The role of Agr quorum sensing on sporulation and solvent formation in *Clostridium beijerinckii*

Lauren Boak, BSc

Thesis submitted to the University of Nottingham for the

degree of Doctor of Philosophy

November 2023

Declaration:

Unless otherwise acknowledged, the work presented in the thesis titled 'The role of Agr quorum sensing on sporulation and solvent formation in *Clostridium beijerinckii*' is my own. No part has been submitted for another degree in the university of Nottingham or any other institute of learning.

Lauren Boak

November 2023

Abstract:

In recent decades there has been a growing demand for the development of biofuels, in particular bioethanol and biobutanol. Such solvents can be produced via acetone-butanol-ethanol (ABE) fermentation in solventogenic *Clostridium* species. For solventogenic *Clostridium*, it has been shown that the switch from acidogenesis to solventogenesis and sporulation along with improved ABE productivity occurs at high cell density in batch culture, such as industrial fermentations, suggesting quorum sensing is involved. Whilst regulation of metabolism and sporulation via QS has been investigated in other clostridia, its role remains unclear in the industrially important species *Clostridium beijerinckii.*

Studies into the *C. beijerinckii* NCIMB 8052 genome sequence suggested that the bacterium is capable of Agr-type QS, a QS system best studied in *Staphylococcus aureus*. Here, the *agrD* gene encodes the precursor for a signal molecule, the autoinducing peptide (AIP); *agrB* encodes the membrane protein which processes the AgrD precursor into the mature AIP, *agrC* encodes the membrane histidine kinase receptor which senses the AIP, and *agrA* encodes a dedicated response regulator. Accumulation of the AIP serves as a measure of population density and at a given threshold concentration triggers a response via the AgrC-AgrA two-component system. The aim of this study was to investigate and further understand if Agr QS has a role in

the regulation of stationary phase phenotypes, primarily sporulation and solvent production, in *C. beijerinckii.*

An initial study into the genomes of various *C. beijerinckii* strains identified highly conserved Agr loci in this species. Through CRISPS-Cas9 based genome editing the *agrB* genes of the three most highly conserved systems, denoted as Agr2, Agr4 and Agr5, in the *C. beijerinckii* NCIMB 8052 genome were inactivated. Phenotypic characterisation showed a significant reduction in heat-resistant endospore formation in the *agr2* and *agr4* mutants, as well as reduced glucose consumption and solvent formation in all three mutants, however no observed change in granulose accumulation when compared to the WT. RNA-seq data comparing mutants to the WT, identified in the *agrB2* and *agrB4* mutants, later stage sporulation genes were hugely downregulated with *spoIVB* and *sigK* genes showing to be a key point of regulatory change. Various solvent related genes were downregulated in the mutants also, and alternative *agr* loci were shown to be differentially regulated in each of the *agrB* mutants, showing the multiple *agr* loci present in one genome influence the regulation of each other.

These findings show that specific *agr* loci do hold a regulatory role in the stationary phase phenotypes sporulation and solvent formation in *C. beijerinckii* NCIMB 8052. However, the depth of each *agr* loci and its role in such processes varies, with a complex *agr* hierarchical system being possible.

Acknowledgements:

Firstly, I would like to thank my primary supervisor, Dr. Klaus Winzer for providing me with this opportunity and for his exceptional guidance, support, and supervision throughout the course of this PhD. I am forever grateful for the experience and knowledge I have acquired under your mentorship as well as the delicious cakes you would bring in. I would also like to thank my secondary supervisor Dr Ying Zhang for her help and support throughout. An extended thank you also to the BDI staff, including Dr David Tooth for his expertise with sample analysis and Shelly for her patience and always being there to answer any question I had no matter how small it was.

Thank you to the B50 office, Liam, Louise, Margaux, Danielle, Temi, Natalia, Bilgesu and Tom for their invaluable support and hilarious company, making the past few years so enjoyable.

Thank you to my friends Willa, Tarry, Cheddar, Ollie, Toby and Pete for keeping me sane throughout with the pub trips and games nights. Also, my housemate over the past 3 years Sarah for her company, particularly through COVID. A huge thank you to my partner Luke for giving up most of your weekends to travel to see me and provide me with your endless encouragement and bizarre yet fun company. A further thank you to my family, particularly my mum Karen for her unwavering support, for whom without, none of this would have been possible.

Table of Contents:

List of Figures:

List of Tables:

List of abbreviations:

µg	micro gram
µl	micro litre
µm	micro metre
µM	micro molar
ABC	ATP-binding cassette
ABE	Acetone-Butanol-Ethanol fermentation
ACE	Allelic coupled exchange
ADP	Adenosine diphosphate
Agr	Accessory gene regulator
AHL	Acyl-homoserine lactone
AI-2	Autoinducer-2
AIP	Autoinducing peptide
ATCC	American type culture collection
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
bp	Base pair
Cas	CRISPR associated
CBM	Clostridial basal medium

CBP	Consolidated bioprocessing
cDNA	complementary DNA
CFU	Colony forming units
CGM	Clostridial growth medium
CoA	Coenzyme A
CRISPR	Clustered regularly interspaced short palindromic repeats
DBD	DNA-binding domain
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EPB	Electroporation buffer
Erm	Erythromycin
FOA	5-Fluoroorotate
g	Gravity
GC	Gas chromatography
gDNA	Genomic DNA
GHG	Greenhouse gases
HDR	Homology directed repair
HK	Histidine kinase
HPLC	High pressure liquid chromatography

IEP	Intron encoded protein
kJ	kilo Joules
KO	Knock out
kV	kilo Volts
LB	Lysogeny broth
LHA	Left homology arm
mg	milli gram
ml	milli litre
mM	milli molar
mm	Millimetre
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
NHEJ	Non-homologous end joining
nm	nano metre
OD	Optical density

ORF	Open reading frame
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
QS	Quorum sensing
RAM	Retrotransposition-activated marker
RHA	Right homology arm
RNA	Ribonucleic acid
RPM	Rotations per minute
rRNA	Ribosomal ribonucleic acid
SAM	S-adenosylmethionine
sgRNA	single-stranded guide RNA
TCS	Two-component system
TEM	Transmission electron microscope
UV	Ultraviolet
V	Volt
WT	wild-type

Chapter 1

Introduction

1.1 Fossil fuels and Global warming

Fossil fuels are non-renewable energy sources developed over millions of years from the remains of microbial, animal and plant organic matter unable to decompose (Ourisson, Albrecht and Rohmer, 1984). Key examples include natural gases, coal and crude oil which together provide around 80% of the world's energy, crude oil being the current largest source of energy globally (Anadón, 2014; BP, 2019).

Energy derived from fossil fuels emits huge amounts of pollutants including nitrogen oxide and sulphur dioxide, and greenhouse gases (GHG) like methane and carbon dioxide (CO_2) (Anadón, 2014). The burning of fossil fuels accounts for a staggering 90% of all CO_2 emissions from human activity (Jackson *et al.*, 2019). Globally, fossil fuel CO_2 emissions have risen every decade since the 1960s, rising a further 2.1% in 2018 and expected to increase in ongoing years (Friedlingstein *et al.*, 2019; BP, 2019). Fossil fuel CO_2 emissions are the upmost serious cause of climate change with average global temperatures already increased by 1.1°C above preindustrial levels. Predictions of warming beyond 1.5°C and potentially 2°C are thought to be met within two decades (Jackson *et al.*, 2019; Masson-Delmotte *et al.*, 2019). The effects of such warming have caused extreme events such as loss of glaciers and snowpack, heavy rainfall and flooding, increased hurricane intensity, heatwaves, wildfires and droughts

(Solomon *et al.*, 2009; Masson-Delmotte *et al.*, 2019; Herring *et al.*, 2018).

Furthermore, fossil fuels are finite. Coal, oil and gas reserves are estimated to be depleted within 132, 50 and 51 years respectively according to BP Statistical Review of World Energy (2019). This along with the GHG effect from fossil-based fuels calls for the need for alternative energy sources.

1.2 Renewable Energy- Biofuels

The BP Statistical Review 2019 showed a global increase in demand for energy by 2.9% in one year, with other sources suggesting an increase in demand by 26% by 2040 (IEA, 2019). To meet such growing demands and reduce the implications of fossil fuel energy, sustainable alternatives have become of increasing interest. Unlike fossil fuels, which will be depleted within decades, renewable sources such as wind, solar, hydropower, geothermal and biomass are regenerative. In 2018, renewable energy output increased by 14.5% pushing up its share in the power sector to 9.3%, biofuel production alone grew by 9.7% globally, showing growing interest (BP, 2019).

Biofuels are an energy source obtained from biomass, organic matter derived from living organisms such as woody plant waste, microalgae, food waste and dried animal waste. They can be categorised into multiple generations. First-generation biofuels are processed from

starch rich food crops like sugar-beet, sugarcane and corn (Sharma, Xu and Qin, 2019). Examples include biodiesel, bioethanol and biogas with characteristic ability of being able to blend with traditional fuel types (Naik *et al.*, 2010). Bioethanol is generated through microbial fermentation of sugars from crops, biodiesel through transesterification of plant- based oils and animal fats, and biogas through anaerobic digestion of liquid-based manure and other organic materials (Naik *et al.*, 2010; Lee and Lavoie, 2013).

Nevertheless, a major issue with first-generation biofuels is the food versus fuel dilemma, diverting the use of land from growing crops for fuel instead of food. Alternatively, second-generation biofuels are generated largely from lignocellulosic material, regarded as non-edible and cheap plant material (Lee and Lavoie, 2013). Woody crops, forest residues and dedicated non-food energy crops can be grown on land ineligible for food crops, eradicating the food versus fuel dilemma. Further advantages of secondary biofuel production include much cheaper production costs and higher abundance of lignocellulosic biomass compared to the starch-rich crops used in primary biofuel production (Lee and Lavoie, 2013). Furthermore, whilst there is still concern for GHG emissions and environmental impacts of first-generation biofuels, second-generation biofuels are expected to largely reduce CO_2 emissions. For example, Brazilian bioethanol produced from sugarcane bagasse can reduce up to 70% CO_2 emissions compared to petrol (SHELL, 2020).

Energy from lignocellulosic biomass can be generated through the thermo-chemical route, which involves direct combustion, gasification, liquefaction and pyrolysis to produce syngas, a mixture of H_2, CO and CO_2. Syngas can be used to produce biofuels such as ethanol and synthetic diesel via Fischer-Tropsch conversion or be a substrate in microbial fermentations (Lee and Lavoie, 2013). Alternatively, the biological route, based on enzymatic and microbial processing can be used to first release sugars from lignocellulosic biomass and then ferment these into alcohols such as bioethanol and biobutanol (Naik *et al.*, 2010; Sims *et al.*, 2010; Kumar and Kumar, 2017).

Lignocellulosic biomass is made up of three main components, cellulose, hemicellulose and lignin (Hendriks and Zeeman, 2009). The biological route of processing lignocellulosic biomass involves four distinctive steps. During the first step, called pre-treatment, the hemicellulose and cellulose components are broken down to release sugars. This process can be done in a variety of different ways. Physical methods include grinding, ultra-sonication and microwaving. Physico- chemical techniques such as steam treatment, hot water or ammonium fiber expansion. Chemical treatment using alkali, acids and organosolv treatments. And finally, through enzymatic hydrolysis, which involves the use of enzymes to degrade polysaccharide sugars into easily fermentable hexose and pentose monosaccharide sugars. It is typical to use a combination of the above methods (Chaturvedi and Verma, 2013; Birgen *et al.*, 2019).

The second step involves detoxification of inhibitory chemicals generated during pre-treatment. Chemicals such as formic acid, acetic acid, phenolics, furan aldehydes and salts are inhibitory to microorganisms and thus must be removed before the fermentation process. Common detoxification steps such as electrodialysis, liming and resin treatments are effective (Jönsson, Alriksson and Nilvebrant, 2013; Birgen *et al.*, 2019).

The third step involves the utilisation of generated sugars and their fermentative conversion into solvents. This is carried out under anaerobic conditions by solvent producing microorganisms including, but not limited to, ethanol fermentation by *Saccharomyces cerevisiae* yeast and acetone-butanol-ethanol (ABE) fermentation by solvent-producing *Clostridium* bacterial species (Birgen *et al.*, 2019).

The final step involves downstream processing to obtain the fermentation product from the culture. Separation techniques include liquid-liquid extraction, distillation, adsorption and membrane extraction (Yakovlev *et al.*, 2013).

This route may have more opportunities to reduce processing costs compared to thermochemical processing, for example through recycling enzymes (Jørgensen and Pinelo, 2017). Alternatively, consolidated bioprocessing (CBP), which can achieve lignocellulosic breakdown and microbial fermentation in the same step (S. Li *et al.*, 2020). A study by Xin *et al.* (2017) found that *Clostridium* sp. strain NJP7 could achieve direct butanol formation from hemicellulose in CBP with saccharification. Whilst ethanol production has been hugely explored,

the development of second-generation biofuels has given rise to other biofuels which could have more potential, one of interest being butanol (Li *et al.*, 2020).

1.2.1 Butanol

Butanol is a carbon-4 alcohol with multiple isomers which have different chemical properties. The isomer n-butanol presents the best properties for use as a fuel and avoids some disadvantages, listed below, that smaller chain alcohols have such as methanol and ethanol, which are currently more established (Trindade and Santos, 2017). Butanol for instance has a longer carbon chain allowing for higher heating values and energy densities, meaning engines would have lower fuel consumption whilst running on n-butanol. Furthermore, n-butanol is less volatile, polar and corrosive compared to ethanol meaning it has reduced ignition problems and is easier to transport via pipelines (Trindade and Santos, 2017; Xue and Cheng, 2019). Finally, diesel engines can run on high butanol to diesel fuel blends. A study by Campos-Fernández *et al.* (2012) reported that 30% butanol-diesel fuel blends could replace 100% diesel fuel, without any modifications to the existing system or loss of performance. Later studies showed higher percentage blends, up to 75% butanol, can also meet diesel fuel standards (Lapuerta *et al.*, 2017).

Currently, commercially available butanol is produced mainly through a propylene oxo synthesis petrochemical pathway, largely impacted by the price of crude oil. The biological production of butanol has therefore become increasingly attractive (S. Li *et al.*, 2020). The *Clostridium* species was first documented by Louis Pasteur in 1862, however it was only in 1914, when a chemist called Weizmann, isolated a *Clostridium* species which could produce yields of butanol and acetone (Jones and Woods, 1986).

1.3 The genus *Clostridium*

The first proposal for the genus *Clostridium* was made in 1880 by Prazmowski. For a long time, only four criteria had to be met for a microorganism to be placed in this genus. These were a Gram-positive cell wall, formation of endospores, strictly anaerobic and non-sulphate reducing (Andreesen, Bahl and Gottschalk, 1989). This resulted in a diversified genus with different species showing a range of characteristics. It was only in the 1970's, by molecular analysis, that the huge diversity of the genus was shown (Johnson and Francis, 1975). Current classification is now largely reliant on the 16s rRNA gene sequences. Since then, classification and genus descriptions of the *Clostridium* genus has been amended by multiple studies (Collins *et al.*, 1994; Lawson and Rainey, 2016; Cruz-Morales *et al.*, 2019). It has been suggested that true *Clostridium* be classified into the cluster I

sensu stricto group and therefore novel species that fall outside of this group should not be included in the genus (Lawson and Rainey, 2016). The genus contains both human pathogens, examples being *Clostridium tetani*, *Clostridium botulinum* and *Clostridium perfringens*, as well as industrial species.

There currently appear to be several key industrial clostridial species of interest, *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Clostridium saccaroperbutylacetonicum, Clostridium saccharoacetobutylicum* and *Clostridium pasteurianum* (Keis et al., 1995; Poehlein et al., 2017; Jones *et al.*, 2023). Whilst C*. acetobutylicum* is the model ABE producer, *C. beijerinckii* shows higher tolerance to process inhibitors such as pH changes and mixed sugars as well as a broader substrate range, making it more advantageous for lignocellulosic biomass fermentations, demonstrated by multiple studies on a variety of lignocellulosic material such as apple pomace (Jin et al., 2019), wheat bran (Liu et al., 2010), sweet potato hydrolysates (Zuleta-Correa, Chinn and Bruno-Bárcena, 2022) and even non-pretreated spent coffee grounds (López-Linares et al., 2021).

1.3.1 Clostridial sporulation

As previously described by Diallo 2021, sporulation in clostridial species can be broken down into different stages (see **Figure 1.1**). Stage I consists of DNA replication and positioning of the Z-rings in preparation

of asymmetric division, the accumulation of granulose causes the cell to swell. Stage II involves septum formation at one cell pole to create the forespore and the mother cell compartments. Once asymmetric division is complete, stage III occurs which involves engulfment of the forespore to form two membranes within the mother cell. The spore coat and cortex are formed during stages IV and V which are then followed by endospore maturation, lysis of the mother cell and spore release in stages VI and VII. The Clostridial endospore is made up of five layers surrounding the spore to create a higher level of protection against environmental factors such as heat, oxygen, enzymes, and various chemicals.

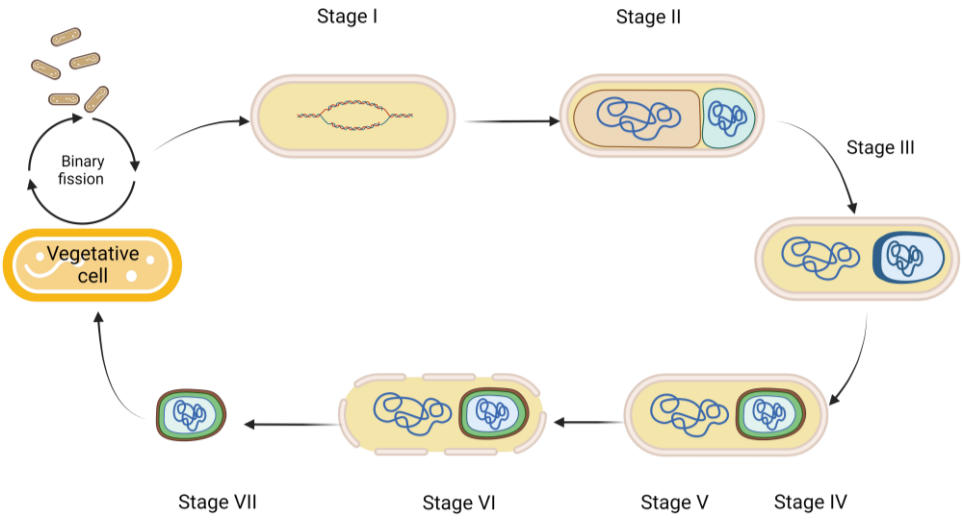


Figure 1.1 Sporulation in Clostridium species (Adapted from Diallo 2021).

Stage I DNA replication and positioning of the Z-rings in preparation of asymmetric division. Stage II asymmetric division. Stage III Engulfment of the forespore. Stages IV and V The spore coat and cortex are formed. Followed by endospore maturation, lysis of the mother cell and spore release in stages VI and VII. Image created with BioRender.com.

1.3.2 Clostridial strain degeneration

Strain degeneration is the term used to describe reduced ability to produce solvents and spores because of repeated subculturing clostridia. Degeneration is thought to occur because of excessive acidification during exponential growth (Kashket and Cao, 1995). Degenerate phenotypes in *C. beijerinckii* NCIMB 8052 have been characterized by Humphreys (2019) and Humphreys *et al.* (2023), who found that the accumulation of degenerate mutants within *C. beijerinckii* cultures was linked to the loss of Spo0A activity, through direct *spo0A* gene mutation or mutations in loci thought to be involved in Spo0A regulation. This study also identified several colony morphologies to be associated with the degenerate phenotype which were round and dark (RD), dark centre with outgrowths (DCOG), caved in centre (CIC) and the flat and white (FW). It was found that whilst the RD, DCOG and CIC types were found to produce similar concentrations of solvents to the wild-type (WT) after 48 hrs of growth, the FW produced significantly less spores, solvents and reduced granulose whilst having an increased acid production compared to the WT. Prevention of degeneration was observed in this study by factors such as reduced growth rate, later inoculation times and preventing the loss of *spo0A* functionality by coupling the gene to an essential gene.

1.4 ABE Fermentation Overview

1.4.1 History of ABE Fermentation

In the early 20th century, ABE fermentation using *Clostridium* was used for mass production of the biofuel biobutanol and acetone. However, in the 1960's fermentation was unable to economically compete with chemical solvent production and was left behind (Jones and Woods, 1986; Jones *et al.*, 2023). Recent interest in using biobutanol as a biofuel, due to world energy and environmental crisis, has driven the desire to develop a more economically efficient and competitive process in the ABE fermentation process in solventogenic *Clostridia* (Xue and Cheng, 2019; Jones *et al.*, 2023). Nevertheless, production of biobutanol through ABE fermentation in solventogenic *Clostridium* is limited to lower titres, which rarely reach beyond 20 g/L in laboratory research (Ezeji, Qureshi and Blaschek, 2004). This is due to factors including butanol toxicity resulting in increased sporulation and degeneration of clostridial cells at higher titers (Maddox, 1989). Butanol tolerance in fermentations is usually limited to 15 g/l, with higher titres leading to damage in the cell membranes (Jones *et al.*, 2023). Clostridia-based fermentations also tend to have low cell densities which may be due to acid inhibition or uncharacterised quorum sensing (QS) mechanisms (Papoutsakis, 2008). Furthermore, high substrate costs and formation of by-products contribute to reduced economic

feasibility. The recent interest in biobutanol has driven research into genetic and metabolic manipulation to improve strain stability and process modifications to optimise culture conditions (Nguyen *et al.*, 2018; Yang *et al.*, 2018; Li *et al.*, 2020). However, there is still poor understanding of physiological processing occurring in clostridia and relatively few genetic engineering tools which are highly efficient available (S. Li *et al.*, 2020).

