gene (footprinting/transcription/sporulation). Genetics *87*, 1801–1805.

Strauch, M.A., Perego, M., Burbulys, D., and Hoch, J.A. (1989). The transition state transcription regulator AbrB of Bacillus subtilis is autoregulated during vegetative growth. Mol. Microbiol. *3*, 1203–1209.

Szulczyk, K.R. (2010). Which is a better transportation fuel–butanol or ethanol? Int J Energy Env. *1*, 501–512.

Tamang, J.P., Watanabe, K., and Holzapfel, W.H. (2016). Review: Diversity of Microorganisms in Global Fermented Foods and Beverages. Front. Microbiol. *7*, 377.

Taylor, B.L., and Zhulin, I.B. (1999). PAS Domains: Internal Sensors of Oxygen, Redox Potential, and Light. Microbiol. Mol. Biol. Rev. *63*, 479–506.

Tomas, C.A., Alsaker, K. V, Bonarius, H.P.J., Hendriksen, W.T., Yang, H., Beamish, J.A., Paredes, C.J., and Papoutsakis, E.T. (2003). DNA array-based transcriptional analysis of asporogenous, nonsolventogenic Clostridium acetobutylicum strains SKO1 and M5. J. Bacteriol. *185*, 4539–4547.

Traag, B.A., Pugliese, A., Eisen, J.A., and Losick, R. (2013). Gene conservation among endospore-forming bacteria reveals additional sporulation genes in Bacillus subtilis. J. Bacteriol. *195*, 253–260.

Tracy, B.P., Jones, S.W., Fast, A.G., Indurthi, D.C., and Papoutsakis, E.T. (2012). Clostridia: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. Curr. Opin. Biotechnol. *23*, 364–381.

Tummala, S.B., Welker, N.E., and Papoutsakis, E.T. (2003a). Design of antisense RNA constructs for downregulation of the acetone formation pathway of Clostridium acetobutylicum. J. Bacteriol. *185*, 1923–1934.

Tummala, S.B., Junne, S.G., and Papoutsakis, E.T. (2003b). Antisense RNA downregulation of coenzyme A transferase combined with alcohol-aldehyde dehydrogenase overexpression leads to predominantly alcohologenic Clostridium acetobutylicum fermentations. J. Bacteriol. *185*, 3644–3653.

Valerie Mercer-Blackman, Hossein Samiei, and K.C. (2007). IMF Survey: Biofuel Demand Pushes Up Food Prices.

Vane, L.M. (2008). Separation technologies for the recovery and dehydration of alcohols from fermentation broths. Biofuels, Bioprod. Biorefining *2*, 553–588.

Visick, K.L., Foster, J., Doino, J., Mcfall-Ngai, M., and Ruby, E.G. (2000). Vibrio fischeri lux Genes Play an Important Role in Colonization and Development of the Host Light Organ. J. Bacteriol. *182*, 4578–4586.

Vollherbst-Schneck, K., Sands, J.A., and Montenecourt, B.S. (1984). Effect of butanol on lipid composition and fluidity of Clostridium acetobutylicum ATCC 824. Appl. Environ. Microbiol. *47*, 193–194.

Volz, K. (1993). Structural conservation in the CheY superfamily. Biochemistry *32*, 11741–11753.

Wackett, L.P. (2015). Specialty chemicals from microbes. Microb. Biotechnol. *8*, 614–615.

Wang, S., Dong, S., Wang, P., Tao, Y., and Wang, Y. (2017). Genome Editing in Clostridium saccharoperbutylacetonicum N1-4 with the CRISPR-Cas9 System. Appl. Environ. Microbiol. *83*.

Wang, S., Hong, W., Dong, S., Zhang, Z.-T., Zhang, J., Wang, L., and Wang, Y. (2018). Genome engineering of Clostridium difficile using the CRISPR-Cas9 system. Clin. Microbiol. Infect.

Wang, Y., Li, X., Mao, Y., and Blaschek, H.P. (2012). Genome-wide dynamic transcriptional profiling in Clostridium beijerinckii NCIMB 8052 using single-nucleotide resolution RNA-Seq. BMC Genomics *13*, 102.

Wang, Y., Zhang, Z.-T., Seo, S.-O., Choi, K., Lu, T., Jin, Y.-S., and Blaschek, H.P. (2015). Markerless chromosomal gene deletion in Clostridium beijerinckii using CRISPR/Cas9 system. J. Biotechnol. *200*, 1–5.