1.4.2 ABE Metabolism

Solventogenic species have similar ABE metabolic pathways involving three main stages of growth, the acidogenic, solventogenic and sporogenic phases. The acidogenic phase is the primary phase occurring during exponential growth and is characterised by highly motile, dividing, thin rods (Jones *et al.*, 1982). Solventogenic *Clostridia* can metabolise pentose and hexose sugars during the acidogenic phase, this results in the production and accumulation of hydrogen, CO_2 and acids like butyrate and acetate, causing a reduced pH of the surrounding media (Jones and Woods, 1986). The metabolic pathways involved in the production of butyrate and acetate also release adenosine triphosphate (ATP) and are therefore important in energy metabolism. As the culture progresses into stationary phase, rapid cell

death from the acidic conditions is prevented as the low pH triggers the transition into solventogenic phase, where butyrate and acetate are reassimilated. These are transformed into the end products acetone, butanol and ethanol, at a ratio of 3:6:1 in the case of *C. acetobutylicum* (Jones and Woods, 1986). Acetone can be further processed into isopropanol in some *C. beijerinckii* strains, a solvent found to be of increased value compared to acetone partially because it can be used in biofuel blends alongside butanol and ethanol (Heap *et al.*, 2012; Dürre, 2014). See a summary of the ABE metabolic pathway in **Figure 1.2** (Lütke-Eversloh, 2014; Patakova *et al.*, 2019).

The solventogenic phase is associated with physiological changes, most notably change in cellular shape, changing from thin rod-shaped to swollen forms with an extracellular capsule (Jones *et al.*, 1982). Upon entering the stationary phase, the onset of sporulation and granulose formation occurs whilst solvent production ceases, this is called the sporogenic phase (Jones and Woods, 1986; S. Li *et al.*, 2020). During this phase, heat-resistant endospores form. These are metabolically inactive structures highly resistant to many factors including oxygen, heat, desiccation, ultraviolet, toxic chemicals and extreme pH (Abel-Santos, 2014). The accumulation of the polysaccharide storage compound granulose appears to occur just before sporulation, during which it may act as an energy source (Reysenbach *et al.*, 1986).

From an industrial point of view, it would be desirable to develop a strain with the following characteristics: improved tolerance to high acid

concentrations to allow prolonged cell growth and higher cell densities. This would increase yields of acids and theoretically increase yields of solvents at later stages. Also, a non-spore forming strain, as this would prevent the cessation of solvent formation, since the cells stop producing solvents once they have sporulated (S. Li *et al.*, 2020).

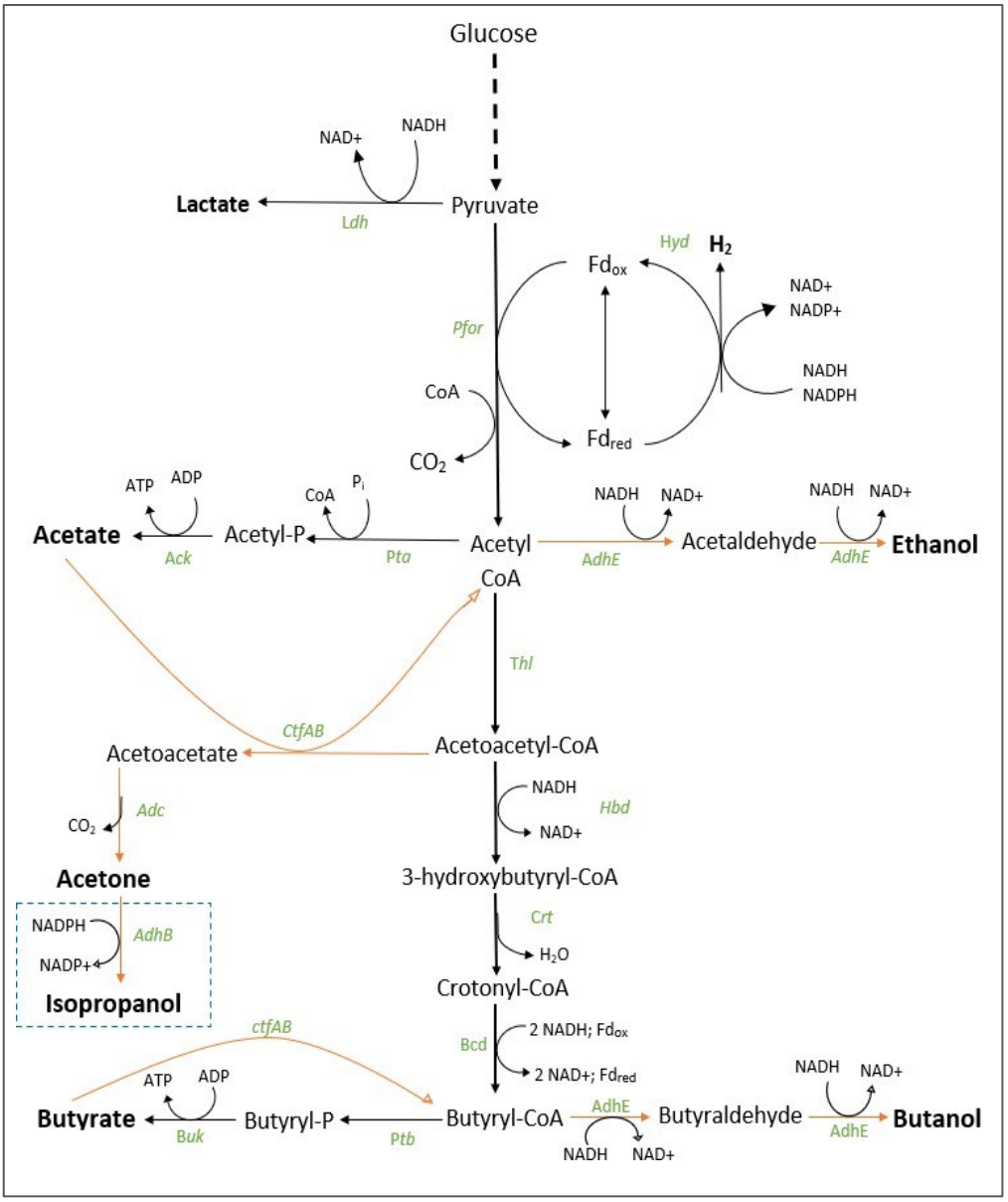


Figure 1.2 Acetone Butanol Ethanol (ABE) fermentation metabolic pathway in *C. acetobutylicum* and *C. beijerinckii*.

ABE metabolism in *C. acetobutylicum* and *C. beijerinckii* adapted from Patakova et al., 2019 and Lütke-Eversloh, 2014. Green labels indicate enzymes: Ldh, Lactate dehydrogenase; Hyd, Hydrogenase; Pfor, pyruvate:ferredoxin oxidoreductase; AdhE, Aldehyde/alcohol dehydrogenase; Pta, Phosphotransacetylase; Ack, Acetate kinase; Thl, Thiolase; CtfAB, CoA transferase; Adc, Acetoacetate decarboxylase; AdhB, Alcohol dehydrogenase; Hbd, 3-hydroxybutyryl-CoA dehydrogenase; Crt, Crotonase; Bcd, Butyryl-CoA dehydrogenase; Ptb, Phosphotransbutyrylase; Buk, Butyrate kinase. Isopropanol production, in the blue dashed box, occurs in some *C. beijerinckii* strains but not *C. acetobutylicum*. Orange arrows indicate stages occurring in solventogenesis.

1.4.3 Regulation of Spore Formation

The regulation of sporulation is well documented in model organisms such as *Bacillus subtilis*. The transcription factor Spo0A is the master regulator of sporulation which works in coordination with a vast number of regulatory elements. In *B. subtilis,* Spo0A must first be phosphorylated into its activated form by a multi-component phosphorelay system (Burbulys, Trach and Hoch, 1991). In this system, various sensor histidine kinases transfer a phosphoryl group to the response regulator Spo0F. In turn the Spo0F-P then transfers a phosphoryl group to the phosphotransferase protein Spo0B which is responsible for phosphorylating Spo0A into its activated form Spo0A-P (Burbulys, Trach and Hoch, 1991). The Spo0A-P can be deactivated by proteins which dephosphorylate Spo0F-P and Spo0B-P such as Rap and Spo0E family phosphatases (Perego *et al.*, 1994; Ohlsen, Grimsley and Hoch, 1994; Piggot and Hilbert, 2004).

However, for *Clostridia* there are some differences. During batch fermentations*, Bacillus subtilis* sporulation is initiated by nutrient starvation, yet in saccharolytic *Clostridia*, like the solventogenic types, sporulation occurs under nutrient excess (Gottschal and Morris, 1981; Long, Jones and Woods, 1984). Furthermore, whilst *Clostridia* also have the Spo0A master regulator, they do not encode a phosphorelay system involving Spo0F/B, like that found in *Bacillus*. In *C.*

acetobutylicum, it appears that two pathways involving orphan histidine kinases (HKs) are instead responsible for both the direct phosphorylation and dephosphorylation of Spo0A. One pathway dependent on HK Cac903 and HK Cac3319, whilst the other is dependent on HK Cac0323 (Steiner *et al.*, 2011). Recently, two orphan histidine kinases, Cbei4484 and Cbei2073, from *C. beijerinckii* have been shown to have high sequence similarity to the three mentioned from *C. acetobutylicum.* Mutation in these histidine kinases, using CRISPR-nCas9 also resulted in reduced sporulation and therefore increased butanol production in *C. beijerinckii,* further characterisation of this pathway is needed to understand fundamental applications (Xin *et al.*, 2020). It is also evident that external factors also influence the regulation of the Spo0A transcription factor. A study by Patakova *et al.* (2021) demonstrated how butanol shock caused the downregulation of Spo0A and most of its associated sigma factors.

Activated Spo0A-P regulates the expression of sporulation genes and other target transition state genes by binding to “0A boxes” (TGNCGAA) positioned at the 5’ promoter regions. Genes are upregulated upon Spo0A-P binding if the OA box is upstream of the promoter, or downregulated if the OA box is overlapping or downstream of the promoter (Strauch *et al.*, 1990; Steiner *et al.*, 2011).

1.4.4 Regulatory Elements of Solventogenesis

During the switch from acidogenesis to solventogenesis key genes are differentially expressed. Genes associated with solvent production, such as *adc*, *aad*, *bdhA*, *bdhB* and the *sol* operon (*adhE-ctfA-ctfB*) have previously been characterised (Gerischer and Durre, 1990; Walter, Bennett and Papoutsakis, 1992; Fischer, Helms and Durre, 1993; Nair, Bennett and Papoutsakis, 1994). The genes, *adhE1*, encoding an NADH-dependent aldehyde dehydrogenase, and *bdhA*/*B,* encoding butanol dehydrogenase, are necessary for butanol formation during solventogenesis (Yang *et al.*, 2018). The expression of the *sol* operon has been shown to be essential for acid reassimilation and solvent production in *Clostridium* (Feng *et al.*, 2020).

A master regulator for the expression of genes associated with solvent production and multiple characteristics associated with the solventogenesis phase has been indicated. For example, a study by Woolley and Morris (1990) demonstrated how *C. acetobutylicum* mutants defective in spore, granulose and solvent production spontaneously reverted to WT capabilities, indicating a single mutation was responsible for the loss of all three traits. It was suggested that a global regulatory gene such as *spo0A*, the master regulator of sporulation, co-ordinates the expression of these three traits. In a similar study, Ravagnani *et al.* (2000) identified that ∆*spo0A* mutants of *C. beijerinckii* were unable to produce solvents, spores or granulose.

Sequence inspection by Ravagnani *et al.* (2000) in both *C. beijerinckii* NCIMB 8052 and *C. acetobutylicum* ATCC 824 revealed that Spo0A may directly control the expression of genes involved in solvent and/or acid formation by binding to 0A boxes in their promoter regions. Further studies have supported Spo0A as a transcriptional regulator that upregulates expression of solvent formation genes (Harris, Welker and Papoutsakis, 2002; Alsaker, Spitzer and Papoutsakis, 2004; Shi and Blaschek, 2008; Seo *et al.*, 2017). In *C. beijerinckii* NCIMB 8052 *and C. acetobutylicum* ATCC 824 Spo0A is assumed to directly regulate the *sol* operon, however in *C. saccharoperbutylacetonicum* N1-4, there is no 0A box within the promoter region of the *sol* operon, suggesting indirect regulation (Kosaka *et al.*, 2007).

Although Spo0A has been shown to upregulate solvent production in *C. beijerinckii* and *C. acetobutylicum*, a different response to *spo0A* inactivation was observed in *C. pasteurianum*. Here, butanol production increased due to Spo0A deficiency showing solvent production is regulated differently in this species (Sandoval *et al.*, 2015).

It should be noted that solvent production is not fully reliant upon sporulation. An early study by Jones *et al.* (1982) identified that *spo* mutants of *C. acetobutylicum,* which were unable to form mature spores, could still produce solvents at the same level as the WT. Other studies found mutants lacking in Spo0A did not abolish expression of genes in the *sol* locus (*aad-ctfA-ctfB* and *adc*) (Harris, Welker and Papoutsakis, 2002). Further studies, in *C. acetobutylicum*, have shown evidence of alternative regulatory factors for solventogenesis include

transcription factors such as Rex, AbrB, CcpA and CsrA (Wietzke and Bahl, 2012; Ren *et al.*, 2012; Tan *et al.*, 2015; Xue *et al.*, 2016). Alternative regulatory factors are not as well studied in *C. beijerinckii* but may include XylR and LytS/YesN (Sun *et al.*, 2015; Wen *et al.*, 2017). Additionally, in solventogenic *Clostridium*, the switch to solventogenesis and sporulation occurs at higher population densities in batch culture suggesting that QS has a role in regulation (Kosaka *et al.*, 2007; Steiner *et al.*, 2012; Feng *et al.*, 2020; Kotte *et al.*, 2020).

1.5 Quorum Sensing

Quorum sensing (QS) is a mechanism of cell-to-cell communication used in both Gram-positive and Gram-negative bacteria. The process is dependent on secreted signalling molecules, called autoinducers, secreted by the bacteria (Miller and Bassler, 2001). As the bacterial population density increases, extracellular autoinducers accumulate in the surrounding environment until a threshold is reached, resulting in differential gene expression across the whole population. QS therefore allows the local cell population to make a coordinated response to changes in population density (Bassler and Losick, 2006). Many processes, in a variety of bacterial species, are known to be affected by QS including virulence, competence, motility, biofilm formation, bioluminescence, sporulation, solvent and antibiotic formation. Such

events are unlikely to occur if only a singular cell was present (Bassler and Losick, 2006; Abisado *et al.*, 2018).

1.6 Gram-Negative Quorum Sensing

The signalling molecules involved differ between Gram-positive and Gram-negative bacteria. In Gram-negative bacteria, the primary signalling molecules are fatty acid derivatives such as acyl-homoserine lactones (AHL's) which are derived from acyl-ACPs and *S*-adenosylmethionine (SAM) (Abisado *et al.*, 2018). AHL-QS was first discovered in *Allivibrio fischeri* (formerly *Vibrio fischeri*) in which bioluminescence was shown to be controlled by the proteins LuxI and LuxR (Engebrecht and Silverman, 1984). The protein LuxI is a signal synthase enzyme responsible for synthesizing the autoinducer N-3-oxo-hexanoyl-homoserine lactone (3OC6-HSL). At high cell density, when 3OC6-HSL concentration has reached a threshold, it diffuses freely back into the bacterial cytoplasm. Here, it is detected by cytoplasmic LuxR, a transcription factor that regulates expression of the *luxCDABEG* operon containing bioluminescent genes (Engebrecht and Silverman, 1984; Abisado *et al.*, 2018).

In most cases of Gram-negative bacteria, the QS mechanism resembles the LuxIR-type AHL QS system of *A. fischeri*. In these species, specific AHL's are synthesised by LuxI homologues and then detected by LuxR homologues. Some of the best studied examples of LuxIR homologues include *Pseudomonas aeruginosa* RhlIR and LasIR systems controlling virulence gene expression, *Erwinia carotovora* CarIR and ExpIR, controlling antibiotic and virulence factors respectively, and the *Agrobacterium tumefaciens* TraIR system controlling transfer and replication of the oncogenic Ti plasmid and hence virulence (Miller and Bassler, 2001; Lang and Faure, 2014).

Alternative non-AHL QS systems are also used in Gram-negative bacteria. Examples are quinolones in *Pseudomonas* species (Pesci *et al.*, 1999), cyclic dipeptides (2,5-diketopiperazines) in multiple Gram-negatives species (Campbell *et al.*, 2009), the diffusible signal factor (DSF) in *Xanthomonas* spp. (Guo *et al.*, 2012), and the AI-2 autoinducer first proposed in *Vibrio harveyi* (Bassler, Wright and Silverman, 1994). Furthermore, the Gram-negative bacterium *Myxococcus xanthus* has a distinctive QS system not coordinated by AHLs. At high cell density the bacteria produce structures termed fruiting bodies, inside of which the bacteria sporulate. The secreted quorum signalling molecule, A-signal, is made up of a combination of proteases, peptides and amino acids and is required for sensing cell starvation and population density. The receptor for A-signal is SasS, a sensor histidine kinase, which upon binding leads to a signalling cascade resulting in the upregulation of genes involved in spore

formation (Miller and Bassler, 2001). The process is similar to the two-component response regulatory systems found in Gram-positive bacteria (**section 1.7.1**).

1.7 Gram-Positive Quorum Sensing

In contrast to Gram-negative bacteria, Gram-positive bacteria use short, secreted peptides as the autoinducer in QS, which can be linear or circular (Bhatt, 2018). The mechanism of peptide-based QS varies regarding the secretion apparatus and the autoinducer sensor. One class involves a two-component response regulatory system primarily made up of a membrane bound sensor kinase as the autoinducer receptor, and an intracellular response regulator (Sturme *et al.*, 2002). Another class involves the internalization of the autoinducer where it interacts directly with a cytoplasmic transcription factor belonging to the RRNPP family (**section 1.7.2**; Bhatt, 2018).

1.7.1 Two-component response regulatory systems

In terms of the two-component response regulatory QS systems (TCS), the general process occurs as follows. The precursor of the autoinducing peptide (AIP) is synthesised in the cell by ribosomes. After

post-translational modifications, the active AIP is secreted from the cell by transporters, a common example being ATP-binding cassette (ABC) transporters, although this varies between systems (Sturme *et al.*, 2002; Bhatt, 2018). The extracellular AIP concentration increases alongside increasing cell density until a concentration threshold is reached. The AIP is then detected by a two-component regulatory system made up of a sensor histidine kinase on the bacterial membrane, opposed to the LuxR-type autoinducer sensors present in most Gram-negatives (Miller and Bassler, 2001). This detection causes autophosphorylation of the histidine residue on the sensor kinase which then leads to phosphorylation of a conserved aspartate residue on a dedicated response regulator inside the cell. The phosphorylated response regulator is the activated form which binds to the promoter region of target genes, altering their expression (Miller and Bassler, 2001; Bassler and Losick, 2006).

The sensor kinase and dedicated response regulator differ between TCS. Examples of systems include ComPA in *B. subtilis* and ComDE in *Streptococcus pneumoniae,* both controlling competence (Piazza, Tortosa and Dubnau, 1999; Shanker and Federle, 2017). The AgrAC in *Staphylococcus* species and VirSR in *C. perfringens*, both regulating virulence factors (Novick and Geisinger, 2008; Cheung *et al.*, 2010). And finally, the NisKR system in *Lactobacillus lactis,* responsible for sensing the bacteriocin nisin (Kuipers *et al.*, 1998).

1.7.1.1 Agr Quorum Sensing

The Agr system in *Staphylococcus* has been extensively studied, **Figure 1.3** shows a general schematic of the Agr system in *S. aureus*. The system is made up of two transcriptional units which are called RNAII expressed from promoter 2, encoding the *agrBDCA* genes, and RNAIII expressed from promoter 3, acting as the effector molecule that regulates virulence gene expression. The system is based upon the recognition of an AIP by a two-component regulatory system. The AgrD is the precursor for the AIP which is a 7-9 amino acid peptide consisting of a short N-terminal tail and a five amino acid thiolactone ring (Novick and Geisinger, 2008). The transmembrane protein AgrB modifies and cleaves AgrD so it can cross the membrane and fold into the mature AIP. The histidine kinase sensor AgrC and designated response regulator AgrA make up a two-component response regulator. At a concentration threshold, the extracellular AIP is sensed by transmembrane receptor domain of AgrC. This leads to phosphorylation of the AgrC cytoplasmic HPK domain. The phosphate group is then transferred to AgrA, resulting in activated AgrA-Pi. The Activated AgrA induces the expression of P2 and P3 promoter gene regions, which upregulates the expression of RNAII and RNAIII, leading to further expression of the *agr* locus and upregulation of virulence genes (Le and Otto, 2015). The *agr* locus can be positively and negatively regulated by a variety of regulatory proteins showing it is part of a complex

regulatory network and plays a role in many pathways (Novick and Geisinger, 2008).

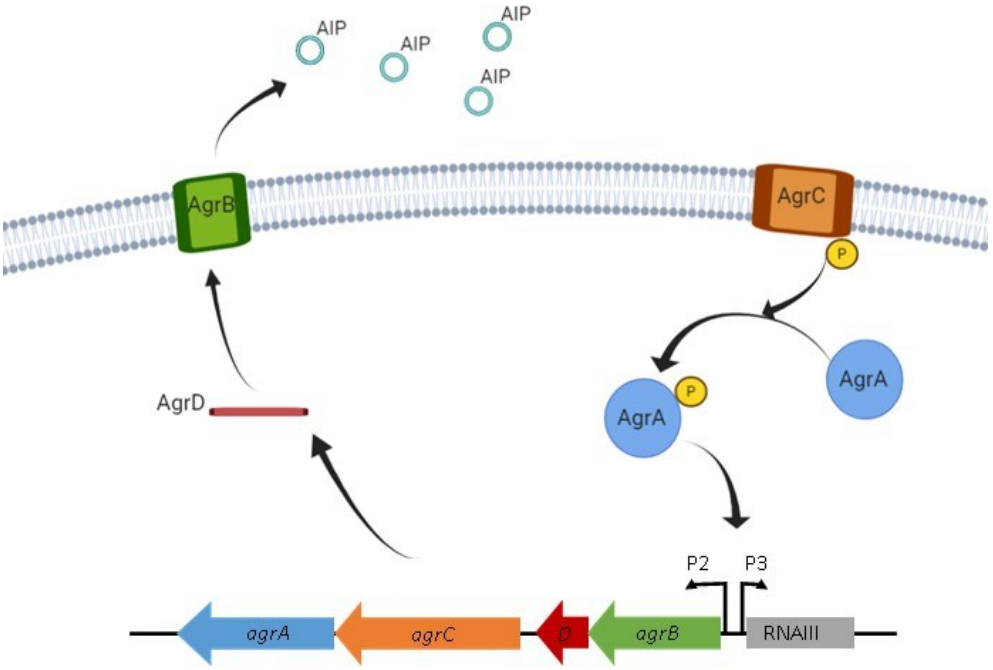


Figure 1.3 Schematic of the Agr quorum sensing signalling system in *Staphylococcus aureus*.

The AgrD precursor peptide is processed into a mature autoinducing peptide (AIP) and excreted to the exterior of the cell by the AgrB transmembrane protein. The mature cyclic AIP eventually reaches a concentration threshold at which it is detected by the designated AgrC, a membrane-bound histidine kinase sensor. This detection results in the phosphorylation and activation of AgrA, the response regulator. The activated AgrA upregulates the expression of the *agr* genes and the RNAIII transcript. The RNAIII transcript regulates the expression of various virulence genes in *S. aureus*. Created with BioRender.com.

1.7.2 Direct detection by cytoplasmic receptors: RRNPP

For the Gram-positive QS systems where the autoinducer binds directly to the target cytoplasmic regulator, there appears to be one key example in literature so far. The superfamily RNPP was originally suggested by Declerck *et al.* (2007) made up of the proteins: regulator aspartate phosphatases (Rap) in *B. subtilis*, neutral protease regulator (NprR) in *B. subtilis*, phospholipase C regulator (PlcR) in *Bacillus cereus*, and pheromone-responsive regulator (PrgX) in *Enterococcus faecalis* (Rocha-Estrada *et al.*, 2010). It was later proposed that the regulator gene of glucosyltransferase (Rgg) in *Streptococcus* be added to the family and it be updated to RRNPP (Parashar *et al.*, 2015).

The cytoplasmic regulators are grouped according to structural and phylogenetic analysis for instance, a study by Declerck *et al.* (2007) demonstrated how Rap, NprR, PlcR and PrgX contained conserved features and originated from a common ancestor.

The RRNPP transcription regulators interact with small linear AIPs between 5-9 amino acids in length with high specificity (Do and Kumaraswami, 2016). The genes encoding for both the regulator and specific AIP are often located together on the bacterial chromosome or in plasmids. The precursor AIP is cleaved by proteolytic enzymes and secreted into the surrounding environment in its mature form where they can interact with other cells. The RNPP regulators are cytoplasmic,

therefore, AIPs must enter cells through oligopeptide permease structures (Rocha-Estrada *et al.*, 2010). All RRNPP family proteins contain a similar tetratricopeptide (TPR)-like repeat domain in their C-terminus which allows interaction with the AIP (Declerck *et al.*, 2007; Parashar *et al.*, 2015). With the exception of Rap phosphatase, binding of the signal peptide to the TPR domain leads to conformational changes resulting in the exposure of or activation of an N-terminal DNA-binding domain (DBD) which interact with target DNA sequences to change gene expression. In the case of Rap phosphatases, these affect gene expression by binding to other transcription regulators and altering their activity (Declerck *et al.*, 2007; Parashar *et al.*, 2015; Do and Kumaraswami, 2016).

RRNPP-type QS has been observed in many *Firmicutes*. A study by Neiditch *et al.* (2017) found that homologues of Rgg and PrgX were mostly present in bacteria belonging to the order *Lactobacillales*, compared to homologues of Rap and NprR which were found mostly in *Bacillales.* Meanwhile PlcR homologues were found to be distributed evenly between both orders.

The RRNPP regulators have a role in various processes, for example sporulation in *B. subtilis.* Here, Rap phosphatases such as RapABEHJ dephosphorylate Spo0F response regulator, which therefore prevent Spo0A activation. Rap phosphatases can be inhibited by the short linear AIP called Phr, therefore allowing Spo0A activation and sporulation to occur (Neiditch *et al.*, 2017). NprR is also thought to play a role in sporulation in *Bacillus thuringiensis*. Structural analysis

showed a Spo0F-binding residue, like that found in Rap phosphatases of *B. subtilis*, between the DBD and TPR domains of the protein. However, it remains unclear as to whether NprR positively or negatively regulates sporulation in *Bacillus* (Cabrera *et al.*, 2014; Neiditch *et al.*, 2017).

Whilst most RRNPP regulators are encoded alongside their respective AIP, there are some exceptions where the RNPP regulator gene is not flanked by the respective AIP gene, these are termed orphan regulators. The regulator of protease B (RopB) from *Streptococcus pyogenes,* induced under acidic conditions (Do *et al.*, 2019), and Rgg from *Streptococcus gordonii*, both Rgg-type regulators controlling virulence gene expression, are both examples of orphan regulators (Neiditch *et al.*, 2017). Furthermore, there are 11 rap genes in *B. subtilis*, whilst most of these are encoded within the *phr* signal cassette, *rapD* and *rapJ* are not encoded in an operon alongside a *phr* signal, hence they are classed as orphan regulators (Parashar, Jeffrey and Neiditch, 2013).

1.8 Inter-Species Communication: AI-2 Based Quorum Sensing

The family of molecules called autoinducer-2 (AI-2) were originally proposed in *Vibrio harveyi,* where it is partially responsible for bioluminescence (Bassler, Wright and Silverman, 1994). The genes

required for AI-2 synthesis, such as *luxS*, are found in both Gram-negative and Gram-positive bacteria, providing a platform for inter-species communication. The LuxS enzyme synthesises AI-2 molecules from the precursor 4,5-dihydroxy-2-pentanedione (Miller and Bassler, 2001; Bassler and Losick, 2006). The structure of mature AI-2 molecules remained unknown for many years however, X-Ray crystallography identified the AI-2 precursor bound to a natural borate to form the mature AI-2 molecule, a furanosyl borate diester, unique to other characterised autoinducers (Chen *et al.*, 2002). There appear to be two key families of AI-2 receptors, LuxP from *Vibrio* species and LsrB from a variety of bacteria such as *Salmonella* sp., *Bacillus* sp. and identified more recently in *Clostridia* (Chen *et al.*, 2002; Miller *et al.*, 2004; Torcato *et al.*, 2019). From here, the AI-2 can be internalised and processed before binding to transcription factors to alter gene expression (Miller and Bassler, 2001). Homologues of LuxS are documented in many bacteria where AI-2 type QS controls a range of processes like biofilm formation in *Streptococcus mutans*, virulence in *Salmonella typhimurium*, biofilm formation in *B. cereus* and motility in *Helicobacter pylori* (Merritt *et al.*, 2003*;* Auger *et al.*, 2006; Rader *et al.*, 2007; Choi *et al.*, 2012).

However, *luxS* homologues in most organisms, except *V. harveyi*, have been found to be encoded within proximity to genes involved with the methionine recycling metabolic pathway, such as *metB*, *metH* and *cysK* (Carter *et al.*, 2005). It is only in *V. harveyi* that concrete evidence for AI-2 being a direct QS autoinducer has been demonstrated. In contrast

to the studies mentioned, this indicates that the *luxS* homologues may play a role in metabolism and the resulting AI-2 is just a by-product, not forming a truly established QS system (Winzer *et al.*, 2002).

1.9 Quorum sensing in *Clostridia*

Currently QS mechanisms in *Clostridia* are not well characterised. Nevertheless, peptide-based QS has been observed in several members of the genus *Clostridium*. This includes two-component regulatory, RRNPP-type and AI-2 QS systems in both pathogenic and solventogenic species.

For two-component regulatory systems, the most frequent in literature mentioned for *Clostridium* is *agr*-dependent QS. This contributes to the regulation of factors in pathogenic species like sporulation and/or virulence in *C. botulinum* (Cooksley *et al.*, 2010), *C. perfringens* (Li *et al.*, 2011; Vidal *et al.*, 2012; Yu *et al.*, 2017; Navarro *et al.*, 2020) and *C. difficile* (Darkoh *et al.*, 2015). It also has a role in sporulation and solvent production in solventogenic *Clostridia* like *C. acetobutylicum* (Steiner *et al.*, 2012; Jabbari *et al.*, 2013). Steiner *et al.* (2012), found that in *C. acetobutylicum* ATCC 824 the *agrBDCA* region is involved in the regulation of both granulose production and sporulation. The system works in a similar fashion to that of the Staphylococcal Agr system in that *agrBD* produces the cyclic AIP and *agrCA* form the two-

component regulatory system. However, unlike in *Staphylococcus*, *agrBD* and *agrCA* are independently transcribed in *C. acetobutylicum* and there does not appear to be a distinct RNAIII transcript. Most members of the genus *Clostridium* appear to contain homologues of *agrBD* in their genome (Steiner *et al.*, 2012). A study by Patakova *et al.* (2019) found three *agr* QS gene clusters in the genome of *C. beijerinckii* NRRL B-598 but did not identify a distinct RNAIII transcript. Meanwhile, Piatek *et al.* (2022) determined that Agr QS influenced the Wood-Ljungdahl pathway in *Clostridium autoethanogenum*. Further genomic analysis has identified 4 putative *agr* loci in *C. saccharoperbutylacetonicum* and 6 in *C. beijerinckii* NCIMB 8052 (Kotte *et al.*, 2020). In *C. beijerinckii*, it has also been suggested that sporulation-related genes are regulated by Agr QS (Patakova *et al.*, 2021). However, whilst the Agr system could play a similar role in other solventogenic *Clostridia*, very little experimental data has yet been generated.

Studies have found putative homologues of the RRNPP regulators Rap, Rgg, PlcR and NprR in *Clostridia*. For instance*, C. acetobutylicum* genome is suggested to contain two *rap-phr* systems, like those found in *B. subtilis,* denoted as CAC0186 and CAC3693 (Pottathil and Lazazzera, 2003; Neiditch *et al.*, 2017). A more recent study by Kotte *et al.* (2020) identified 8 putative RRNPP-type QS systems (QssA-H) in *C. acetobutylicum* ATCC 824. Inactivation of 7 of these systems resulted in a change in solvent production with QssB indicated to regulate early

solventogenesis and sporulation. Comparison to other solventogenic-type genomes found 5 putative RRNPP-type systems in *C. saccharoperbutylacetonicum*, however no complete RRNPP-type systems in *C. beijerinckii* NCIMB 8052. The analysed genomes also encoded a number of orphan RRNPP-type regulators, for instance CA_C0957 and CA_C0958 proteins in *C. acetobutylicum* ATCC 824, which played a regulatory role in solventogenesis and sporulation (Kotte *et al.*, 2020). Another recent study by Feng *et al.* (2020) found that multiple RRNPP-type systems (QSS1235) in *C. saccharoperbutylacetonicum* N1-4 upregulated the *sol* operon and therefore solvent production. These recent findings show RRNPP-based QS is a hot-topic within the field.

AI-2 QS has also been indicated in *Clostridium*. A study by Carter *et al.* (2005) located a *luxS* homologue in *C. difficile* which is responsible for AI-2 synthesis. Nearby genes encoding a response regulator, *rolA*, and a sensor kinase, *rolB*, making up a TCS were located. These were shown to regulate the transcription of *luxS* and therefore synthesis of AI-2 through a TCS, a common mechanism in Gram-positive bacteria (see **section 1.7.1**). However, as recently discussed (section. 1.8), *luxS* homologues in bacteria other than *V. harveyi* are thought to play a role in metabolism rather than a truly established QS system. Nevertheless, recent identification and characterisation of a functional LsrB-type receptor in *C. saccharobutylicum* regenerates the idea of AI-2 being established in *Clostridium* (Torcato *et al.*, 2019).

It is evident that *Clostridium* species have adopted multiple QS systems, a feature which may allow the regulation of a complex system of genes at different density thresholds. It is evident QS plays a role in ABE metabolism, with studies demonstrating increased productivity at higher cell densities (S. J. Choi *et al.*, 2012). It has also been suggested that uncharacterised QS mechanisms may be responsible for lower cell-densities during ABE fermentations and therefore lower solvent yields (Papoutsakis, 2008). To add to complexity, recent studies have identified various molecules that may act as signalling molecules, affecting the regulation of sporulation, granulose formation, and butanol production. A study by (Herman *et al.*, 2017), demonstrated how mutation of the gene encoding the polyketide clostrienose in *C. acetobutylicum* resulted in increased solvent production and reduced sporulation. Another study by Chen *et al.* (2020) found that ranthipeptides such as beijerinckin in *C. beijerinckii* NCIMB 8052, play important roles in regulating cell population and solvent production. The study found that mutants which could not produce beijerinckin could produce n-butanol at a faster and higher rate compared to the WT. Further transcriptomic analysis showed that the *agr* loci genes in mutants were upregulated during the acidogenic phase, suggesting a link to QS. Once again, this is an area which would require further research clarification.

1.10 Genetic Manipulation Tools in *Clostridium*

The genus *Clostridium* has a limited number of stable genetic manipulation techniques developed. The low plasmid transformation and DNA repair efficiency in *Clostridium* species hindered both the development and application of alternative genetic manipulation tools for a relatively long period of time. Until recently, genetic manipulation in *Clostridium* species was largely dependent on techniques based on homologous recombination techniques, involving plasmids with targeted homology regions. Due to a lack of counterselection markers, the process often relied on the occurrence of a single cross-over insertion, in which the whole plasmid would insert into the chromosome however, this resulted in unstable mutants. Advancements have taken place within the last decade with key manipulation techniques being based on intron insertion, further developed homologous recombination tools and CRISPR-Cas9 systems.

1.10.1 Intron-Based Genetic Manipulation Tools

A popular alternative to homologous recombination-based methods were retargeted group II intron tools such as the TargeTron. This genomic manipulation tool is based on the insertion of group II intron RNA targeted to specific genes of interest. These group II introns

consist of a catalytically active RNA (ribozyme) and an intron-encoded protein (IEP) which has reverse transcriptase activity (Wen *et al.*, 2020). The intron can fold into a secondary structure which consists of 6 domains from which domain VI can achieve self-splicing, resulting in an excised lariat RNA. The Intron lariat RNA and IEP can form a complex, called the RNP particle, which then recognise specific DNA target sites (Zhong and Lambowitz, 2003; Wen *et al.*, 2020). The intron lariat RNA splices into one strand of the target DNA whilst the IEP component cleaves and unwinds the second strand to allow for reverse transcription of the inserted intron RNA. This results in an integrated cDNA in the target region of the host genome through the host DNA repair systems (Zhong and Lambowitz, 2003). Depending on the site of RNA intron insertion, gene inactivation through this method can be made conditional, where the intron RNA is removed from the mRNA transcript via self-cleavage and thus the transcribed target gene is successfully translated and remains functional. Alternatively, unconditional inactivation can occur, when the intron RNA is inserted into the antisense strand or towards the end of the target gene, this condition means the intron RNA cannot be removed (Frazier *et al.*, 2003; Heap *et al.*, 2010; Wen *et al.*, 2020).

The TargeTron was originally developed by Karberg *et al.* (2001), based on the mobile group II intron LI.ltrB from *Lactococcus lactis,* and used in *E. coli* to demonstrate highly specific chromosomal gene disruption. The method was later adapted to work in other

microorganisms, for instance just a few years later, such intron-based technologies were initially implemented in *C. perfringens* by Chen *et al.* (2005). Further studies by Heap *et al.* (2007) adapted the technique to be used in a variety of *Clostridia*. The ClosTron was initially based on the pMTL007 plasmid, containing an intron derivative of the Ll.ltrB from *L. lactis* and a retrotransposition-activated marker (RAM). The RAM typically consists of an inactive copy of the gene *ermB,* a self-cleavage group I intron, which inserts into the resistance gene, and group II introns. Upon successful insertion of the group II intron into the target DNA, the group I intron self-splices from the resistance gene which then becomes activated, providing resistance to the erythromycin antibiotic. This allowed for quick identification for positive intron insertion without the need for phenotypic screening (Heap *et al.*, 2007). Two years later a study by Heap *et al.* (2009) developed a standardised modular system for the construction of *Clostridium-E. coli* shuttle vectors based on the pMTL80000 plasmid. Further streamlining of the ClostTron mutagenesis system was carried out by Heap *et al.* (2010) to allow for marker-recycling which would enable multiple insertions to be carried out in the same genome. This involved the addition of an FLP recombinase recognition target (FRT) to the intron allowing the subsequent removal of the selectable marker from the genome. The intron-based mutagenesis technology has been used on a variety of *Clostridium* species including, but not limited to, *C. acetobutylicum*, *C. beijerinckii, C. autoethanogenum C. botulinum*, *C. sporogenes* and *C. difficile* (Shao *et al.*, 2007; Heap *et al.*, 2010; Kuehne *et al.*, 2011;

Steiner *et al.*, 2012; Wang *et al.*, 2013; Liew *et al.*, 2017; Gu *et al.*, 2018; Little *et al.*, 2018).

There are some clear benefits to using the ClosTron based technology in *Clostridium*. For example, it overcomes the issue of low homologous recombination and transformation efficiency faced in *Clostridia* as it has little reliance upon these. Furthermore, the improvements introduced by Heap *et al.* (2010) resulted in the method being very quick and easy to perform with very little labour required. The efficiency of the ClosTron is so high that the positive intron insertion mutants can be easily identified by PCR and gel electrophoresis, therefore it does not even require a selectable marker (Wen *et al.*, 2020).

Whilst ClosTron technology has shown to be of huge value in research on *Clostridia*, there have been various limitations. A major problem with using this technology is the occurrence of polar effects on downstream genes. One issue is that it leaves a "scar", the integrated intron, in the host chromosome which could potentially lead to unprecedented polar effects, especially if the target gene is part of an operon (Heap *et al.*, 2010; Wen *et al.*, 2020). Polar effects have also been known to occur due to off-target insertion (Cooksley *et al.*, 2012). Furthermore, the ClosTron cannot knock-out the whole open reading frame (ORF), it only disrupts transcription and therefore some gene functionality may remain (Wen *et al.*, 2020). In response to these issues, a marker-less method of in-frame gene deletions was desirable.

1.10.2 Allelic-Coupled Exchange

In response to the issues faced with other genetic manipulation techniques, a newly improved marker-less homologous recombination technique was developed for *Clostridia* (Heap *et al.*, 2012). This technique involves the use of plasmids consisting of homology arms which are complementary to the flanking regions of the chromosomal target genes. These homology arms allow for target DNA to be inserted or removed from the chromosome through a double-crossover homologous recombination event without leaving behind a “scar” in the genome. An important strategy which relies on this principle was termed Allelic Coupled Exchange (ACE) (Heap *et al.*, 2012; Ehsaan *et al.*, 2016). This method relies on a plasmid-based allele becoming ‘coupled’ with a genome-based allele during the second recombination event, resulting in a new selectable allele. This then allows selection of double-crossover mutants (Ehsaan *et al.*, 2016; Minton *et al.*, 2016).

There are important advantages in using ACE compared to intron-based genetic manipulation techniques. In particular, the ACE technique leaves no ‘scarring’ in the genome, making it a highly desired markerless method which leads to no unprecedented polar effects. Furthermore, unlike the intron-based methods described, ACE can remove the whole ORF, ensuring no gene functionality remains. As it

stands, the main downfall of ACE technology is that homologous recombination in *Clostridia* has a low efficiency, making ACE time-consuming and labour intensive (Wen *et al.*, 2020).

These more recent homologues recombination technologies have been improved through the development of a variety of counter-selection markers shown to work within other bacterial species. In Gram-positive bacteria, common counterselection markers rely on genes encoding enzymes involved in purine or pyrimidine metabolism. For example, *codA* (cytosine deaminase), *pyrE* (orotate phosphoribosyltransferase), *hpt* (hypoxanthine phosphoribosyltransferase) and *upp* (phosphoribosyltransferase) (Martinussen and Hammer, 1994; Argyros *et al.*, 2011; Ehsaan *et al.*, 2016; Huang, Liebl and Ehrenreich, 2018). Such systems rely on the idea that purine or pyrimidine analogues are converted into toxic compounds by conversion enzymes. Because of this, only bacterial cells which lack the gene for the conversion enzymes can survive in the presence of the analogue (see **sections 1.10.2.1** and **1.10.2.2**). Alternative counterselection agents are based on endogenous toxin/antitoxin systems such as *mazE/mazF* (Al-Hinai, Fast and Papoutsakis, 2012).

1.10.2.1 PyrE System

The *pyrE* gene encodes orotate phosphoribosyltransferase, an enzyme involved in *de novo* pyrimidine biosynthesis. The enzyme converts the pyrimidine intermediate orotic acid into orotidine monophosphate (OMP). The *pyrE* gene becomes essential in the absence of exogenous pyrimidines and renders 5-fluoroorotate (FOA) toxic to cells. FOA is an analogue of orotic acid which is converted, by the PyrE enzyme, into 5-fluoroorotidine monophosphate (5-FOMP) which is further converted into 5-fluorouridine monophosphate (5-FUMP) (Minton *et al.*, 2016). This becomes toxic to cells as it can be misincorporated into cellular DNA and RNA in place of thymine or uracil (Ehsaan, Kuehne and Minton, 2016). These features allow *pyrE* to be used as a counter-selection marker in *Clostridia*. The strategy involves the inactivation of the WT *pyrE* allele by replacing it with a plasmid-derived truncated version of the allele which is missing the first 300 bp of the 3' end of the gene (Heap *et al.*, 2012; Ehsaan, Kuehne and Minton, 2016). This is done using ACE, demonstrated by **Figure 1.4**. Upon successful allele exchange, the *pyrE* mutant cells can then be selected for due to their resistance to FOA.