Wang, Y., Zhang, Z.T., Seo, S.O., Lynn, P., Lu, T., Jin, Y.S., and Blaschek, H.P. (2016). Bacterial Genome Editing with CRISPR-Cas9: Deletion, Integration, Single Nucleotide Modification, and Desirable "clean" Mutant Selection in Clostridium beijerinckii as an Example. ACS Synth. Biol. *5*, 721–732.

Webb, J.S., Givskov, M., and Kjelleberg, S. (2003). Bacterial biofilms: prokaryotic adventures in multicellularity. Curr. Opin. Microbiol. *6*, 578–585.

West, A.H., and Stock, A.M. (2001). Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem. Sci. *26*, 369–376.

West, S.A., Griffin, A.S., Gardner, A., and Diggle, S.P. (2006). Social evolution theory for microorganisms. Nat. Rev. Microbiol. *4*, 597–607.

West, S.A., Diggle, S.P., Buckling, A., Gardner, A., and Griffin, A.S. (2007a). The Social Lives of Microbes. Annu. Rev. Ecol. Evol. Syst. *38*, 53–77.

West, S.A., Griffin, A.S., and Gardner, A. (2007b). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J. Evol. Biol. *20*, 415–432.

Wilkinson, S.R., and Young, M. (1994). Targeted integration of genes into the Clostridium acetobutylicum chromosome. Microbiology *140*, 89–95.

Wilkinson, S.R., Young, D.I., Morris, J.G., and Young, M. (1995). Molecular genetics and the initiation of solventogenesis in Clostridium beijerinckii (formerly Clostridium acetobutylicum) NCIMB 8052. FEMS Microbiol. Rev. *17*, 275–285.

Willson, B.J., Kovács, K., Wilding-Steele, T., Markus, R., Winzer, K., and Minton, N.P. (2016). Production of a functional cell wall-anchored minicellulosome by recombinant Clostridium acetobutylicum ATCC 824. Biotechnol. Biofuels *9*, 109.

Wolanin, P.M., Webre, D.J., and Stock, J.B. (2003). Mechanism of Phosphatase Activity in the Chemotaxis Response Regulator CheY. Biochemistry *42*, 4075–14082.

Woods, D.R., and Jones, D.T. (1987). Physiological Responses of Bacteroides and Clostridium Strains to Environmental Stress Factors. Adv. Microb. Physiol. *28*, 1–64.

Woolley, R.C., and Morris, J.G. (1990). Stability of Solvent Production by Clostridium-Acetobutylicum in Continuous Culture - Strain Differences. J. Appl. Bacteriol. *69*, 718–728.

Xue, Q., Yang, Y., Chen, J., Chen, L., Yang, S., Jiang, W., and Gu, Y. (2016). Roles of three AbrBs in regulating two-phase Clostridium acetobutylicum fermentation. Appl. Microbiol. Biotechnol. *100*, 9081–9089.

Yamaguchi, Y., and Inouye, M. (2009). Chapter 12 mRNA Interferases, Sequence-Specific Endoribonucleases from the Toxin–Antitoxin Systems. In Progress in Molecular Biology and Translational Science, pp. 467–500.

Yang, X., and Tsao, G.T. (1995). Enhanced acetone-butanol fermentation using repeated fed-batch operation coupled with cell recycle by membrane and simultaneous removal of inhibitory products by adsorption. Biotechnol. Bioeng. *47*,

444–450.

Ye, Z., Song, J., Zhu, E., Song, X., Chen, X., and Hong, X. (2018). Alginate Adsorbent Immobilization Technique Promotes Biobutanol Production by Clostridium acetobutylicum Under Extreme Condition of High Concentration of Organic Solvent. Front. Microbiol. *9*, 1071.

Yutin, N., and Galperin, M.Y. (2013). A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. Environ. Microbiol. *15*, 2631–2641.

Zeikus, J.G. (1980). Chemical and fuel production by anaerobic bacteria. Ann. Rev. Microbiol *34*, 423–464.

Zeisler, J., Kleinhappl, M., and Hofbauer, H. (2010). Reliable sampling of impurities in product gas and syngas.

Zhang, Y., Grosse-Honebrink, A., and Minton, N.P. (2015). A universal mariner transposon system for forward genetic studies in the genus clostridium. PLoS One.

Zhang, Y., Jiao, S., Lv, J., Du, R., Yan, X., Wan, C., Zhang, R., and Han, B. (2017). Sigma Factor Regulated Cellular Response in a Non-solvent Producing Clostridium beijerinckii Degenerated Strain: A Comparative Transcriptome Analysis. Front. Microbiol. *8*, 23.

Zhong, J., Karberg, M., and Lambowitz, A.M. (2003). Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. Nucleic Acids Res. *31*, 1656–1664.

7. Appendix

7.1. Vector maps for *pyrE* knock out and *pyrE* repair

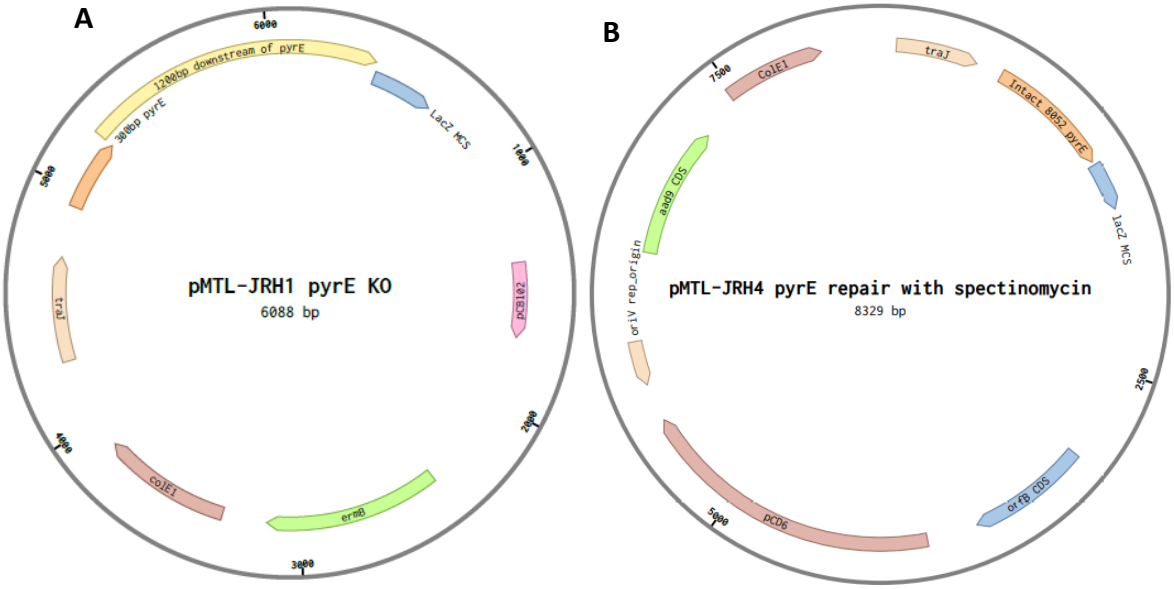


Figure A.1 pMTL-JRH1 *pyrE* knock out vector (A) and pMTL-JRH4 *pyrE* repair vector (B) maps.

7.2. Statistical analysis of wild type compared to isolated degenerates

7.2.1. 2 day solvents and acids

	Acetone		Ethanol		Butanol		Acetate		Butyrate	
	Significant	P value	Significant	P value	Significant	P value	Significant	P value	Significant	P value
WT vs. RD1	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. RD2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. RD3	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. RD4	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. RD5	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. RD1-1	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. RD1-2	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. RD2-1	Yes	****	No	ns	Yes	****	Yes	****	Yes	****