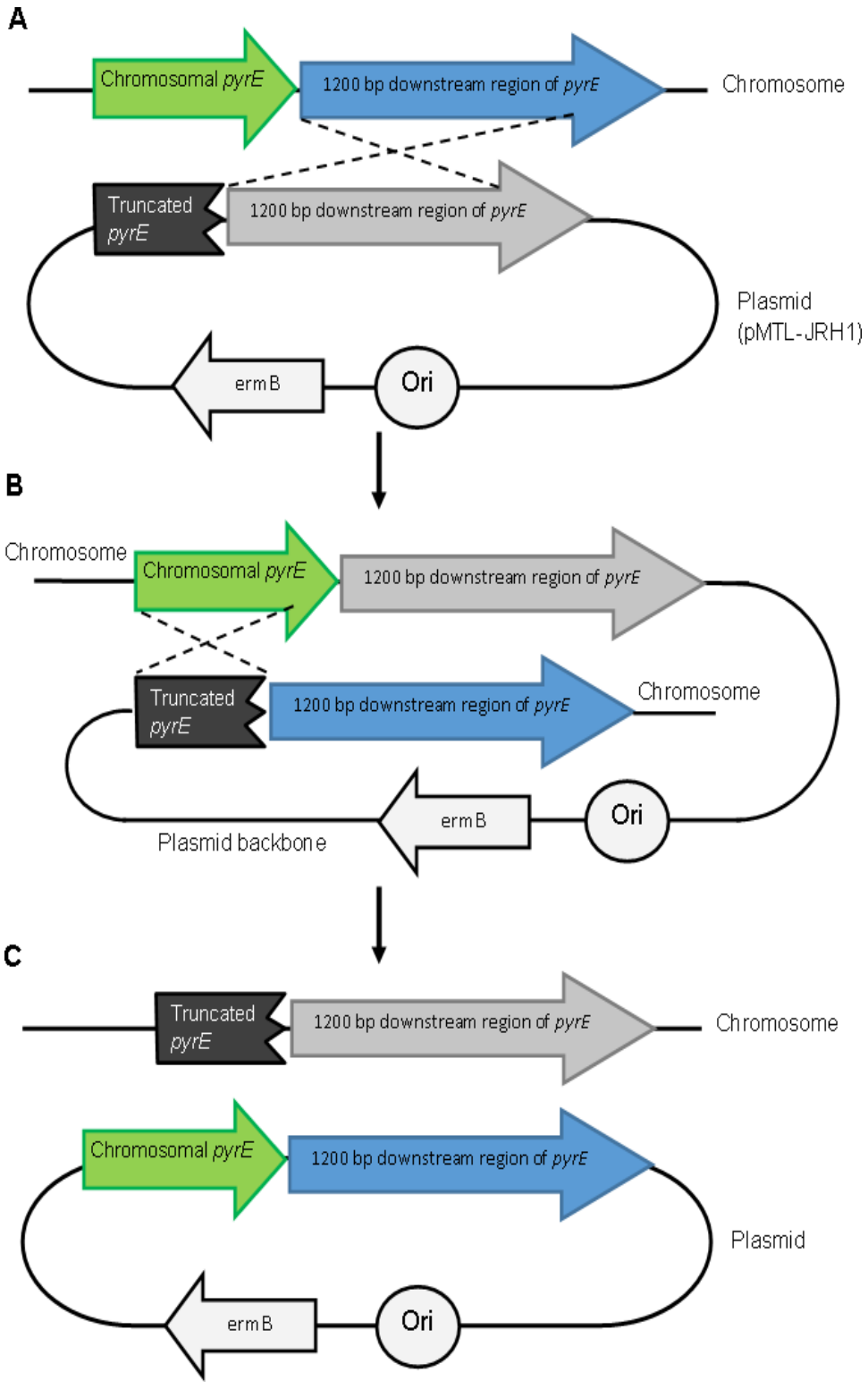


Figure 1.4 Creation of the pyrE mutant using allelic-coupled exchange (ACE). Adapted from Heap et al. 2012.

Knockout vectors containing in-frame deletion homology arms and the functional *pyrE* gene can then be designed. The knockout vectors are transformed into *pyrE* mutant cells, where homologous recombination between the plasmid-based homology arms and the target gene region occurs, demonstrated in **Figure 1.5.** After the second recombination event, FOA acts as the counterselection agent for plasmid loss (Heap *et al.*, 2012; Ehsaan *et al.*, 2016; Ehsaan, Kuehne and Minton, 2016; Minton *et al.*, 2016). The *pyrE* mutant cells also require growth media to be supplemented with uracil as the exogenous pyrimidine source (Ehsaan, Kuehne and Minton, 2016). An important feature of this method is that the truncated *pyrE* allele can be restored to the fully functional WT allele using an ACE correction vector (Heap *et al.*, 2012; Ehsaan *et al.*, 2016). This allows the specified target gene mutation to be analysed without the issue of unprecedented polar effects which can occur due to 'scarring' in the genome through other genetic manipulation methods (Ehsaan *et al.*, 2016).

The *pyrE* counterselection system has been developed for use in gene substitutions, addition or deletion in many *Clostridia* including but not limited to *C. difficile*, *C. perfringens*, *C. botulinum*, *C. acetobutylicum, C. pasteurianum, C. autoethanogenum, C. ljungdahlii* and *C. beijerinckii* (Ehsaan *et al.*, 2016; Ehsaan, Kuehne and Minton, 2016; Little *et al.*, 2018).

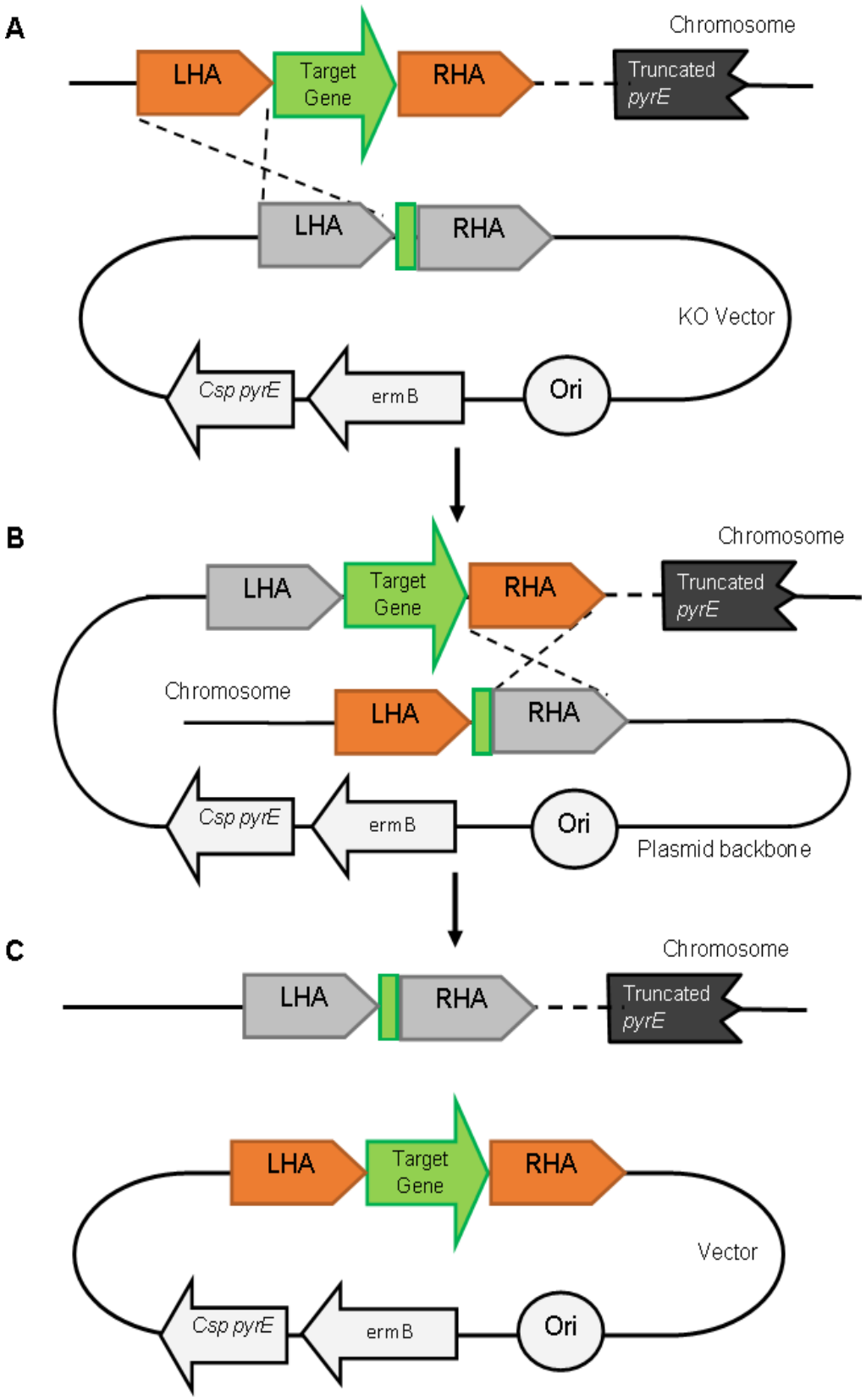


Figure 1.5 Generation of in-frame deletions using the *pyrE* system. Adapted from Ehsaan, Kuehne and Minton (2016).

1.10.2.2 The *codA* System

The *codA* system can be used in a similar way to the *pyrE* system however, the *codA* gene of *E. coli* encodes cytosine deaminase, an enzyme that is required for converting cytosine to uracil but also renders the cytosine analogue 5-fluorocytosine (5-FC) toxic to cells due to the accumulation of 5-fluorouracil (5-FU) which results in the misincorporation of fluorinated nucleotides into both RNA and DNA. Therefore 5-FC acts as the counterselection marker in this system (Cartman *et al.*, 2012; Huang, Liebl and Ehrenreich, 2018). The *codA* system was developed for use in *C. difficile* by Cartman *et al.* (2012) who used *codA*-mediated allele exchange to restore and induce mutations in the *tcdC* gene. The codA system has also been developed for solventogenic *Clostridia*, for instance Huang, Liebl and Ehrenreich (2018) used the *codBA* genes from *Clostridium ljungdahlii* as a counterselection marker to introduce chromosomal deletions into *C. saccharobutylicum* NCP 262. Before this, Ehsaan *et al.* (2016) developed the both the *codA* and *pyrE* system to knockout target genes in the *C. acetobutylicum* ATCC 824 chromosome. Whilst it was concluded that both systems were equally effective as a genomic manipulation technique in *Clostridium*, the *pyrE* system was favoured since, through ACE, the *pyrE* mutation could undergo rapid restoration compared to the *codA* mutation.

1.10.2.3 mazE/mazF System

The MazEF toxin-antitoxin system from *E. coli* is encoded by the *mazEF* operon. The *mazF* gene encodes a stable ribosome-independent endoribonuclease which cleaves at the ACA sequence. Meanwhile, the *mazE* gene encodes a labile antitoxin, rapidly degraded by cellular proteases. The MazF toxin, referred to as an mRNA interferase, leads to selective mRNA cleavage under ectopic expression, which results in growth arrest and cell death (Zhang *et al.*, 2003; Yamaguchi and Inouye, 2009; Al-Hinai, Fast and Papoutsakis, 2012). The MazEF system has been successfully used for genetic manipulation in a variety of *Clostridium* species including *C. acetobutylicum* (Al-Hinai, Fast and Papoutsakis, 2012).

1.10.3 CRISPR

Whilst homologous recombination-based methods have been the main approach towards targeted genetic manipulation historically, the efficiency of homologous recombination in many Gram-positive bacteria is low. It was later found that double-stranded breaks in target DNA could increase the efficiency of homology-directed repair and therefore targeted nucleases have become of huge interest to create target gene

mutations. The targeted nuclease can be used alongside a repair DNA template to create the desired mutation. In recent decades the CRISPR-associated 9 (Cas9) nucleases have been increasingly used to create targeted DNA breakages (H. Li *et al.*, 2020).

The term CRISPR stands for clustered regularly interspaced short palindromic repeats. This system was first identified in *E. coli* (Ishino *et al.*, 1987). The CRISPR locus has been identified in a variety of bacterial species and almost all species of archaea. The locus has a highly conserved structure to it which consists of short repeated sequences that are interspersed with spacer sequences. A study by Jansen *et al.*, in 2002 identified protein-encoding genes, now designated as *cas* (CRISPR-associated) genes, which were adjacent to the locus. However, it was only in 2005 when scientists observed that the spacer regions of the CRISPR locus and sequences of bacteriophage and plasmids had high similarity. Later studies described the CRISPR system as an adaptive immunity against bacteriophage and plasmids, found in a variety of different microorganisms (Bolotin *et al.*, 2005; Mojica *et al.*, 2005; Pourcel, Salvignol and Vergnaud, 2005).

The CRISPR-Cas systems can usually be designated into class 1 or class 2, which are based on the structural organisation of the Cas proteins associated with that system. For systems in class 1, the Cas proteins occur in the form of a multiprotein complex. In contrast, the Cas proteins for class 2 function as a single effector protein. The well-known CRISPR-Cas9 system, originating from *Streptococcus pyogenes* (SpCas9), is in class 2 (Jiang *et al.*, 2013).

Gene editing using CRISPR-Cas9 systems involves two main steps which are DNA restriction followed by DNA repair. The main components involved in these steps are the single-stranded guide RNA (sgRNA) and the Cas9 endonuclease. The sgRNA contains a short complementary sequence called a protospacer, usually 20 bp, which is specific to the target DNA immediately followed by a binding site for specific Cas9 proteins, known as a protospacer adjacent motif (PAM), for example 'NG'. The Cas9 endonuclease, guided by the sgRNA, makes a double-stranded break at the DNA target site (Gasiunas *et al.*, 2012). The double-stranded break triggers the cellular DNA repair mechanisms which can be either homology-directed repair (HDR) or non-homologous end joining (NHEJ). In some bacterial cells, the NHEJ mechanism was found to be a weak system and thus not enough to repair a double-stranded break (Peters *et al.*, 2015). The HDR mechanisms rely on the presence of donor DNA having homology to the regions flanking the cleaved DNA to repair the double-stranded break at high fidelity. The use of CRISPR-Cas9 has been popular due to the ease in engineering the sgRNA to new targets without having to alter the Cas9 nuclease (H. Li *et al.*, 2020).

Variations of the CRISPR-Cas9 system have been developed to improve the technique for a larger variety of organisms. This system has been adapted for use in *Clostridium* species such as *C. saccharoperbutylacetonicum*, *C. acetobutylicum* and *C. beijerinckii* for target gene knock-out (Li *et al.*, 2016; Diallo, Hocq, *et al.*, 2020a; Feng

et al., 2020; Xin *et al.*, 2020). Some examples of Cas9 variants include the dead Cas9 (dCas9), a nuclease-deficient version of Cas9, meaning that no DNA cleavage can occur, instead it is used to regulate transcription of target genes via activation (CRISPRa) or inhibition (CRISPRi) (de la Fuente-Núñez and Lu, 2017). This can be advantageous over the original Cas9 nuclease as it does not rely on cellular repair pathways and therefore makes it more desirable in microbes which are difficult to engineer. The Cas9 nickase (nCas9) variant, generated by creating a D10a mutation in the RuvC domain or a H840A mutation in the HNH domain of Cas9, can only produce single-stranded breaks in the target DNA. This is advantageous as ssDNA breaks are less lethal to the cell compared to dsDNA breaks. Creating a ssDNA nick does not induce the non-homologous end joining repair pathway, which is highly useful in prokaryotes which lack NHEJ expression. Instead, the ssDNA nick induces homologous recombination at a high efficiency (de la Fuente-Núñez and Lu, 2017).

Recent nCas9 tools have been developed for use in Clostridia. One example being developed in acetogens and then demonstrated in *C. beijerinckii* and *C. ljungdahlii*. This involved using an nCas9 fused with an activation-induced cytosine deaminase, which converts cytosine to thymine, and generating an early stop codon in the target gene (Li *et al.*, 2019; Xia *et al.*, 2020). Base editors such as this example are useful as they can target single nucleotides without the need for a DNA double-stranded break. This can allow for minute changes to lead to a

disruption in target protein expression (Seys *et al.*, 2023). Another example of nCas9 tool development is described by Diallo *et al.*, (2020b) who used a dual-plasmid inducible nCas9 system to insert large cargo DNA into the desired genome target site in *C. beijerinckii* NCIMB 8052.

1.11 Project Aims

The central objective of this work is to understand the role of Agr QS in stationary phase phenotypes, primarily sporulation and solvent production, in *C. beijerinckii.* For the solventogenic *Clostridium*, it has been shown that the switch from acidogenesis to solventogenesis and sporulation along with improved ABE productivity occurs at high cell density in batch culture, suggesting QS is involved (S. J. Choi *et al.*, 2012). Both RRNPP-type and Agr QS has been studied in *C. acetobutylicum* ATCC 824 and demonstrated to play a regulatory role in solventogenesis and/or sporulation (Steiner *et al.*, 2012; Kotte *et al.*, 2020). However, whilst putative Agr and RRNPP proteins have been observed in other solventogenic *Clostridia*, currently there is no experimental evidence demonstrating the role of QS in *C. beijerinckii*. This organism has high potential to produce biobutanol from lignocellulosic biomass compared to other solventogenic strains in industry as it demonstrates good tolerance to process inhibitors and pH changes.

This study will concentrate on the Agr systems in *C. beijerinckii* NCIMB 8052 as no complete RRNPP has been found in the species. Furthermore, this strain is a precursor strain for hyper-butanol producing strains (Formanek, Mackie and Blaschek, 1997; Qureshi and Blaschek, 2001; Xin *et al.*, 2020) and is still widely used in industry. Working towards unraveling the potential contribution of cell-cell communications in the regulation of solventogenesis and sporulation will provide a platform of knowledge for future strain engineering strategies and improved culturing conditions. These will aim to maximise butanol production whilst also maintaining strain stability and solvent resistance. It therefore stands that the specific objectives of this PhD are to:

1. Identify putative homologues of the *C. beijerinckii* NCIMB 8052 AgrB and AgrD proteins in *C. beijerinckii* strains and other species in the genus *Clostridium* through bioinformatic analysis.

2. Inactivate Agr signalling systems in *C. beijerinckii* NCIMB 8052 individually or if required combined.

3. Characterise the phenotypic changes in the mutants with emphasis on stationary phase features such as sporulation, solvent production and granulose formation.

4. Identify the system(s) with the most interesting effects and characterise in more detail with respect to:
   a- AIP chemical characterisation
   b- Expression profile of different genes
   c- The regulatory mechanisms of differentially expressed genes.

Chapter 2: Materials and Methods

2.1 List of bacterial strains, plasmids, oligonucleotides and peptides used.

Table 2.1 List of genome and sequencing information for *C. beijerinckii* strains analysed for agr loci.

Strain	Assembly (Genbank)	Release Date	Sequencing Platform	Sequence size	Genome representation	Genome coverage
4J9	GCA_002006135.1	2017	Illumina	5,888,124	FULL	145x
ATCC 39058	GCA_002006125.1	2017	Illumina	5,953,339	FULL	233x
BAS/B2	GCA_002006485.1	2017	Illumina	5,982,920	FULL	307x
BAS/B3/I/124	GCA_002003345.1	2017	Sanger dideoxy sequencing; 454; Illumina	6,123,550	FULL	123x
BGS1	GCA_001685175.1	2016	Illumina HiSeq	5,880,896	FULL	90x
BIOML-A11	GCA_009876465.1	2020	Illumina NextSeq	6,176,719	FULL	127.482x
Br21	GCA_002009885.1	2017	Illumina HiSeq	5,992,945	FULL	230x
DJ032	GCA_013294385.1	2020	PacBio	5,607,524	FULL	210.0x

DJ033	GCA_013149505.1	2020	PacBio	6,065,472	FULL	72x
DJ051	GCA_013149485.1	2020	PacBio	6,260,681	FULL	113x
DJ076	GCA_013149515.1	2020	PacBio	6,093,716	FULL	286x
DJ077	GCA_013149565.1	2020	PacBio	6,103,722	FULL	75x
DJ078	GCA_013149405.1	2020	PacBio	6,477,311	FULL	152x
DJ091	GCA_013149425.1	2020	PacBio	6,126,292	FULL	178x
DSM 53	GCA_002006205.1	2017	Illumina	5,773,247	FULL	225x
DSM 791	GCA_002006445.1	2017	Illumina	5,781,472	FULL	131x
G117	GCA_000280535.1	2012	Illumina HiSeq 2000	5,811,816	FULL	90x
HUN142	GCA_000621745.1	2014	Illumina HiSeq 2000	6,106,710	FULL	unknown
NBRC 109359	GCA_007992515.1	2019	Illumina HiSeq 1000	5,660,510	FULL	101x
NCIMB 14988	GCA_000833105.2	2016	PacBio RSII; 2xSMRT; Illumina	6,485,394	FULL	500x
NCIMB 8052	GCA_000016965.1	2007		6,000,632	FULL	N/A

NCP 260	GCA_002006285.1	2017	Illumina	5,968,330	FULL	228x
NCTC13035	GCA_900447025.1	2018		6,010,505	FULL	100x
NRRL B-528	GCA_002006405.1	2017	Illumina	6,255,488	FULL	68x
NRRL B-591	GCA_002006295.1	2017	Illumina	5,874,824	FULL	264x
NRRL B-593	GCA_002006325.1	2017	Illumina	6,156,662	FULL	192x
NRRL B-596	GCA_002006435.1	2017	Illumina	6,220,133	FULL	152x
NRRL B-598	GCA_000506785.4	2018	454; PacBio	6,186,993	FULL	100x
SA-1 ATCC 35702	GCA_000767745.1	2014	Sanger dideoxy sequencing; Illumina	5,999,050	FULL	914x
WB	GCA_002954125.1	2018	Illumina HiSeq	5,781,106	FULL	100x
WB01_NA02	GCA_012843965.1	2020	Illumina MiSeq	5,627,834	FULL	300x
WB53	GCA_003129525.1	2018	PacBio	4,287,191	FULL	99.98x
Clostridium diolis DSM	GCA_008705175.1	2019	PacBio RSII; Illumina NextSeq	5,940,808	FULL	700x

15410						
Clostridium diolis NJP7	GCA_002176895.1	2017	Illumina GAII	5,756,492	FULL	150x
Clostridium diolis WST	GCA_003015255.1	2018	Illumina HiSeq	5,852,174	FULL	0.9999x

All genome sequence data was downloaded from the *National Center for Biotechnology Information (NCBI*, 2020) website.

Table 2.2 List of putative *agrB* and *agrD*-precursor peptides sequence information from *C. beijerinckii* NCIMB 8052 (genbank accession: CP000721.1).

Putative protein	GenBank Accession/ AA sequence	Locus tag (old locus tag)
AgrB (1)	ABR32458.1 / WP_011967620.1	CBEI_RS26420 (Cbei_0269)
AgrB (2)	ABR32845.1 / WP_011968006.1	CBEI_RS03605 (Cbei_0658)
AgrB (3)	ABR32847.1 / WP_011968008.1	CBEI_RS03615 (Cbei_0660)
AgrB (4)	ABR35303.1 / WP_012059354.1	CBEI_RS16315 (Cbei_3171)
AgrB (5)	ABR36075.1 / WP_012060123.1	CBEI_RS20300 (Cbei_3965)
AgrB (6)	ABR36687.1	CBEI_RS23365 (Cbei_4578)
AgrD-precursor (1)	WP_085953352.1	CBEI_RS26850
AgrD-precursor* (2)	MKIKNLLGKVLMMLSLSMIIFAPASAATAGIEEMPESMKKLR	N/A
AgrD-precursor* (3)	MIKKLIGLLLEKMSNSMVTLAPASFARFGVEDMPESMKKLR	N/A
AgrD-	WP_077836881.1	CBEI_RS26650

precursor (4)		
AgrD-precursor (5)	WP_085953365.1	CBEI_RS26925
AgrD-precursor (6)	WP_077829127.1	CBEI_RS26810
AgrD-precursor* (7)	MGIKKMIANLLLNLSKSMTKSPMVSYCFAGIEDMPESMKKLR	N/A

Sequences were obtained from the NCBI online database and used as the query sequences when using the NCBI database to find AgrBD homologues. * Indicates AA sequences were sourced through tBLASTn search and have no associated accession number/ locus tag.

Table 2.3 List of bacterial strains used.

Organism	Description	Source
Escherichia coli Top10	*E. coli* strain used for cloning	Invitrogen
Escherichia coli dh5α	*E. coli* strain used for cloning	
Clostridium beijerinckii NCIMB 8052	The wild type *C. beijerinckii* NCIMB 8052 strain (Culture collection 19 yoong strain)	NCIMB
Clostridium beijerinckii NCIMB 8052 *agrD5*::CTermB	The *C. beijerinckii* NCIMB 8052 *agrD5 (CBEI_RS26925)* ClosTron mutant.	This Study
Clostridium beijerinckii NCIMB 8052 *ΔpyrE*	The *C. beijerinckii* NCIMB 8052 mutant with a truncated *pyrE* gene (Cbei_1006).	JRH/This study
Clostridium beijerinckii NCIMB 8052 *ΔpyrE agrB2* KO	The *C. beijerinckii* NCIMB 8052 mutant with a truncated *pyrE* gene (Cbei_1006) and disrupted *agrB2* gene (CBEI_RS03605/Cbei_0658).	This Study
Clostridium beijerinckii NCIMB 8052 *ΔpyrE ΔagrD4*	The *C. beijerinckii* NCIMB 8052 mutant with a truncated *pyrE* gene (Cbei_1006) and disrupted *agrD4* gene (CBEI_RS26650).	This Study
Clostridium beijerinckii NCIMB 8052 *ΔpyrE agrB4* KO	The *C. beijerinckii* NCIMB 8052 mutant with a truncated *pyrE* gene (Cbei_1006) and disrupted *agrB4* gene (CBEI_RS16315/Cbei_3171).	This Study
Clostridium beijerinckii NCIMB 8052 *ΔpyrE ΔagrD5*	The *C. beijerinckii* NCIMB 8052 mutant with a truncated *pyrE* gene (Cbei_1006) and disrupted *agrD5* (CBEI_RS26925) gene.	This Study

Clostridium beijerinckii NCIMB 8052 *ΔpyrE agrB5* KO	The *C. beijerinckii* NCIMB 8052 mutant with a truncated *pyrE* gene (Cbei_1006) and disrupted *agrB5* gene (CBEI_RS20300/Cbei_3965).	This Study
Clostridium beijerinckii NCIMB 8052 *agrB2* KO	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006) and disrupted *agrB2* gene (CBEI_RS03605/Cbei_0658).	This Study
Clostridium beijerinckii NCIMB 8052 *agrB2* KO_comp	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), a disrupted *agrB2 gene* (CBEI_RS03605/Cbei_0658) which is complemented at the *pyrE* locus.	This Study
Clostridium beijerinckii NCIMB 8052 *agrD4* KO	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006) and disrupted *agrD4* gene (CBEI_RS26650).	This Study
Clostridium beijerinckii NCIMB 8052 *agrD4* KO_comp	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), disrupted *agrD4* gene (CBEI_RS26650) and a complemented *agrD4* gene at the *pyrE* locus.	This Study
Clostridium beijerinckii NCIMB 8052 *agrB4* KO	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006) and disrupted *agrB4* (CBEI_RS16315/Cbei_3171) gene.	This Study
Clostridium beijerinckii NCIMB 8052 *agrB4* KO_comp	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), a disrupted *agrB4* gene (CBEI_RS16315/Cbei_3171) and a complemented *agrB4* gene at the *pyrE* locus.	This Study
Clostridium beijerinckii NCIMB 8052 *agrD5* KO	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006) and disrupted *agrD5* (CBEI_RS26925) gene.	This Study

Clostridium beijerinckii NCIMB 8052 *agrD5* KO_comp	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), a disrupted *agrD5 (CBEI_RS26925)* gene and a complemented *agrD5* gene at the *pyrE* locus.	This Study
Clostridium beijerinckii NCIMB 8052 *agrB5* KO	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006) and disrupted *agrB5* gene (CBEI_RS20300/Cbei_3965).	This Study
Clostridium beijerinckii NCIMB 8052 *agrB5*_comp	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), a disrupted *agrB5* gene (CBEI_RS20300/Cbei_3965) and a complemented *agrB5* gene at the *pyrE* locus.	This Study
Clostridium beijerinckii NCIMB 8052 *ΔpyrE agrB2agrB4* KO	The *C. beijerinckii* NCIMB 8052 with a truncated *pyrE* gene (Cbei_1006), *agrB2* gene (CBEI_RS03605/Cbei_0658) and a disrupted *agrB4* (CBEI_RS16315/Cbei_3171) gene.	This Study
Clostridium beijerinckii NCIMB 8052 *agrB2agrB4* KO	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), *agrB2* gene (CBEI_RS03605/Cbei_0658) and a disrupted *agrB4* (CBEI_RS16315/Cbei_3171) gene.	This Study
Clostridium beijerinckii NCIMB 8052 *agrB2agrB4* KO_comp B2	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), *agrB2* gene (CBEI_RS03605/Cbei_0658), a disrupted *agrB4* (CBEI_RS16315/Cbei_3171) gene, and a complemented *agrB2* gene at the *pyrE* locus.	This Study

Clostridium beijerinckii NCIMB 8052 *agrB2agrB4* KO_comp B4	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), *agrB2* gene (CBEI_RS03605/Cbei_0658), a disrupted *agrB4* (CBEI_RS16315/Cbei_3171) gene, and a complemented *agrB4* gene at the *pyrE* locus.	This Study
Clostridium beijerinckii NCIMB 8052 *ΔpyrE agrB2agrB5* KO	The *C. beijerinckii* NCIMB 8052 with a truncated *pyrE* gene (Cbei_1006), *agrB2* gene (CBEI_RS03605/Cbei_0658) and a disrupted *agrB5* gene (CBEI_RS20300/Cbei_3965).	This Study
Clostridium beijerinckii NCIMB 8052 *agrB2agrB5*	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), *agrB2* gene (CBEI_RS03605/Cbei_0658) and a disrupted *agrB5* gene (CBEI_RS20300/Cbei_3965).	This Study
Clostridium beijerinckii NCIMB 8052 *agrB2agrB5* KO_comp B2	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), *agrB2* gene (CBEI_RS03605/Cbei_0658), a disrupted *agrB5* gene (CBEI_RS20300/Cbei_3965), and a complemented *agrB2* gene at the *pyrE* locus.	This Study
Clostridium beijerinckii NCIMB 8052 *agrB2agrB5* KO_comp B5	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), *agrB2* gene (CBEI_RS03605/Cbei_0658), a disrupted *agrB5* gene (CBEI_RS20300/Cbei_3965), and a complemented *agrB5* gene at the *pyrE* locus.	This Study

Clostridium beijerinckii NCIMB 8052 *ΔpyrE agrB4agrB5* KO	The *C. beijerinckii* NCIMB 8052 with a truncated *pyrE* gene (Cbei_1006), *agrB4* gene (CBEI_RS16315/Cbei_3171) and a disrupted *agrB5* gene (CBEI_RS20300/Cbei_3965).	This Study
Clostridium beijerinckii NCIMB 8052 *agrB4agrB5* KO	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), *agrB4* gene (CBEI_RS16315/Cbei_3171) and a disrupted *agrB5* gene (CBEI_RS20300/Cbei_3965).	This Study
Clostridium beijerinckii NCIMB 8052 *agrB4agrB5* KO_comp B4	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), *agrB4* gene (CBEI_RS16315/Cbei_3171) and a disrupted *agrB5* gene (CBEI_RS20300/Cbei_3965), and a complemented *agrB4* gene at the *pyrE* locus.	This Study
Clostridium beijerinckii NCIMB 8052 *agrB4agrB5* KO_comp B5	The *C. beijerinckii* NCIMB 8052 with a repaired *pyrE* gene (Cbei_1006), *agrB4* gene (CBEI_RS16315/Cbei_3171) a disrupted *agrB5* gene (CBEI_RS20300/Cbei_3965), and a complemented *agrB5* gene at the *pyrE* locus.	This Study

Table 2.4 List of plasmids used.

Plasmid	Properties	Source
pMTL83251	*Clostridium* modular plasmid containing a pCB102 Gram-positive replicon, a *ermB* marker, ColE1 + TraJ Gram-negative replicons and a *lacZ* multiple cloning site.	Heap et al. 2009
pMTL-JRH1	*pyrE* truncation vector containing a pCB102 Gram-positive replicon, a *ermB* marker, ColE1 + TraJ Gram-negative replicons and a *lacZ* multiple cloning site containing the first 300bp of the *pyrE* gene and 1200bp downstream of the *pyrE* gene from the *C. beijerinckii* NCIMB 8052 genome.	JRH
pMTL-JRH3	Gene knock-out vector containing a pCB102 Gram-positive replicon, a *ermB* marker, ColE1 + TraJ Gram-negative replicons and a *lacZ* multiple cloning site.	JRH
pMTL84351	*Clostridium* modular plasmid containing a pCD6 Gram-positive replicon, an *aad9* marker, ColE1 + TraJ Gram-negative replicons and a *lacZ* multiple cloning site.	Heap et al. 2009
pMTL-JRH4	*pyrE* repair vector containing a pCD6 Gram-positive replicon, an *aad9* marker, ColE1 + TraJ Gram-negative replicons and a *lacZ* multiple cloning site containing the intact *pyrE* gene and 1200bp downstream of the *pyrE* gene from *C. beijerinckii* NCIMB 8052 genome.	JRH
pMTL007S-E2::CBEI_agrD5-79|80a	ClosTron plasmid targeting the agrD5 gene in the C. beijerinckii NCIMB 8052 genome. Containing a pCB102 Gram-positive replicon, a *aad9* marker and a FRT-flanked *ermB* RAM intron marker.	ATUM
pMTL007S-E2::CBEI_agrD2-65|66a	ClosTron plasmid targeting the agrD2 gene in the C. beijerinckii NCIMB 8052 genome. Containing a pCB102 Gram-positive replicon, a *aad9* marker and a FRT-flanked *ermB* RAM intron marker.	ATUM
pMTL-JRH3_agrD2 KO	Gene knock-out vector containing a pCB102 Gram-positive replicon, a *ermB* marker, ColE1 + TraJ Gram-negative replicons, a *lacZ* multiple cloning site containing the first 9bp and 750 bp upstream and the last 9bp and 750 bp downstream of the *agrD2* gene from *C. beijerinckii* NCIMB 8052 genome.	This Study
pMTL-JRH3_agrD5 KO	Gene knock-out vector containing a pCB102 Gram-positive replicon, a *ermB* marker, ColE1 + TraJ Gram-negative replicons, a *lacZ* multiple cloning site containing the first 45bp and 750 bp upstream and the last 27bp and 750 bp downstream of the *agrD5* gene from *C. beijerinckii* NCIMB 8052 genome.	This Study

pMTL-ACE suicide_agrD5 KO	Gene knock-out vector containing a *ermB* marker, ColE1 + TraJ Gram-negative replicons, a *lacZ* multiple cloning site containing the first 45bp and 750 bp upstream and the last 27bp and 750 bp downstream of the *agrD5* gene from *C. beijerinckii* NCIMB 8052 genome.	This Study
pMTL-ACE suicide_agrD2 KO	Gene knock-out vector containing a *ermB* marker, ColE1 + TraJ Gram-negative replicons, a *lacZ* multiple cloning site containing the first 9bp and 750 bp upstream and the last 9bp and 750 bp downstream of the *agrD2* gene from *C. beijerinckii* NCIMB 8052 genome.	This study
vFS77	NG-AID with PfdxE translational repression system with a *catP* marker. Expected to be toxic once expressed.\nPfdxE_nCas9_AID_LacIq_ColE1+tra_ParaE_repH(pCB102)\n	Francois Seys
vLRB1	NG-AID with PfdxE translational repression system with a *ermB* marker. Expected to be toxic once expressed.\nPfdxE_nCas9_AID_LacIq_ColE1+tra_ParaE_repH(pCB102)\n	This Study
vLRB1-agrD5	NG-AID with PfdxE translational repression system with a *ermB* marker. Expected to be toxic once expressed.\nPfdxE_nCas9_AID_LacIq_ColE1+tra_ParaE_agrD5sgRNA_repH(pCB102)\n; target agrD5 (CBEI_RS26925)	This Study
vLRB1-agrB5	NG-AID with PfdxE translational repression system with a *ermB* marker. Expected to be toxic once expressed.\nPfdxE_nCas9_AID_LacIq_ColE1+tra_ParaE_agrB5sgRNA_repH(pCB102)\n; target *agrB5* (CBEI_RS20300/Cbei_3965)	This Study
vLRB1-agrD4	NG-AID with PfdxE translational repression system with a *ermB* marker. Expected to be toxic once expressed.\nPfdxE_nCas9_AID_LacIq_ColE1+tra_ParaE_agrD4sgRNA_repH(pCB102)\n; target agrD4 (CBEI_RS26650)	This Study
vLRB1-agrB4	NG-AID with PfdxE translational repression system with a *ermB* marker. Expected to be toxic once expressed.\nPfdxE_nCas9_AID_LacIq_ColE1+tra_ParaE_agrB4sgRNA_repH(pCB102)\n; target agrB4 (CBEI_RS16315/Cbei_3171)	This Study
vLRB1-agrB2	NG-AID with PfdxE translational repression system with a *ermB* marker. Expected to be toxic once expressed.\nPfdxE_nCas9_AID_LacIq_ColE1+tra_ParaE_agrB2sgRNA_repH(pCB102)\n; target agrB2 (CBEI_RS03605/Cbei_0658)	This Study
pMTL-JRH4-agrB4	pMTL-JRH4 complementation vector containing the *agrB4* (CBEI_RS16315/Cbei_3171) gene plus 200 bp upstream for the promoter region.	This Study

pMTL-JRH4-agrB5	pMTL-JRH4 complementation vector containing the agrB5 (CBEI_RS20300/Cbei_3965) gene plus 200 bp upstream for the promoter region.	This Study
pMTL-JRH4-agrD4	pMTL-JRH4 complementation vector containing the agrD4 (CBEI_RS26650) gene plus 200 bp upstream for the promoter region.	This Study
pMTL-JRH4-agrD5	pMTL-JRH4 complementation vector containing the agrD5 (CBEI_RS26925) gene plus 200 bp upstream for the promoter region.	This Study
pMTL-JRH4-agrB2	pMTL-JRH4 complementation vector containing the agrB2 (CBEI_RS03605/Cbei_0658) gene plus 200 bp upstream for the promoter region.	This Study

Table 2.5 List of oligonucleotides used.

Oligonucleotide	Purpose	Sequence 5'-3'	Source
Construct primers			
agrDG2 KO LHA_F	Generation of the LHA for *agrD2* KO using HiFi assembly into pMTL-JRH3	CAGGAAACAGCTATGACCGCTAATCATCTAATTAATGTTGGAAAAATATTTTTAC	This Study
agrDG2 KO LHA_R	Generation of the LHA for *agrD2* KO using HiFi assembly into pMTL-JRH3	CTTATCTCAAAAAATCCATTCCCCTTTCAATTTTG	This Study
agrDG2 KO RHA_F	Generation of the RHA for *agrD2* KO using HiFi assembly into pMTL-JRH3	AATGGATTTTTTGAGATAAGTCACTTTCACG	This Study
agrDG2 KO RHA_R	Generation of the RHA for *agrD2* KO using HiFi assembly into pMTL-JRH3	CATGGTCATATGGATACAGCTGTGCTCCGTAGTCATGC	This Study
agrDG4 CBEI_RS26650 KO LHA_F	Generation of the 1000bp LHA for *agrD4* (CBEI_RS26650) KO using HiFi assembly into pMTL-JRH3	CAGGAAACAGCTATGACCGCGCCTTAAAATACATAATTGAGGGAATG	This Study
agrDG4 CBEI_RS26650 KO LHA_R	Generation of the 1000bp LHA for *agrD4* (CBEI_RS26650) KO using HiFi assembly into pMTL-JRH3	ATTAATAATTTAAAATATTGTTCATTAATTTCATCTCCTC	This Study
agrDG4 CBEI_RS26650 KO RHA_F	Generation of the 1000bp RHA for *agrD4* (CBEI_RS26650) KO using HiFi assembly into pMTL-JRH3	GAACAATATTTTAAATTATTAATATCAATGTTAAAAGTAAAAATTGTC	This Study
agrDG4 CBEI_RS26650 KO RHA_R	Generation of the 1000bp RHA for *agrD4* (CBEI_RS26650) KO using HiFi assembly into pMTL-JRH3	CATGGTCATATGGATACAGCTTTTCAAATAACATATCATAACAAATTTTATTATTTAATAATAG	This Study

agrDG5 CBEI_RS26925 KO LHA_F	Generation of the 1000bp LHA for *agrD5* (CBEI_RS26925) KO using HiFi assembly into pMTL-JRH3	CAGGAAACAGCTATGACCGCAAATAAATAGTGTTTCCTCTAAACATAC	This Study
agrDG5 CBEI_RS26925 KO LHA_R	Generation of the 1000bp LHA for *agrD5* (CBEI_RS26925) KO using HiFi assembly into pMTL-JRH3	CAGGATAATCTAACTTAAGAAGTATAGCTGAAAAAATG	This Study
agrDG5 CBEI_RS26925 KO RHA_F	Generation of the 1000bp RHA for *agrD5* (CBEI_RS26925) KO using HiFi assembly into pMTL-JRH3	TCTTAAGTTAGATTATCCTGAAGAATTATTGAAATAAATATC	This Study
agrDG5 CBEI_RS26925 KO RHA_R	Generation of the RHA for *agrD5* (CBEI_RS26925) KO using HiFi assembly into pMTL-JRH3	CATGGTCATATGGATACAGCTTGGTATTATGATAACATAGTAATTTTAAAC	This Study
ermB_fwd vFS77 hifi	Hifi oligonucleotide containing the FseI restriction site to generate the *ermB* gene from pMTL-83251 modular plasmid for insertion into vFS36, vFS77 or vFS100.	TATGGATTATAAGCGGCCGGCCGAAGCAAACTTAAGAGTGTG	This Study
ermB_rev vFS77 hifi	Hifi oligonucleotide containing the HpaI restriction site to generate the *ermB* gene from pMTL-83251 modular plasmid for insertion into vFS36, vFS77 or vFS100.	GTTACAGACAAACCTGAGTT*AAC*AAGCAGCAGATTACGCGC	This Study
agrB5 oligo 1	Produce sgRNA for *agrB5*, Hifi fragment insertion into vFS77 vecor.	ATCTTAAGGAGGAGTTTTCGTCGACAATTCAATATACATTAGAAGGTTTTAGAGCTAGAAATAGCAAGTT	This Study

agrB5 oligo 2	Produce sgRNA for *agrB5*, Hifi fragment insertion into vFS77 vecor.	AACTTGCTATTTCTAGCTCTAAAACCTTCTAATGTATATTGAATTGTCGACGAAAACTCCTCCTTAAGAT	This Study
agrD5 oligo 1	Produce sgRNA for *agrD5*, Hifi fragment insertion into vFS77 vecor.	ATCTTAAGGAGGAGTTTTCGTCGACCCCCATACGCCTGCACAGCTGTTTTAGAGCTAGAAATAGCAAGTT	This Study
agrD5 oligo 2	Produce sgRNA for *agrD5*, Hifi fragment insertion into vFS77 vecor.	AACTTGCTATTTCTAGCTCTAAAACAGCTGTGCAGGCGTATGGGGGTCGACGAAAACTCCTCCTTAAGAT	This Study
agrB2 oligo 1	Produce sgRNA for *agrB2*, Hifi fragment insertion into vFS77 vecor.	ATCTTAAGGAGGAGTTTTCGTCGACGAACAAATGGAATATATTCTGTTTTAGAGCTAGAAATAGCAAGTT	This Study
agrB2 oligo 2	Produce sgRNA for *agrB2*, Hifi fragment insertion into vFS77 vecor.	AACTTGCTATTTCTAGCTCTAAAACAGAATATATTCCATTTGTTCGTCGACGAAAACTCCTCCTTAAGAT	This Study
agrD4 oligo 1	Produce sgRNA for *agrD4*, Hifi fragment insertion into vFS77 vecor.	ATCTTAAGGAGGAGTTTTCGTCGACCCCCAACCACATGCTGACGCGTTTTAGAGCTAGAAATAGCAAGTT	This Study
agrD4 oligo 2	Produce sgRNA for *agrD4*, Hifi fragment insertion into vFS77 vecor.	AACTTGCTATTTCTAGCTCTAAAACGCGTCAGCATGTGGTTGGGGGTCGACGAAAACTCCTCCTTAAGAT	This Study
agrB4 oligo 1	Produce sgRNA for *agrB4*, Hifi fragment insertion into vFS77 vecor.	ATCTTAAGGAGGAGTTTTCGTCGACGAACAGATAATCGTATATGGGTTTTAGAGCTAGAAATAGCAAGTT	This Study
agrB4 oligo 2	Produce sgRNA for *agrB4*, Hifi fragment insertion into vFS77 vecor.	AACTTGCTATTTCTAGCTCTAAAACCCATATACGATTATCTGTTCGTCGACGAAAACTCCTCCTTAAGAT	This Study

agrB5 + 200bp up_fwd	Produce the HiFi fragment for *agrB5* + 200 bp upstream, containing SalI restriction site, for insertion into JRH4 complementation vector.	TACCCGGGGATCCTCTAGAG*TCGAC*ATTACTAAACCTAAATAAATAGTGTTTC	This Study
agrB5 + 200bp up_rev	Produce the HiFi fragment for *agrB5* + 200 bp upstream, containing SalI restriction site, for insertion into JRH4 complementation vector.	ATCTCCATGGACGCGTGACG*TCGAC*TTAAGAAGTATAGCTGAAAAAATG	This Study
agrB4 + 200bp up_fwd	Produce the HiFi fragment for *agrB4* + 200 bp upstream, containing SalI restriction site, for insertion into JRH4 complementation vector.	TACCCGGGGATCCTCTAGAG*TCGAC*AAAGATAAAATAGCAATTAATATAGAAAGTC	This Study
agrB4 + 200bp up_rev	Produce the HiFi fragment for *agrB4* + 200 bp upstream, containing SalI restriction site, for insertion into JRH4 complementation vector.	ATCTCCATGGACGCGTGACG*TCGAC*TTACTTTGCCTTTGGTGTC	This Study
agrB2 + 200bp up_fwd	Produce the HiFi fragment for *agrB2* + 200 bp upstream, containing SalI restriction site, for insertion into JRH4 complementation vector.	TACCCGGGGATCCTCTAGAG*TCGAC*ATTAAAGAAAAAGCATCCTATTAATAG	This Study
agrB2 + 200bp up_rev	Produce the HiFi fragment for *agrB2* + 200 bp upstream, containing SalI restriction site, for insertion into JRH4 complementation vector.	ATCTCCATGGACGCGTGACG*TCGAC*TTAAGCATCTTTTTTATTAAACATTAATATG	This Study