WT vs. RD2-2	Yes	*	No	ns	No	ns	No	ns	No	ns
WT vs. RD3-1	Yes	*	No	ns	No	ns	No	ns	No	ns
WT vs. RD3-2	Yes	**	No	ns	No	ns	No	ns	No	ns
WT vs. RD4-1	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. RD4-2	Yes	*	No	ns	No	ns	No	ns	No	ns
WT vs. RD5-1	No	ns	No	ns	Yes	**	No	ns	No	ns
WT vs. RD5-2	Yes	*	No	ns	Yes	*	No	ns	No	ns
WT vs. RD6-1	No	ns	No	ns	Yes	***	No	ns	No	ns
WT vs. RD6-2	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. DCOG1	Yes	**	No	ns	Yes	****	No	ns	Yes	****
WT vs. DCOG2	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. DCOG3	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG4	No	ns	No	ns	Yes	*	No	ns	No	ns
WT vs. DCOG5	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. DCOG1-1	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. DCOG1-2	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. DCOG2-1	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. DCOG2-2	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. DCOG3-1	No	ns	No	ns	No	ns	No	ns	Yes	*
WT vs. DCOG3-2	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. DCOG4-1	Yes	*	No	ns	No	ns	No	ns	No	ns
WT vs. DCOG4-2	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. DCOG5-1	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. DCOG5-2	Yes	*	No	ns	No	ns	No	ns	No	ns
WT vs. DCOG6-1	Yes	*	No	ns	No	ns	No	ns	No	ns
WT vs. DCOG6-2	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. CIC1	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC3	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC4	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. CIC5	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC6	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. CIC1-1	No	ns	No	ns	Yes	*	No	ns	No	ns
WT vs. CIC1-2	No	ns	No	ns	Yes	*	No	ns	No	ns
WT vs. CIC2-1	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. CIC2-2	No	ns	No	ns	Yes	****	Yes	*	Yes	****
WT vs. CIC3-1	No	ns	No	ns	Yes	*	No	ns	No	ns
WT vs. CIC3-2	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. CIC4-1	No	ns	No	ns	Yes	****	Yes	**	Yes	****
WT vs. CIC4-2	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. CIC5-1	Yes	*	No	ns	No	ns	No	ns	No	ns
WT vs. CIC5-2	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. CIC6-1	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. CIC6-2	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. FW1	No	ns	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW2	Yes	***	No	ns	Yes	****	No	ns	Yes	****

WT vs. FW3	Yes	***	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW4	No	ns	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW5	No	ns	No	ns	Yes	****	No	ns	Yes	*
WT vs. FW6	Yes	***	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW7	Yes	***	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW1-1	No	ns	No	ns	No	ns	Yes	****	Yes	****
WT vs. FW1-2	Yes	***	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW2-1	Yes	***	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW2-2	Yes	***	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW3-1	Yes	*	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW3-2	Yes	***	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW4-1	Yes	***	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW4-2	No	ns	No	ns	Yes	*	No	ns	Yes	*
WT vs. FW5-1	Yes	***	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW5-2	Yes	**	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW6-1	Yes	***	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW6-2	Yes	**	No	ns	Yes	****	No	ns	Yes	****

7.2.2. 5 day solvents and acids

	Acetone		Ethanol		Butanol		Acetate		Butyrate	
	Significant	P value	Significant	P value	Significant	P value	Significant	P value	Significant	P value
WT vs. RD1	Yes	**	No	ns	Yes	****	No	ns	No	ns
WT vs. RD2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. RD3	No	ns	No	ns	Yes	**	No	ns	No	ns
WT vs. RD4	No	ns	No	ns	Yes	***	No	ns	No	ns
WT vs. RD5	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. RD1-1	No	ns	No	ns	Yes	**	No	ns	No	ns
WT vs. RD1-2	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. RD2-1	Yes	****	No	ns	Yes	****	Yes	****	Yes	****
WT vs. RD2-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. RD3-1	No	ns	No	ns	Yes	*	No	ns	No	ns
WT vs. RD3-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. RD4-1	No	ns	No	ns	Yes	**	No	ns	No	ns
WT vs. RD4-2	No	ns	No	ns	Yes	***	No	ns	No	ns
WT vs. RD5-1	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. RD5-2	No	ns	No	ns	Yes	**	No	ns	No	ns
WT vs. RD6-1	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. RD6-2	No	ns	No	ns	Yes	**	No	ns	No	ns
WT vs. DCOG1	Yes	**	No	ns	Yes	****	No	ns	No	ns