Sequencing primers			
NEW JRH3 insertion check_F	sequencing primer for successful insertion into the pMTL JRH3 vector MCS. Also used to check for RHA single cross over isolates during ACE.	AGCTGGTGAAGTACATCAC	This Study
JRH3 insertion check_R	sequencing primer for successful insertion into the pMTL JRH3 vector MCS. Also used to check for LHA single cross over isolates during ACE.	TCTATTCAGCACTGTTATGC	This Study
C-agrDG2 Cross LHA Check_F	Confirm both double cross over isolates and single cross over on LHA of *agrD2* gene in *C. beijerinckii* NCIMB 8052	AATACCGATTTTTGCATTAGT	This Study
C-agrDG2 Cross RHA Check_R	Confirm both double cross over isolates and single cross over on RHA of *agrD2* gene in *C. beijerinckii* NCIMB 8052	AACTTCTACAGCTGTTGGC	This Study
C-agrDG5 cross LHA Check_F	Confirm both double cross over isolates and single cross over on LHA of *agrD5* CBEI_RS26925 gene in *C. beijerinckii* NCIMB 8052	ACCTCTCCCCTCTTAATTAC	This Study
C-agrDG5 cross RHA Check_R	Confirm both double cross over isolates and single cross over on RHA of *agrD5* CBEI_RS26925 gene in *C. beijerinckii* NCIMB 8052	ATAGTCAAGGAAATTTGTGTG	This Study
16S forward	16s sequencing primer for Clostridia	GCGAGAGTTTGATCCTGGCTCAG	This Study
16S reverse	16s sequencing primer for Clostridia	CGCGGTTACCTTGTTACGACTT	This Study
agrD2 KO confirmation_F	Confirm the truncation of the *agrD2* gene in *C. beijerinckii* NCIMB 8052	AGAATAAGAGGTCTAAGTATTG	This Study

agrD2 KO confirmation_ R	Confirm the truncation of the *agrD2* gene in *C. beijerinckii* NCIMB 8052.	TTCCAAATATACTTGTAA AGC	This Study
agrB2 CRISPR KO confirmation F1	Confirm the mutation of the *agrB2* (CBEI_RS03605/Cbei _0658) gene in *C. beijerinckii* NCIMB 8052.	ATAATCATCTAATTAATG TTGG	This Study
agrD4 KO confirmation_ F	Confirm the truncation of the *agrD4* CBEI_RS26650 gene in *C. beijerinckii* NCIMB 8052.	AGTGGCTTAAATATAGAT ACAG	This Study
agrD4 KO confirmation_ R	Confirm the truncation of the *agrD4* CBEI_RS26650 gene in *C. beijerinckii* NCIMB 8052.	TCATATGTATTATTATGC TACTC	This Study
agrB4 CRISPR KO confirmation R1	Confirm the mutation of the *agrB4* (CBEI_RS16315/Cbei _3171) gene in *C. beijerinckii* NCIMB 8052.	ATGATTCTGGAATAAGTC	This Study
agrB4 CRISPR KO confirmation F1	Confirm the mutation of the *agrB4* (CBEI_RS16315/Cbei _3171) gene in *C. beijerinckii* NCIMB 8052.	TCAATTTGATTTTTCTGT TC	This Study
agrD5 KO confirmation_ F	Confirm the mutation of the *agrD5* CBEI_RS26925 gene in *C. beijerinckii* NCIMB 8052.	TCTTTAATATCGCTAACT TTTTG	This Study
agrD5 KO confirmation_ R	Confirm the mutation of the *agrD5* CBEI_RS26925 gene in *C. beijerinckii* NCIMB 8052.	AGATGGATTACCCTGTA G	This Study
agrB5 KO confirmation_ F	Confirm the mutation of the *agrB5* gene in *C. beijerinckii* NCIMB 8052.	ATTAGTTGGTATTATGAT AACATAG	This Study

agrB5 KO confirmation_ R	Confirm the mutation of the *agrB5* gene in *C. beijerinckii* NCIMB 8052.	AATCAAGTAGTAAATCCT AGTG	This Study
agrD1 KO confirmation_ R	Confirm the truncation of the *agrD1* CBEI_RS26850 gene in *C. beijerinckii* NCIMB 8052	AATGAAACATCTCATTTA ATAG	This Study
agrB1 KO confirmation_ F	Confirm the mutation of the *agrB1* gene in *C. beijerinckii* NCIMB 8052.	CCACTTTCATGTTAACAC	This Study
agrD3 KO confirmation_ R	Confirm the truncation of the *agrD3* gene region in *C. beijerinckii* NCIMB 8052	GGTTGATGATCAAAATCT TC	This Study
agrB3 KO confirmation_ F	Confirm the mutation of the *agrB3* gene in *C. beijerinckii* NCIMB 8052.	TGAAGGATTTTTAGTGGT G	This Study
agrD6 KO confirmation_ R	Confirm the truncation of the *agrD6* region CBEI_RS26810 gene in *C. beijerinckii* NCIMB 8052	AATGATTACAAGTCATTC ATAG	This Study
agrB6 KO confirmation_ F	Confirm the mutation of the *agrB6* gene in *C. beijerinckii* NCIMB 8052.	CTTGTAGGATTATTTTGT AC	This Study
pyrE KO confirmation_ F	Confirm the truncation of the *pyrE* gene in *C. beijerinckii* NCIMB 8052.	CAAAGTAGATATTGGAG GAATT	This Study
pyrE KO confirmation_ R	Confirm the truncation of the *pyrE* gene in *C. beijerinckii* NCIMB 8052.	AACACTTGCAGTCTTATG AAC	This Study
oFS073_sgRN A3.0_F	sequencing primer to check sgRNA region of CRISPR NG plasmids.	CTAGATTTATATTTAGTC CCTTGCCTTGC	This Study

oFS074_pCB102_R1	sequencing primer to check sgRNA region of CRISPR NG plasmids.	CTGTTATGCCTTTTGACTATC	This Study
8052 LAC_F	Confirm the isolates were *C. beijerinckii* NCIMB 8052 by amplification of the Cbei_1632 (DNA methylase) gene and not contaminant strains.	AGAGTGCTGTGAACTAATGGA	This Study
8052 LAC_R	Confirm the isolates were *C. beijerinckii* NCIMB 8052 by amplification of the Cbei_1632 (DNA methylase) gene and not contaminant strains.	TACCTCCGTGTTCTTTCGTT	This Study
Intact pyrE_F	Confirm the presence of *pyrE* gene from *C. beijerinckii* NCIMB 8052 in the pMTL-JRH4 plasmid MCS.	CATCAAGAAGAGCGACTT	This Study
Intact pyrE_R	Confirm the presence of pyrE gene from C. beijerinckii NCIMB 8052 in the pMTL-JRH4 plasmid MCS.	GAATTCGTAATCATGGTCATA	This Study
CBEI_RS12555|pepT FWD Set 1	qPCR primer for pepT gene expression.	TCCAGGGTCTGCAAAGAATAC	This Study
CBEI_RS12555|pepT REV Set 1	qPCR primer for pepT gene expression.	CCTTCTGTATGCTCAGGTCTTT	This Study
CBEI_RS08920|spo0A FWD Set 1	qPCR primer for spo0A gene expression.	ATCAGCAGTAGGCCAAGATAAG	This Study
CBEI_RS08920|spo0A REV Set 1	qPCR primer for spo0A gene expression.	GAACTGCTGCAGAGCTATTAAAC	This Study
CBEI_RS10750|CotJA FWD Set 2	qPCR primer for cotJA gene expression.	GTACAGACGCGACGATGATG	This Study
CBEI_RS10750|CotJA REV Set 2	qPCR primer for cotJA gene expression.	ACGAAACCTTCTTTGCAAGAATAC	This Study

CBEI_RS24005|PTS IIA FWD Set 1	qPCR primer for PTS IIA gene expression.	GCGAAGTTCCTGATCCAGTAT	This Study
CBEI_RS24005| PTS IIA REV Set 1	qPCR primer for PTS IIA gene expression.	AGCGTGTCCACTATCCATTATT	This Study
CBEI_RS05390|pflB FWD Set 1	qPCR primer for *pflB* gene expression.	AGACTCACAACGAAGGTGTTT	This Study
CBEI_RS05390|pflB REV Set 1	qPCR primer for *pflB* gene expression.	GCATCTGGTAGGCCTGTTAAT	This Study
CBEI_RS08915|spoIVB FWD Set 1	qPCR primer for *spoIVB* gene expression.	GGAAATCCAGTCGGAGTTAGAG	This Study
CBEI_RS08915|spoIVB REV Set 1	qPCR primer for *spoIVB* gene expression.	AGCCTTACCAGGACTTTCTTC	This Study
CBEI_RS01685|adhE REV Set 1	qPCR primer for *adhE* gene expression.	TCCATCCTGCATTTGGGTATT	This Study
CBEI_RS01685|adhE FWD Set 1	qPCR primer for *adhE* gene expression.	CACCATGTACCTCACGGAATAG	This Study
CBEI_RS05910|sigK FWD Set 1	qPCR primer for *sigK* gene expression.	TCGATGCAGGTAAGGGTACTA	This Study
CBEI_RS05910|sigK REV Set 1	qPCR primer for *sigK* gene expression.	CTCCGATTGGATCTTGGAGATAAA	This Study

2.2 Microbiological methods

2.2.1 Growth Media

2.2.1.1 Lysogeny Broth (LB)

For growth of *E. coli*, LB media was used which consisted of 5 g/l yeast extract, 5 g/l Sodium Chloride, 10 g/l tryptone made to pH 7.2 using HCl or NaOH. For LB agar, 15 g/l agar was added.

2.2.1.2 2 x YTG

For electro-transformations *C. beijerinckii* NCIMB 8052 was cultured in liquid 2 × YTG. The 2 × YTG liquid medium consisted of 16 g/l tryptone, 10 g/l yeast extract and 5 g/l NaCl. This was made up to a volume of 800 - 983.4 ml (depending on desired final glucose concentration) with deionised water and then the pH made to pH 7.4 using HCl or NaOH, before autoclaving. After autoclaving the mixture was made up to 1 L using a 30% glucose solution.

2.2.1.3 Clostridial Growth Medium (CGM)

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 $MgSO_4 \times 7\ H_2O$, 750 µl 20 g/l $FeSO_4 \times 7\ H_2O$, 500 µl 20 g/l $CaCl_2$, 500 µl 20 g/l $MgSO_4 \times H2O$, 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 L.

2.2.1.4 Clostridial Basal Medium (CBM)

For phenotypic analysis, Clostridial Basal Medium (CBM) was used which was made up of 20 ml of 20 g/l $MgSO_4 \times 7\ H_2O$, 379 µl of 20 g/l $MnSO_4$, 500 µl of 20 g/l $FeSO_4$, 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 of casamino acids. This mixture was made up to a 760 ml volume with deionised water before autoclaving. For CBM agar, 13 g/l of technical agar 1 (Oxoid, UK) was added just before autoclaving. After autoclaving, 10 ml of 1 g/l KH_2PO_4, 10 ml of 1 g/l K_2HPO_4, 20 ml of 250 g/l $CaCO_3$ and 200 ml of a 30% glucose solution were added to make up the final solution to 1 L.

2.2.2 Growth conditions and storage

E. coli was grown at 37 °C static for solid media or shaking at 200 RPM in liquid media. Overnight cultures were stored as 15% glycerol stocks

at -80 °C. When required the following antibiotic selection was commonly used, 500 µg/ml erythromycin, 250 µg/ml spectinomycin.

C. beijerinckii was grown at 37 °C in an anaerobic cabinet (MG1000 anaerobic workstation, Don Whitley Scientific) maintained at an atmosphere of 80% nitrogen, 10% carbon dioxide and 10% hydrogen. Colonies from a fresh plate were used to inoculate overnight cultures which were used to make 15 % glycerol stocks stored at -80 °C or spore stocks. Spore stocks were prepared by inoculating 5 ml of liquid medium containing 6 % glucose with 5 colonies and then left to grow for 5-7 days. After 5-7 days, the culture was centrifuged at 10, 000 × g for 5 minutes, the supernatant was discarded, and the pellet resuspended into 500 µl PBS. This was centrifuged at 10, 000 × g for 1 minute, the supernatant discarded, and the pellet resuspended into 1 ml PBS. This was then stored at 4 °C or -20 for long term storage. The *C. beijerinckii* spore stocks were revived by heating 50 µl of spore stock suspension at 80 °C for 10 minutes in a heating block. This was then immediately plated onto CGM media and left to grow for 24 to 48 hours.

For *C. beijerinckii*, the following antibiotic selection and supplementation was used, 10 µg/ml erythromycin, 750 µg/ml spectinomycin. For *pyre* mutants, 20 µg/ml uracil was used to supplement growth media, this was changed to 2 µg/ml in the presence of 600 µg/ml 5-Fluoroorotic acid. Liquid medias were left to reduce overnight in the anaerobic cabinet before use, whereas solid media only required 3 hours before use.

2.3 Molecular biological methods

2.3.1 Isolation of DNA

2.3.1.1 Plasmid DNA extraction from *E. coli*

Plasmid DNA was extracted from a bacterial host using the QIAGEN plasmid miniprep kit following the manufacturer's instructions. All DNA purifications were eluted into 50-100 µl of dH_2O.

2.3.1.2 Phenol-chloroform extraction of genomic DNA from *C. beijerinckii*

For general use, the QIAGEN DNAeasy blood and tissue kit was used to extract gDNA from an overnight culture of *C. beijerinckii,* grown in 5 ml of 2xYTG. The only alteration to the manufacturer's instructions for Gram-positive bacteria was the final elution into 50 µl of dH_2O. The eluted gDNA concentration was measured and then stored at -20 °C.

For genomic DNA used for sequencing, the relevant *C. beijerinckii* strain was grown on CGM overnight and colonies used to inoculate 10 ml of 2 × YTG (0.5 % glucose). The overnight culture was then centrifuged at 6000 × g for 5 minutes and the supernatant discarded. The pellet was resuspended in 180 µl of lysis buffer (10 mg/ml lysozyme) and incubated at 37 ^{0}C for 30 minutes. Next, 4 µl of RNase A

was added and the mixture incubated at room temperature for 30 minutes. To the mixture, 25 µl proteinase K, 85 µl dH_2O and 110 µl of 10% SDS was added before heating at 65 ^{0}C for 30 minutes. After incubation, 400 µl of phenol:chloroform:isoamyl alcohol (25:24:1, saturated with 10 mM Tris, pH8, 1 mM EDTA, Sigma:P3803) was added and mixed thoroughly. The whole mixture was transferred to a phase-lock tube and centrifuged at 21,000 × g for 3 minutes. The top phase was carefully transferred to a new phase-lock tube containing 400 µl of phenol:chloroform:isoamyl alcohol, this was once again mixed thoroughly and centrifuged at 21,000 × g for 3 minutes. The top phase was transferred to a new phase-lock tube and 10 µl RNase A was added and incubated at room temperature for 30 minutes. After incubation 400 µl of phenol:chloroform:isoamyl alcohol was added, this was mixed thoroughly and centrifuged as before. The top phase was then transferred to a 1.5 ml Eppendorf tube containing 40 µl of 3 M NaAc and 800 µl of 100% ethanol and mixed gently. This was placed onto ice where the DNA fell out of suspension. The DNA pellet was then collected using a sterile glass pipette and transferred to a fresh 1.5 ml Eppendorf tube containing 1 ml of 70% ethanol. Finally, the pellet was transferred to an empty Eppendorf and allowed to dry at room temperature for 45 minutes. The resulting DNA was re-suspended into 50 µl dH_2O and stored at -20 C.

2.3.2 DNA analysis and manipulation

2.3.2.1 DNA quantification and purity determination

Purified DNA was analysed for quantity and purity using a NanoDrop Lite (Thermo Scientific). The DNA was quantified by measuring the absorption at 260 nm. The purity of DNA was estimated by the 260/280 nm absorption ratio, DNA of good quality had a ratio between 1.8-2.0. Purified gDNA which was to be sent for whole genome sequencing was measured using a Quibit fluorometric quantification device (Invitrogen), all steps were carried out using the manufacturer's instructions.

2.3.2.2. Polymerase chain reactions

For general screening of bacterial colony, mutants and plasmids, the Thermo Scientific DreamTaq DNA polymerase was used for PCR reactions. Each 20 µl reaction was setup as follows; 10 µl DreamTaq Green PCR MasterMix (2×), 1 µl 10 µM forward primer, 1 µl 10 µM reverse primer, 1 µl template DNA at a concentration between 1-100 ng/µl, 7 µl nuclease-free H_2O. The thermocycler conditions were as follows; 95 °C for 2 minutes, or 5 minutes for colony PCR. Then followed by 25 cycles of 95 °C for 10 seconds denaturation, 5 °C below the primer annealing temperature for 30 seconds, extension carried out at 72 °C for 1 minute up to 2 Kb of DNA and then extended 1 minute

per kb after this. Finally, there was a final extension at 72 °C for 10 minutes.

For DNA to be used in cloning or to be sent off for sequencing, Phusion High-Fidelity DNA Polymerase (New England BioLabs) was used for PCR reactions. Each 50 µl reaction was setup as follows; 25 µl of Phusion high-fidelity PCR master mix (2X), 2.5 µl 10 µM forward primer, 2.5 µl 10 µM reverse primer, 1 µl template DNA at a concentration below 250 ng/µl, 1.5 µl DMSO and 17.5 µl nuclease-free H_2O. The thermocycler conditions were as follows; 98 °C for 30 seconds. Then followed by 30-35 cycles of 98 °C for 10 seconds, primer annealing temperature for 30 seconds, 72 °C for 30 seconds per 1 Kb of DNA and a final extension of 72 °C for 10 minutes.

2.3.2.3 Agarose gel electrophoresis

To separate DNA by size, samples were routinely run on gels made of 1× TAE buffer (40 mM Tris, 1 mM EDTA and 0.1% (v/v) glacial acetic acid) with 0.75-2% agarose (w/v) and 0.1% (v/v) SYBR safe (ThermoScientific). The gels were loaded into a Compact gel tank (Biometra) and electrophoresis carried out using a PowerPac 300 (Bio-rad) at 100 V, 400 mA for 30-60 minutes. DNA was visualised with a Gel Dock XR+ System (Biorad) or, when excising DNA bands from the gel, with a blue light transilluminator. Individual DNA band sizes were determined using a quick-load purple 2-log DNA ladder (NEB).

2.3.2.4 DNA purification

DNA from a reaction mix was purified using the QIAGEN PCR cleanup kit following the manufacturer's instructions. If DNA was to be extracted from an agarose gel, the QIAGEN Gel extraction kit was used following the manufacturer's instructions.

2.3.2.5 DNA restriction enzyme digests

Restriction endonucleases were purchased from New England Biolabs (NEB). Purified DNA was then cleaved under the specified reaction conditions for each enzyme. Digested DNA was purified and quantified.

2.3.2.6 DNA ligation reactions

T4 DNA ligase enzyme and reaction buffer were purchased from Promega and used following the manufacturer's instructions. The reactions were carried out in 10 µl volumes with vector:insert ratios of 1:1, 1:3 or 1:6. Reaction conditions were as follows 100 ng of vector DNA, 1 µl of 10× ligation buffer, 3 U T4 DNA ligase followed by incubation at 4 °C overnight.

2.3.2.7 HiFi DNA assembly

Plasmid DNA and fragment DNA were assembled using the NEBuilder HiFi DNA Assembly Master Mix. The fragments for assembly were designed using the online NEBuilder assembly tool (https://nebuilder.neb.com/). For a 2-3 fragment assembly, reactions were carried out at a DNA molar ratio, vector:insert, of 1:6. Generally each reaction consisted of 0.03 pmols of vector and 0.2 pmols of each insert, 10 µl of HiFi DNA Assembly Master Mix and the total volume made up to 20 µl using deionised H_2O. The mixture was incubated at 50 °C for 1 hour and then stored at -20 °C until needed.

2.3.2.8 DNA sequencing

DNA fragments/plasmids used for cloning were sent for sanger sequencing through either Source BioScience Gene Service (Nottingham) or Euorfins genomics. The DNA was prepared following the service requirements.

For genomic DNA, whole genome sequencing was conducted by MicrobesNG through the Illumina sequencer standard Nextera protocol. The gDNA was prepared following the MicrobesNG DNA preparation guidelines.