WT vs. DCOG2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG3	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG4	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG5	No	ns	No	ns	Yes	*	No	ns	No	ns
WT vs. DCOG1-1	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG1-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG2-1	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG2-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG3-1	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG3-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG4-1	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG4-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG5-1	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG5-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG6-1	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. DCOG6-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC1	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC3	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC4	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC5	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC6	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC1-1	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC1-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC2-1	No	ns	No	ns	Yes	*	No	ns	No	ns
WT vs. CIC2-2	No	ns	No	ns	Yes	****	Yes	**	Yes	***
WT vs. CIC3-1	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC3-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC4-1	Yes	*	No	ns	Yes	****	Yes	**	Yes	**
WT vs. CIC4-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC5-1	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC5-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. CIC6-1	No	ns	No	ns	No	ns	No	ns	No	ns
WT vs. CIC6-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. FW1	Yes	*	No	ns	Yes	****	No	ns	No	ns
WT vs. FW2	Yes	****	No	ns	Yes	****	No	ns	No	ns
WT vs. FW3	Yes	****	No	ns	Yes	****	No	ns	No	ns
WT vs. FW4	Yes	**	No	ns	Yes	****	No	ns	No	ns
WT vs. FW5	Yes	**	No	ns	Yes	****	No	ns	No	ns
WT vs. FW6	Yes	****	No	ns	Yes	****	No	ns	No	ns
WT vs. FW7	Yes	****	No	ns	Yes	****	No	ns	No	ns
WT vs. FW1-1	No	ns	No	ns	Yes	****	Yes	**	Yes	**
WT vs. FW1-2	Yes	****	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW2-1	Yes	****	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW2-2	Yes	****	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW3-1	Yes	***	No	ns	Yes	****	No	ns	Yes	***

WT vs. FW3-2	Yes	****	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW4-1	Yes	****	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW4-2	Yes	**	No	ns	Yes	****	No	ns	Yes	***
WT vs. FW5-1	Yes	****	No	ns	Yes	****	No	ns	Yes	****
WT vs. FW5-2	No	ns	No	ns	Yes	****	No	ns	No	ns
WT vs. FW6-1	Yes	*	No	ns	Yes	****	No	ns	No	ns
WT vs. FW6-2	No	ns	No	ns	Yes	****	No	ns	No	ns

7.3. End point statistics for the 1:10 and 1:100 subculturing

7.3.1. 1:10 regime

	Acetone		Ethanol		Butanol		Acetate		Butyrate	
	Significant	P value	Significant	P value	Significant	P value	Significant	P value	Significant	P value
End vs start	Yes	****	No	ns	Yes	****	No	ns	Yes	****

7.3.2. 1:100 regime

	Acetone		Ethanol		Butanol		Acetate		Butyrate	
	Significant	P value	Significant	P value	Significant	P value	Significant	P value	Significant	P value
End vs start	No	ns	No	ns	No	ns	No	ns	No	ns

7.4. Fermentation statistics for the ClosTron mutants and complements

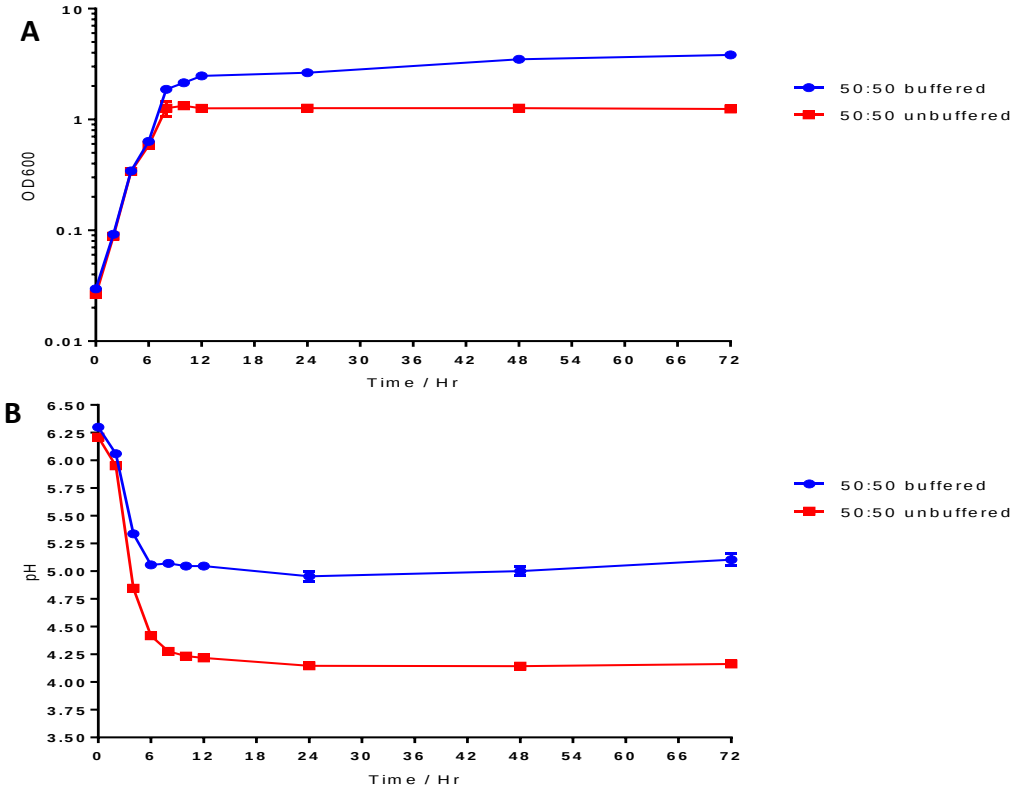
	Acetone		Ethanol		Butanol		Acetate		Butyrate	
	Significant	P value	Significant	P value	Significant	P value	Significant	P value	Significant	P value
P5 vs. Cbei_0017	No	ns	No	ns	No	ns	No	ns	No	ns
P5 vs. Cbei_0017_comp	No	ns	No	ns	No	ns	No	ns	No	ns
P5 vs. Cbei_3078	No	ns	No	ns	No	ns	No	ns	No	ns
P5 vs. Cbei_3078:_comp	No	ns	No	ns	No	ns	No	ns	No	ns
P5 vs. spo0A	Yes	****	No	ns	Yes	****	No	ns	Yes	****
P5 vs. spo0A_comp	No	ns	No	ns	No	ns	No	ns	No	ns