2.3.3 Plasmid Construction

2.3.3.1 Construction of the pMTL-JRH3 in-frame *agrD* gene KO vectors

The plasmid used for pyrE-based gene KO was pMTL-JRH3, based on the pMTL-83251 modular plasmid. The pMTL-JRH3 plasmid was first digested with NotI-HF restriction enzyme, run on 1% agarose gel and then purified. The left homology arm (LHA) and right homology arm (RHA) for each target gene was amplified from the *C. beijerinckii* NCIMB 8052 WT genome. Both were 750 bp plus the number of base pairs into the start and stop codon required to make an in-frame deletion without disrupting any overlapping genes. The LHA was amplified with Phusion high-fidelity DNA polymerase using the primers agrDG KO LHA_F and agrDG KO LHA_R, whilst the RHA was amplified using the primers agrDG KO RHA_F and agrDG KO RHA_R, respective to each target gene. The digested pMTL-JRH3 plasmid and respective RHA/ LHA were then assembled using HiFi DNA assembly kit (New England BioLabs). Assembled vectors were either stored at -20 °C or chemically transformed into Top10 *E. coli*. Positive transformants were sent off for Sanger sequencing through Source Bioscience.

Suicide vectors of these KO constructs were made by digesting the relevant vector with AscI and PflfI restriction enzymes. The digest was

run on a 1% agarose gel and the 5.1 kb band, which was the linear KO vector without the Gram-positive replicon, was extracted. This digest product was then treated with T4 DNA polymerase following the manufacturer's instructions, to create blunt ends. The blunt end digest product was then treated with T4 DNA ligase following the manufacturer's instructions. Assembled vectors were chemically transformed into Top10 *E. coli* and sent off for sanger sequencing.

2.3.3.2 Construction of ClosTron plasmids

ClosTron plasmids were designed following the instructions provided by the ClosTron.com website (ClosTron.com - ClosTron mutagenesis). Intron insertion target sites were selected based on the highest scores given through the Perutka algorithm. Plasmids were then synthesised by ATUM (https://www.atum.bio/eCommerce/login) and transformed into E. coli TOP10 for storage.

2.3.3.3 Construction of CRISPR KO vectors

The plasmids used for CRISPR based gene KO were based on vFS77. The vector vFS77 was first digested with FseI and HpaI restriction enzymes to remove the resistance marker *catP* and the digested vector purified via gel electrophoresis. The *ermB* resistance marker was then amplified with Phusion high-fidelity DNA polymerase from the Clostridia modular plasmid pMTL83251 using the HiFi primers ermB_fwd vFS77

hifi and ermB_rev vFS77 hifi. The digested vFS77 vector and *ermB* amplicon were then assembled using HiFi DNA assembly kit. Assembled vectors with the *ermB* marker, referred to as vLRB1 were stored at -20 °C or chemically transformed into Top10 *E. coli*. Positive transformants were sent off for Sanger sequencing through Source Bioscience.

The vLRB1 vector was targeted to specific genes by replacing the sgRNA. The guide sequences were designed through the Benchling website by isolating the target gene sequence and using the CRISPR single guide design function. The guide length was set to 20 bp with a target activation-induced cytidine deaminase (AID) NG PAM 3' side. The protospacer sequences were then screened for those which have a cytosine in the editing window (-19 to -16) and when this cytosine mutated to a thymine, a STOP codon (TAA, TAG, TGA) would be produced. Confirmed protospacer sequences were then ordered as a HiFi DNA oligonucleotide by inserting the sequence into the 20 N seed region in the following sequence ATCTTAAGGAGGAGTTTTCGTCGACNNNNNNNNNNNNNNNNNNNN GTTTTAGAGCTAGAAATAGCAAGTT. These 70 bp DNA oligonucleotides, alongside their reverse complimentary oligonucleotide e.g. agrX oligo1 and agrX oligo2, were ordered through the IDT website and then annealed to each other to create a double-stranded sgRNA. The vector vLRB1 was digested with the SalI restriction enzyme and the linearised plasmid purified via gel electrophoresis. The digested vector and sgRNA HiFi oligonucleotide were assembled using HiFi DNA

assembly kit (New England BioLabs). Assembled vectors, referred to as vLRB1-agr were either stored at -20 °C or chemically transformed into dh5α *E. coli*. Positive transformants were sent off for Sanger sequencing through Source Bioscience or Euorofins genomics.

2.3.3.4 Construction of agr complementation vectors

The JRH4 vector was digested with SalI to create a linearised vector. The *agr* genes and 200 bp upstream were amplified using primer pair agr + 200bp up_fwd and agr + 200bp up_rev for the relevant *agr* gene, from the *C. beijerinckii* NCIMB 8052 WT gDNA. The *agr* gene and linearised JRH4 vector were assembled using HiFi assembly to give the JRH4-agr complementation vector. These were either stored at -20 °C or transformed into dh5α *E. coli* and confirmed through sanger sequencing.

2.3.4 Isolation of RNA

Total RNA was extracted from *C. beijerinckii* NCIMB 8052 using the FAST RNA pro-blue kit and QIAGEN RNeasy kit. Bacterial cells were harvested by transferring 2 ml of culture into 4 ml of RNA protect (bacterial reagent), this was mixed by vortex and then left to incubate at room temperature. After 5 minutes, the mixture was centrifuged at 5000 x g at 4 °C for 10 minutes, the supernatant was discarded and the pellet stores at -80 °C until required.

To extract RNA, the bacterial pellet was resuspended in 1 ml of RNA pro solution in a fume hood, on ice. The whole mixture was then transferred into an RNA pro matric tube on ice followed by homogenisation at 6400 rpm for 45 seconds. The samples were centrifuged at 13, 000 rpm, at 4 °C for 15 minutes. The supernatant, ensuring to avoid cell debris, was carefully transferred to a 1.5 ml Eppendorf and left at room temperature. After 5 minutes, the tubes were transferred to a fume hood where 300 µl of ice-cold chloroform was added to each sample and vortexed. This was incubated at room temperature for 5 minutes and then centrifuged at 13,000 rpm, at 4 °C for 15 minutes. The upper phase was then slowly transferred to an Eppendorf tube containing 500 µl of absolute ethanol using a p200 pipette and then mixed thoroughly by inversion before incubating at -80 °C for 1 hour.

After RNA precipitation, the samples were centrifuged at 13,000 rpm, at 4 °C for 15 minutes. The supernatant was discarded, and the pellet washed in 500 µl of 70% ethanol (in DEPC RNase free water). This was once again centrifuged at 13,000 rpm, at 4 °C for 15 minutes. The supernatant was discarded, and the pellet allowed to air dry at room temperature for 30 minutes. Dried pellets were resuspended into 50 µl of DEPC RNase free water and stored at -80 °C.

Genomic DNA removal was carried out by adding 5 µl of Turbo DNase buffer and 2 µl of Turbo DNase to each sample. These were incubated for 30 minutes at 37 °C with some agitation. The RNA mixture was then

cleaned using the RNeasy Mini kit following the manufacturer's instructions.

2.3.4.1 RNA quantification and integrity

RNA integrity and quantity was analysed using the Agilent bioanalyser by following the manufacturer's instructions. The samples were also checked for gDNA contamination by carrying out PCR and gel electrophoresis. If any DNA was detected, the RNA sample was treated again with DNase as described in **section 2.3.4**.

2.3.4.2 Total RNA sequencing

Total RNA samples were sent to DeepSeq at the University of Nottingham for full RNA sequencing. The RNA was quality controlled by checking RNA integrity on the Agilent TapeStation 4200 and the Agilent RNA ScreenTape Assay Kit (Agilent; 5067-5576 and 5067-5577. The RNA concentration was measured using the Qubit fluorometer and RNA BR Assay kit (ThermoFisher Scientific; Q10211). For all samples bacterial 5S/16S and 23S rRNA removal was achieved using the QIAseq FastSelect -5S/16S/23S kit (Qiagen; 335925). Standard library preparation of RNA samples from the bacterial total RNA was undertook using the NEBNext Ultra II Directional RNA Library Prep Kit for illumina (NEB; E7760L) and NEBNext Multiplex oligos for illumina (96 unique dual index primer pairs- NEB; E6440L)- 8 cycles of PCR.

The libraries were then quantified independently with Qubit dsDNA HS kit (ThermoFisher Scientific; Q32854). The library fragment-length distributions were analysed with the Agilent TapeStation 4200 and the Agilent High Sensitivity D1000 ScreenTape Assay. Adapter dimer fragments were removed using the Blue Pippin (SAGE Science) and 2% Pippin Gel Cassette (Sage Science; BDF2010). These were then pooled together in equimolar concentrations. The concentration of the pool of libraries was confirmed using the KAPA Library Quantification Kit for Illumina (Roche; KK4824). Sequencing of these libraries on the Illumina NextSeq500 platform was conducted on a NextSeq 500 Mid Output V2.5 150, producing approximately 10 million 75bp single end reads for each library.

After sequencing raw reads were trimmed of Illumina adapters and low quality (Q<20) nucleotides using TrimGalore (v 0.6.7). Reads shorter than 15 bp were discarded. The trimmed reads were aligned to *Clostridium beijerinckii* NCIMB 8052 (GenBank accession: NC_009617.1 with additional ORFs) using HISAT2 (v 2.2.1). Then StringTie (v 2.2.1) was used to assemble genes and calculate gene abundance. The analysis of differential expression was conducted using DESeq2. The Spearman correlation method was used to calculate the distance between samples as the method measures similarities in features even if the observed values may be far apart and can mitigate the possible impact of outliers.

2.3.4.3 Reverse transcription qPCR

The Qiagen Omniscript RT Kit was used to synthesise cDNA from the relevant RNA samples, following the manufacturer's instructions. Following this, 15 ul qPCR reactions were set up using SYBR Green Master Mix (ThermoFisher Scientific), the gene-specific PCR primers at a final concentration of 0.5 uM, and a 1:10 dilution of the cDNA. These were run on a LightCycler 480.

2.4 Methods of DNA Transfer

2.4.1 Preparation of chemically competent Top10 *E. coli* cells and subsequent chemical transformation

Top10 E. coli was inoculated into 5 ml of LB broth containing the relevant selective agents. The inoculation was left to grow overnight at 37 °C, shaking at 200 RPM. The next day, 2 ml of overnight culture was used to inoculate 200 ml of LB broth on a 500 ml conical flask. This was left at 37 °C, shaking at 200 RPM until an optical density of 0.3-0.4 was reached. Once this OD was reached the flask was chilled on ice for 30 minutes. The culture was then separated into 4 x 50 ml falcon tubes and centrifuged at 3000 xg, 4 °C for 10 minutes. The supernatant was discarded, and the pellets were combined into 2× falcon tubes and each resuspended in a total of 20 ml of ice cold 100 mM $CaCl_2$-$2H_20$ (10 ml each falcon tube). The resuspension was centrifuged again at 3000 xg,

4 °C for 10 minutes. Afterwards, the supernatant was discarded, and the resulting pellet resuspended into 1.5 ml of ice cold 100mM $CaCl_2$-$2H_20$. This was then left on ice for at least 3 hours. The competent cells were then stored in 100 µl, 20% glycerol aliquots at -80 °C, until needed.

Aliquots were thawed on ice and then 50 µl of competent cells were gently mixed with 5 µl of plasmid DNA. The mixture was incubated on ice for 45 minutes and then underwent heat shock by heating for exactly 45 seconds in a 42 °C water bath. After, the mixture was immediately placed on ice for at least 2 minutes before being added to 950 µl of pre-warmed LB broth in a 1.5 ml Eppendorf. This was then incubated at 37 °C shaking at 200 RPM for 1-2 hours. After recovery, 200 µl and 400 µl of each transformation was plated on LB agar containing the appropriate antibiotic selection. Plates were then incubated at 37 °C for 24-48 hours. Transformants were screened via PCR and gel electrophoresis before being sent for sequencing.

2.4.2 Preparation of electro-competent *C. beijerinckii* cells and subsequent electro-transformation

Fresh colonies of the desired *C. beijerinckii* strain were used to inoculate 10 ml of 2 × YTG broth (0.5% glucose) and used to make a 1:10 serial dilution down to 10^{-4}. The overnight culture with an OD600nm of around 0.6 was then used to inoculate 100 ml 2 × YTG broth (20 µg/ml Uracil, 0.5% v/v glucose) to a starting OD600nm of

0.02. This was grown at 37 °C under anaerobic conditions until an OD600nm of 0.5-0.6 was reached. At this point, the culture was separated into 50 ml falcon tubes and cooled on ice for 10 minutes. The culture was then centrifuged for 10 minutes at 4 °C and 2000 × g. Subsequently, the supernatant was discarded, and the bacterial pellet immediately resuspended into 10 ml of anaerobically reduced ice-cold electroporation buffer (270 mM sucrose, 7 mM Na_2HPO_4, 7 mM NaH_2PO_4, 1 mM $MgCl_2$, pH 7.4). The resuspension was then centrifuged for 10 minutes at 4 °C and 2000 × g and then the supernatant discarded. The final bacterial pellet was then resuspended into 3 ml of ice-cold electroporation buffer. Competent cells were stored as 600 µl aliquots containing 10% (v/v) DMSO and stored at -80 °C in brown 1.5 ml Eppendorf tubes.

If frozen, aliquots were thawed under anaerobic conditions on ice and then up to 20 µl of plasmid DNA was added to 600 µl of competent cells. The mixture was transferred into a pre-cooled electroporation cuvette (Bio-rad, 4 mm) and left on ice for 8 minutes. The cuvette was then immediately placed into an electroporation chamber and electroporation carried out at 2 kV, 25 µF and ∞ Ω. The cells were immediately added to 9 ml of 2 × YTG broth and incubated for 4 hours under anaerobic conditions at 37 °C. After recovery, the cells were centrifuged for 10 minutes at room temperature and 2000 × g. The final pellet was resuspended into 100 µl of 2 × YTG broth and plated onto CGM agar containing the appropriate supplements and antibiotic

selection. Plates were incubated at 37 °C under anaerobic conditions for 24-48 hours.

2.5 Genetic manipulation of *C. beijerinckii*

2.5.1 Truncation of the *pyrE* gene in *C. beijerinckii*

Truncation of the *pyrE* gene in *C. beijerinckii* NCIMB 8052 was carried out by firstly transforming cells with the plasmid pMTL-JRH1 and selecting for plasmid uptake on CGM + 10 µg/ml erythromycin agar. After 24 hours, the larger colonies were restreaked onto CBM + 600 µg/ml 5-FOA + 2 µg/ml uracil agar. After 24 hours, the largest colonies were once again restreaked onto the same media and left to grow overnight. Resulting colonies were screened using PCR primer pair *pyrE* KO confirmation_F + *pyrE* KO confirmation_R and gel electrophoresis. Confirmed pyre mutants were then inoculated into 5 ml CBM + 600 µg/ml 5-FOA + 2 µg/ml uracil broth and incubated overnight. 1 ml of overnight culture was used to make glycerol stocks, whilst the rest was used to extract gDNA to further confirm the *pyrE* truncation through PCR/ gel electrophoresis and sanger sequencing.

2.5.2 In-frame deletion of target genes using allelic exchange

The in-frame deletion method ACE was adapted from (Ehsaan *et al.*, 2016). The relevant pMTL-JRH3 *agrD* gene KO vector was transformed into *C. beijerinckii* NCIMB 8052 ΔpyrE via electroporation. The transformants were plated onto CGM + 10 µg/ml erythromycin and left to grow overnight. Confirmed transformant colonies were restreaked twice onto CGM + 10 µg/ml erythromycin to encourage single-crossover isolates. After the second round of streaking onto selective media colonies were screened for a single cross-over event using primers JRH3 insertion check_R + c-agrD cross RHA check_R for the relevant target gene. Pure single cross-over colonies were then plated onto CBM + 20 µg/ml uracil and incubated for 2-3 days to encourage a double-cross over event. After incubation, all of the growth on the plate was collected and resuspended into 500 µl of PBS. This was used to make a serial dilution down to 10^{-6} and 100 µl of each was plates onto CBM + 600 µg/ml 5-FOA + 2 µg/ ml uracil to encourage the loss of the plasmid. The colonies that grew were patch plated onto CGM + 10 µg/ml erythromycin and CGM + 20 µg/ml uracil plates. If no growth occurred on the former, the loss of the plasmid was confirmed. The colonies with confirmed plasmid loss were then screen for the relevant target gene mutation using primer *agrD* KO confirmation_F + *agrD* KO confirmation_R. Confirmed mutants were subject to gDNA extraction which was then sent off for sequencing.

2.5.3 ClosTron mutagenesis

C. beijerinckii NCIMB 8052 *ΔpyrE* was transformed with the respective ClosTron vector. The transformations were spread-plated onto CGM agar with 750 µg/ml spectinomycin, supplemented with 20 µg/ml uracil in the case of the *ΔpyrE* mutant, these were allowed to grow overnight. The resulting colonies were screened for plasmid uptake through PCR, and positive colonies were restreaked onto the same media. All growth was then resuspended into 400 µl of PBS and plated onto CGM + 10 µg/ml erythromycin (+ 20 µg/ml uracil) and left to grow for 24-48 hours. Resulting colonies were inoculated into CGM + 10 µg/ml erythromycin (+ 20 µg/ml uracil) overnight and then used to make glycerol stocks and extract gDNA. The gDNA was used to confirm intron insertion at the target site using the necessary PCR primers flanking the target gene and gel electrophoresis.

2.5.4 In-frame deletion of target genes via CRISPR

C. beijerinckii NCIMB 8052 *ΔpyrE* was transformed with the respective vLRB1-agr CRISPR KO vector. The transformations were spread-plated onto CGM + 10 µg/ml erythromycin + 20 µg/ml uracil and allowed to grow for 24 hours. The resulting colonies were screened for plasmid uptake via PCR. Positive colonies were patch plated onto CGM + 10 µg/ml erythromycin + 20 µg/ml uracil + 5 mM theophylline to

induce the CRISPR-Cas9 activity. After 24 hours, those which grew were once again patch plated onto the same media. Resulting colonies were screened for the necessary *agr* mutation by Phusion HF PCR amplification of the relevant gene region and sent for sanger sequencing. Plasmid loss was confirmed by patch plating onto CGM + 10 µg/ml erythromycin + 20 µg/ml uracil and CGM + 20 µg/ml uracil, if no growth occurred on the former, the loss of the plasmid was confirmed.

2.5.5 Repair of the *pyrE* gene in *C. beijerinckii*

Repair of the *pyrE* gene in *C. beijerinckii* NCIMB 8052 *pyrE* mutants was carried out by firstly transforming cells with the plasmid pMTL-JRH4 and selecting for plasmid uptake on CGM + 750 µg/ml spectinomycin + 20 µg/ml uracil agar. After 24 hours, the larger colonies were restreaked onto CBM agar. After 24 hours, the largest colonies were once again restreaked onto the same media and left to grow overnight. Resulting colonies were screened using PCR primer pair *pyrE* KO confirmation_F + *pyrE* KO confirmation_R and sanger sequencing. Confirmed *pyrE* repair isolates were then inoculated into 5 ml CBM broth and incubated overnight. 1 ml of overnight culture was used to make glycerol stocks, whilst the rest was used to make spore stocks.

2.5.6 genetic complementation of mutants

The *agr* mutants containing a truncated *pyrE* gene were transformed with the relevant JRH4_agr complementation vector. Transformants were selected for by plating onto CGM + 750 µg/ml spectinomycin + 20 µg/ml uracil agar. After 24 hours complementation through the *pyrE* locus was undertaken through the same method used to repair the pyre gene (see 2.5.5). PyrE repair and *agr* gene complementation were confirmed by screening with PCR primer pair *pyrE* KO confirmation_F + *pyrE* KO confirmation_R and sent for sanger sequencing. Confirmed complementation isolates were stored in both glycerol stocks and as spore stocks.

2.6 Phenotypic characterisation of C. beijerinckii and associated mutants

2.6.1 Preparation of pre-cultures for growth analysis

The relevant *C. beijerinckii* strains were grown on CBM agar overnight. Fresh colonies were then used to inoculate 10 ml of CBM broth, which was used to prepare a 1:10 serial dilution, down to 10^{-4}, and left to grow overnight. The overnight serial dilution with an optical density measurement between 0.2-0.8 (OD 600nm) was used to inoculate

technical triplicates of the main culture to a starting OD 0.01. Culture volumes started at 60 ml CBM in a 100 ml conical flask.

2.6.2 Optical density measurements

The optical density at 600 nm of cultures was taken to analyse cell growth. This was measured by collecting a 1 ml sample and using a 1.5 ml plastic cuvette (Fisherbrand) in a Thermo Scientific Biomate 3 spectrophotometer, with sterile liquid media as a blank control. When the OD 600nm reached 0.8, the sample was diluted to 1:10 and measured again. The growth rate was calculated for each culture using the following equation:

Growth Rate = (ln [OD2/OD1]) / (T2-T1)

2.6.3 Granulose detection

The relevant *C. beijerinckii* strains were grown on CBM agar overnight. Fresh colonies were then used to inoculate 10 ml of CBM broth, which was used to prepare a 1:10 serial dilution, down to 10^{-4}, and left to grow overnight. A 10 µl inoculation loop was used to take cells from the overnight serial dilution with an optical density measurement between 0.2-0.8 (OD 600nm) and plate a heavy swap onto CBM agar without $CaCo_3$. These were left to grow for 72 hours to allow for the

accumulation of granulose. After incubation, Lugol's iodine solution (Sigma-Aldrich) was dropped onto the area of growth to visualise the presence of granulose.

2.6.4 Sporulation assay

To analyse for heat resistant spore formation, 100 µl of culture was taken after 5 days of growth. This was then heat shocked in a heating block at 80 °C for 10 minutes, before being transferred to the anaerobic cabinet. PBS was used to then make a 1:10 serial dilution down to 10^{-6}, 50 µl of each dilution was spread plated onto CGM agar. Heat resistant CFU/ ml was calculated after 24 hours.

2.6.5 High performance liquid chromatography (HPLC)

HPLC was used to quantify glucose and lactate present at specific time points in cultures of *C. beijerinckii* NCMIB and derived mutants. This required 2 ml of culture to be extracted and placed into a 1.5 ml Eppendorf. This was then centrifuged >20,000 × g for 2 minutes, the supernatant was then collected into a separate 2 ml centrifuge tube and frozen at -20 °C until needed. For analysis, 250 µl of supernatant was mixed with 250 µl of HPLC diluent (5 mM sulphuric acid, 50 mM Valerate), which acted as the internal standard. The mixture was then

filtered using a 0.22 µM filter into a HPLC glass vial containing a 300 µl glass insert and then sealed with a HPLC snap cap. Samples were stored at 4 °C if not run on the same day. All standards, prepared by the technical team, were made in HPLC grade water and then prepared in the same manner as samples.