Cbei_0017 vs Comp	No	ns	No	ns	No	ns	No	ns	No	ns
Cbei_3078 vs Comp	No	ns	No	ns	No	ns	No	ns	No	ns
spo0A vs Comp	Yes	****	No	ns	Yes	****	No	ns	Yes	****

7.5. End point statistics for the WT/FW7 monoculture

	OD		pH		CFU	
	Significant	P value	Significant	P value	Significant	P value
WT buffered vs. WT unbuffered	Yes	**	No	ns	No	ns
WT buffered vs. FW buffered	Yes	****	No	ns	Yes	****
WT buffered vs. FW unbuffered	Yes	****	Yes	****	Yes	****
WT unbuffered vs. FW buffered	Yes	*	No	ns	Yes	****
WT unbuffered vs. FW unbuffered	Yes	***	Yes	**	Yes	****
FW buffered vs. FW unbuffered	No	ns	Yes	****	No	ns

7.6. 50:50 mixed culture pilot experiment



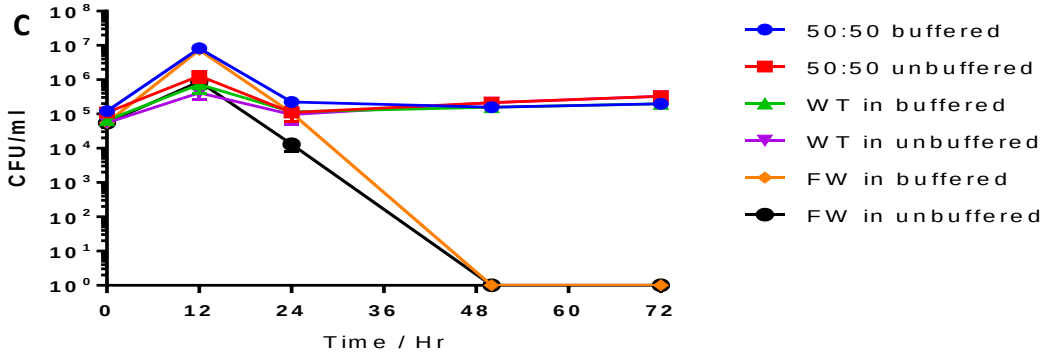


Figure A.2 Mixed culture (50:50) growth of wild type and FW7 mutant grown in CBM-S 6% glucose with and without $CaC0_3$ buffering.

Growth measured as increase in optical density (OD_{600}) (A). pH (B) and CFU/ml (C) were also measured. No detectable CFU/ml was given a value of 1 (10^0) as 0 cannot be plotted on a log scale. Technical replicas represent standard error of the mean. *n* = 3.