Samples were run through a Dionex Ultimate 3000 HPLC system (Thermo Fisher Scientific), fitted with a BioRad Aminex HPX-87H column (BioRad), a refractive index and diode array detector at UV 210 nm with an isocratic flow rate of 0.5 ml/min of 5 mM sulphuric acid as mobile phase. Column temperature was 35 °C and injection volumes of 20 µl were used. The concentrations of solvents analysed were calculated using Chromeleon 7.3 Chromatography Data system (Thermo Fisher Scientific).

2.6.6 Gas Chromatography (GC)

The GC was used to quantify acetate, butyrate, acetone, butanol and ethanol present at specific time points in cultures of *C. beijerinckii* NCMIB and derived mutants. This required 2 ml of culture to be extracted and placed into a 2 ml centrifuge tube. This was then centrifuged >20,000 × g for 2 minute, the supernatant was then collected into a separate 2 ml centrifuge tube and frozen at -20 °C until needed. Upon analysis, 4 µl of 10 M sulphuric acid was added to 400 µl of sample supernatant and mixed. To the mixture, 400 µl of propyl propionate containing 50 mM valerate was added, which acted as the

internal standard. The mixture was vortexed for at least 10 seconds and then centrifuged >20,000 × g for 1 minute. Next, 300 µl of the resulting upper phase was pipetted into a 300 µl glass insert inside a GC glass vial. This was sealed with a GC snap cap and stored at -20 °C if not run on the same day. The 1 M stock solution containing acetate, butyrate, acetone, butanol and ethanol was used to prepare 2 sets of standards in ELGA water to the following concentrations 0 mM, 1 mM, 3 mM, 5 mM, 10 mM, 30 mM, 50 mM, 100 mM and 150 mM. These were then prepped in the same way as described for samples.

Samples were run through a Thermo Focus gas chromatograph containing a 30 m TR-FFAP column with 0.25 mm internal diameter, 0.25 µm film thickness and a flame ionisation detector. The hydrogen carrier gas supply was 0.8 ml/min with the flame maintained by compressed air at 350 ml/min, hydrogen 35 ml/ml and nitrogen 30 ml/min. An auto sampler was used to carry out 1 µl sample injection volumes at an injector temperature of 240 °C and detector temperature of 270 °C. The oven temperature was running at 40 °C for 2 minutes, followed by a temperature increase of 80 °C/min to 150 °C and then a temperature increase of 45 °C/min to 210 °C, which was held for 1 minute. The concentrations of solvents analysed were calculated Chromeleon 7.3 Chromatography Data system (Thermo Fisher Scientific).

2.6.7 Electron Microscopy

The required *C. beijerinckii* NCIMB 8052 isolates were grown in 60 ml of CBM and collected at 24 hours of growth. Aliquots of the cells were centrifuged at 3500 x g and the supernatant removed. The bacterial pellet was resuspended in fixative (3% Glutaraldehyde in 0.1M Cacodylate buffer) and left for 1 hour. After, the cells were centrifuged as before, and the fixative removed. The pellet was washed in 0.1M Cacodylate buffer three times before being stored in Cacodylate buffer at 4^0C overnight. The next day, the cells were centrifuged and the Cacodylate buffer removed, cells were then resuspended in 1% osmium tetroxide and incubated at room temperature for 1 hour. After, the osmium tetroxide was removed via centrifugation and the cells washed four times in 0.1M Cacodylate buffer. The cells were then dehydrated in a graded ethanol series, 2 x 15 mins 50% ethanol, 2 x 15 mins 70% ethanol, 2 x 15 mins 90% ethanol, 3 x 20 mins 100% ethanol. After the dehydration series, the ethanol was removed, and the pellet resuspended in propylene oxide two times for 15 minutes. The pellet was then resuspended in 1:1 propylene oxide: resin overnight. The next day the cells were resuspended in 100% resin for 4 hours. After 4 hours cells were resuspended in fresh resin and then left in the embedding oven for 48 hours at 60 ^{0}C. Samples were further processed and viewed by University of Nottingham, School of Life Sciences Imaging (SLiM).

2.6.8 Super resolution Microscopy

For the fluorescent microscopy analysis, cells were washed by centrifugation (2000 x g, 1 min) and resuspended in 300 µl of PBS. Following washing, cells were supplemented with the membrane dye FM 4-64 FX and the DNA dye Hoechst 33342 (Invitrogen molecular probes). Cells were left to incubate for 1 hour and mounted on 1.7% agarose coated glass slides. Fluorescent signals were visualized with a Zeiss Elyra PS1 Super resolution system, with structured illumination using PCO Edge camera; operated by Zeiss Zen Black 2012 Software. The same software was used to reconstruct the images.

2.7 In silico work

2.7.1 Analysis Raw Numerical data

Graphpad Prism was used to analyse raw numerical data and to construct graphic representation.

2.7.2 Visualising DNA and Protein Sequences

Amino acid and DNA sequences were visualised using the free online software Benchling. This was also used for oligonucleotide design.

2.7.3 Analysing *Clostridium beijerinckii* strain genomes for *agr* genes

The *C. beijerinckii* strain genomes analysed for putative *agr* loci were those which had been classified into taxid 1520 on the *National Center for Biotechnology Information (NCBI,* 2020) website. They were also determined as *C. beijerinckii* through journal articles which had confirmed the strain identification via 16srRNA/genomic sequencing (Keis, Shaheen and Jones, 2001; Wu *et al.*, 2012; Poehlein *et al.*, 2017; Chen, Sun and Wu, 2018; Jiang *et al.*, 2018; Little *et al.*, 2018; Sun *et al.*, 2018; Zhang, Li and He, 2018; Fonseca *et al.*, 2019; Kobayashi *et al.*, 2020). The genome and sequencing information for each strain is presented in **Table 2.1**.

2.7.4 Finding Agr protein and nucleotide homologues in *C. beijerinckii*

The NCBI online search tool BLASTp was used to observe similar protein sequences. Additionally, tBLASTn was used to observe for non-annotated/ unidentified AgrD-precursor nucleotide homologues (Altschul *et al.*, 1997). For both programmes the search parameters were set to default except, the database search set to 'non-redundant' protein sequences (nr)', organism search '*Clostridium beijerinckii* (taxid:1520)', exclude 'uncultured/environmental sample sequences'

and max target sequences parameter set to 1000. The AA sequences for AgrBD peptides were originally obtained from the NCBI website from 11th May 2020.

The resulting sequences were determined to be significant homologues based on conserved residues, percent identity, query cover and expected value (E-value). Both AgrB and AgrD-precursor peptides in other *Clostridium* species and *Staphylococcus* have conserved residues in the AA sequences. The AgrB has a conserved motif, RXXXXGXHXXXXXXC, where R can be replaced by K or H and tend to be around 200 AA long (Qiu *et al.*, 2005). Whereas AgrD-precursor peptides tend to be 35-60 AA in length, have a positively charged N-terminus followed by a largely hydrophobic region, a conserved cysteine (C) or serine (S) and also a highly conserved proline (P) in the C-terminus (Thoendel and Horswill, 2009, 2013; Steiner *et al.*, 2012). Sequences were manually observed for all or some of these properties, prioritising the conserved residues as an indicator, to confirm the identification. Percent identity/similarity indicates the percentage of amino acids that were the same or similar between the query and the target sequence therefore, a higher percent identity/similarity showed a more significant match. Query cover identifies what proportion of the query sequence is covered by the target. The E-value measures the probability that the alignment occurred by chance, a smaller E-value means a more significant alignment, the cut-off E-value was set to the default value of 10. Putative AgrC were later identified based on proximity to the *agrBD* genes and existing annotations.

2.7.5 Generating multiple sequence alignments and analysing phylogenetic relationships

The putative AgrB, AgrD-precursor peptide and AgrC sequences collected from the NCBI protein database were submitted to multiple sequence alignments in FASTA format. The full AgrD-precursor AA sequences were aligned to one another using the multiple sequence alignment tool, Clustal Omega with default parameters (Madeira *et al.*, 2019). For the putative AgrB and AgrC peptide sequence alignments the programme NG.Phylogeny.fr (Lemoine *et al.*, 2019) was used to conduct a multiple sequence alignment through Clustal Omega using a 5x iteration to improve accuracy for the longer sequences (Sievers *et al.*, 2011). AgrB and AgrC alignments were visualised using MSAviewer (Yachdav *et al.*, 2016).

Phylogenetic trees were generated based on the Clustal Omega alignments. The tool PhyML-SMS for high-throughput analysis was used to predict statistically maximum likelihood phylogenetic trees based on mathematical algorithms. (Guindon *et al.*, 2010; Lefort, Longueville and Gascuel, 2017; Lemoine *et al.*, 2018). The trees were generated as Newick Display (Junier and Zdobnov, 2010) and then visualised using the software Interactive Tree Of Life: iTOL (Letunic and Bork, 2019).

2.7.6 AIP sequence prediction and analysis

The full AgrD sequences were submitted to the online software SignalP-5.0 using default settings (Almagro Armenteros et al., 2019). This software provides predictions on the presence of a signal peptide and the location of the cleavage point. The automated predictions were also used to manually estimate the AIP-ring size using the AgrD alignment data. The predicted AIP ring sequences, excluding the N-terminal tail, were used to predict positive and negative strain interactions. It was assumed that those strains containing an identical or very similar AIP sequence would positively interact.

2.7.7 AgrB protein topology predictions

The outer-, inner- and trans- membrane regions of the AgrB protein homologues were predicted using the online tool TOPCONs using default parameters (Bernsel *et al.*, 2009).

Chapter 3: Analysis of Agr Quorum Sensing Components in Clostridium beijerinckii Strains

3.1 Introduction

Many studies have now identified QS systems in a variety of both pathogenic and saccharolytic species of the genus *Clostridium.* Some key examples include the RRNPP systems, Agr systems or AI-2 signaling, discussed in **section 1.8**.

The Agr QS system has been characterised to a great extent in *Staphylococcus* species, however further work is to be done to characterise the Agr QS systems of *Clostridium* species. It therefore stands that the mechanism by which Agr protein components interact with one another is largely based on *S. aureus* as a model organism. In *S. aureus*, the Agr system is made up of two transcriptional units, RNAII and RNAIII. The RNAII unit encodes the *agrBDCA* genes, each carrying out a different function, the system is discussed in detail in **section 1.7.1.1**. The Agr systems in the genus *Clostridium* are thought to run in a similar manner, however the small RNAIII transcript has not been identified (Steiner *et al.*, 2012; Patakova *et al.*, 2019).

There have been multiple Agr QS systems previously identified in the genomes of the genus *Clostridium*, including those of pathogenic species including, but not limited to, *C. perfringens, C. botulinum* and *C. difficile,* where the systems have been linked to the regulation of various virulence factors and sporulation. The Agr system has also been identified in solventogenic species like *C. beijerinckii, C. acetobutylicum*, *C. saccharoperbutylacetonicum and C.*

autoethanogenum (Steiner *et al.*, 2012; Jabbari *et al.*, 2013; Kotte *et al.*, 2020; Piatek *et al.*, 2022). Studies on solventogenic *Clostridia* have primarily been conducted on Agr systems in *C. acetobutylicum* (Scott, 2012; Steiner *et al.*, 2012; Jabbari *et al.*, 2013) where it was found that the systems may be linked to the regulation of solventogenic phase phenotypes, such as granulose production, and sporulation. However, very few have conducted an in-depth analysis into the Agr systems of *C. beijerinckii*. Some studies have conducted genomic analyses which have identified *agr* loci within the genome of multiple *C. beijerinckii* strains. As previously mentioned in **section 1.9**, when conducting genomic analysis Patakova *et al.* (2019) identified 3 putative Agr QS gene clusters in *C. beijerinckii* NRRL B-598 whereas Kotte *et al.* (2020) identified 6 putative Agr systems in *C. beijerinckii* NCIMB 8052. Another study by Magal (2019) has looked into the variation of Agr components between *Clostridium* species. The findings indicated that some groups of strains in the study, particularly when comparing those from *C. botulinum* and *C. sporogenes,* appeared to have similar sequence variations in Agr components. It was suggested that this showed divergence into different component variations and these variants could potentially allow cross-regulation of *agr* operons. For instance, *S. aureus* has four Agr types. Agr signalling between AgrI and II types and AgrI and III types can be inhibited by the diverse AIP variants produced in the different Agr systems (Ji, Beavis and Novick, 1997; Novick and Geisinger, 2008; Olson et al., 2014; Canovas et al., 2016). The variation of Agr components identified in *Clostridium* species could

potentially lead to categorisation of Agr components into different groups. (Magal, 2019). However, the Agr components for individual species should be further investigated in more detail to support these suggestions. Further detailed analysis into the putative Agr components of *C. beijerinckii* could provide novel evidence supporting or disproving such claims of cross-interaction between Agr variants within the same species.

Aims of study

The overall aim of this study was to investigate the distribution and variety of putative Agr QS systems in the fully genome sequenced strains of *C. beijerinckii* and to make predictions regarding their ability to communicate with one another. The specific objectives were to:

- Identify putative *agr* loci within the genomes of different *C. beijerinckii* strains.
- Categorise the Agr systems encoded by these loci based on their sequences and their surrounding gene regions.
- Analyse the predicted AgrB and AgrD protein sequences for conserved AA residues.
- Analyse the variation between the Agr protein components and their predicted evolutionary relationships.
- Predict the different strains' ability to communicate with each other: identify commonly shared systems and unique systems based on the predicted AIP sequence.

3.2 Results

3.2.1 Generating the AgrB and AgrD protein database to observe the distribution of *agr* gene clusters in *C. beijerinckii*

To analyse the distribution of Agr QS systems in *C. beijerinckii* the 6 putative AgrB and 7 putative AgrD AA sequences (see Table 2.2) from *C. beijerinckii* strain NCIMB 8052 (GenBank accession: CP000721.1) were used to identify putative homologues of AgrB and AgrD-precursors in other *C. beijerinckii* strains.

The AgrB component was used to identify Agr systems initially as this is the most unique component to Agr systems, holding limited similarity to other proteins in online databases (Thoendel *et al.*, 2011). A final database of AgrBD variants was generated which consisted of 35 different *C. beijerinckii* strains (see **Table 2.1**). Multiple *agr* loci were found in most strains for instance, between 2 and 7 different *agr* loci were usually identified.

The putative *agr* loci were then categorised into different groups from Agr1 to Agr15. Categorisation was based on similar neighboring genes situated just upstream and downstream of the putative *agr* loci. **Figure 3.1** shows a general schematic for the genomic arrangement of

putative *agr* loci and the surrounding gene regions, for each Agr group identified in this study. For most *agr* loci found, the *agrB* gene was correctly annotated as 'accessory gene regulator B family' with the size of the encoded protein ranging between 185-238 AA. The *agrD* gene was usually annotated as 'cyclic lactone autoinducer peptide' and encoded peptide sizes ranged between 39-58 AA. The *agrCA* genes were not as clearly annotated. However, downstream of the putative *agrBD* genes, most systems had a GHKL domain protein between 415-439 AA, assumed to be AgrC. There was also usually a gene encoding a response regulator transcription factor protein between 240-258 AA, assumed to produce AgrA. This was assumed due to their proximity to the *agrBD* and relative AA size, 447 AA and 241 AA respectively, to the counterparts identified in *C. acetobutylicum* ATCC 824 on the NCBI website. In this study, the identified Agr systems consisting of all four components were always arranged in the order *agrBDCA*. Within these clusters, *agrBD* genes would maintain the same orientation in the genome as each other and this was also independently true for *agrCA* which supports the idea that *agrBD* and *agrCA* are independently transcribed. A common example of a complete Agr system found in this study were those from the Agr5 group. These were highly prevalent, appearing in 97.14% of analysed strains (34/35). The strain *C. beijerinckii* WB53 was the only strain in which an Agr5-type system was not identified. Some loci did not appear to encode a complete Agr system and instead had one of the following arrangements identified, *agrBD*, *agrDC* or *agrBDC* (see **Figure 3.1**).

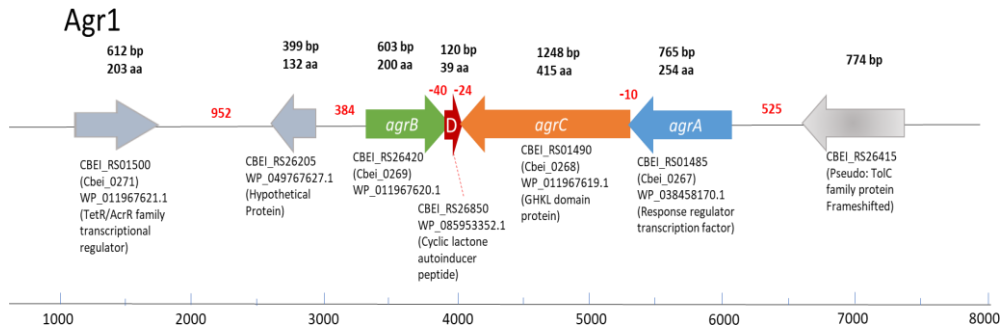
Where individual *agrCA* genes were not predicted, there were often potential surrounding genes which may carry out the functional equivalent. For example, in this study no *agrCA* genes were annotated within Agr4, 8, 10 and 11-type systems, however, putative regulatory proteins were often located nearby. Around 91.43% (32/35) of strains analysed were found to harbour Agr4-type systems. The genetic arrangement surrounding the Agr4-type systems were more variable, especially downstream genes (up to 4000 bp) from the *agr* locus. For most there was a conserved gene encoding various homologues of a TetR/AcrR family transcriptional regulator (Accession WP_012059356.1) based upstream of the putative *agrBD* genes. Some Agr4-type systems also had a Rrf2 family transcriptional regulator (Accession WP_039773646.1) situated downstream. An immediate suggestion is that one or both of these regulatory proteins could hold a linked regulatory function to the Agr4- type systems. For systems categorised into Agr4 there was usually a key characteristic genetic arrangement. Within the *agr* gene region there was usually a conserved gene, encoding homologues of a 155 AA hypothetical protein (Accession: WP_012059353.1), situated between the putative *agrD* and *agrB* genes. A Blastp search did not uncover any further information about the putative function of this hypothetical protein. The same hypothetical protein also appeared in the Agr10 group however, the gene was situated upstream of the *agrB* gene in this system.

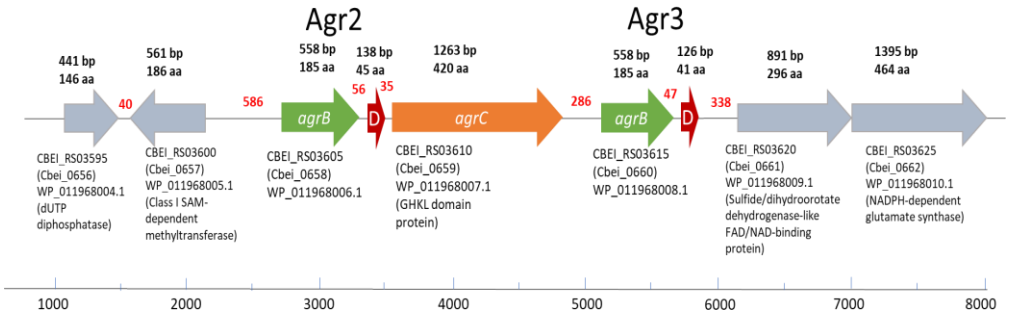
The most common grouping of Agr systems identified from *C. beijerinckii* strains in this study was Agr2, which was identified in all the

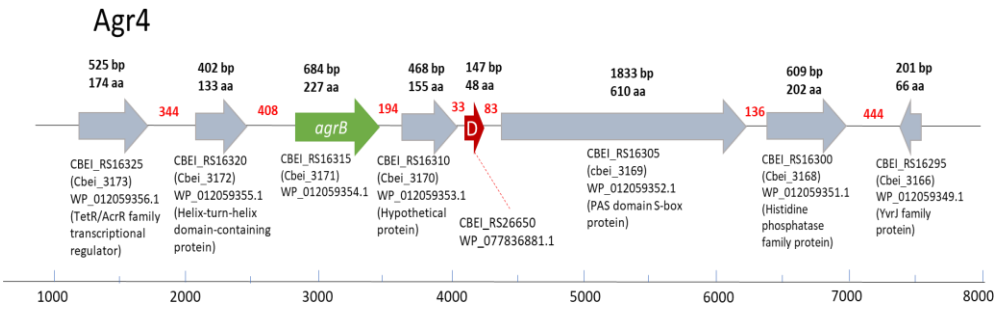
35 strains analysed. This system was arranged as *agrBDC* and had no putative regulator transcription factor genes identified nearby which could correspond to *agrA*. Furthermore, in 40% (14/35) of analysed strains, the *agrBD genes* from Agr2 systems appeared to share the putative *agrC* with a closely associated (immediately downstream located) *agrBD*, designated as Agr3. It should be noted that Agr3-type systems were never present in the absence of Agr2-type.

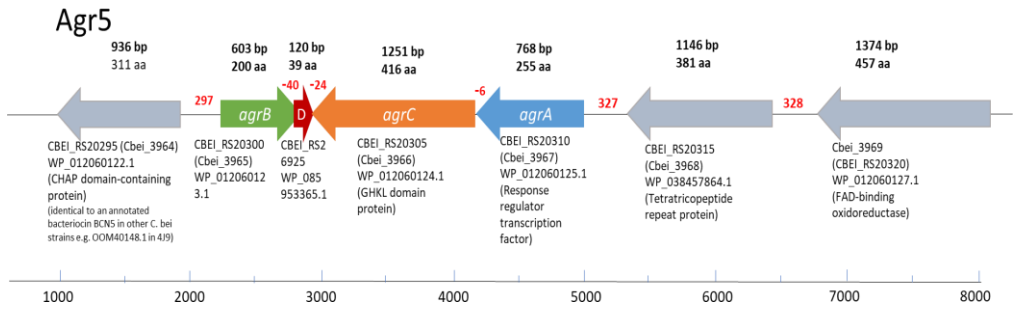
Only one group identified had an *agrD* gene which was not associated with *agrB*, these systems were in the group Agr7. The Agr7-type systems, identified in 71.43% of strains (25/35), only appeared to have putative *agrDC* and were consequently termed an orphan AgrD since this may not be processed into a mature AIP in the absence of an AgrB. However, it may be that the AgrD peptides from these systems are instead processed by AgrB proteins from alternative systems, or there is a corresponding gene function elsewhere in the genome.

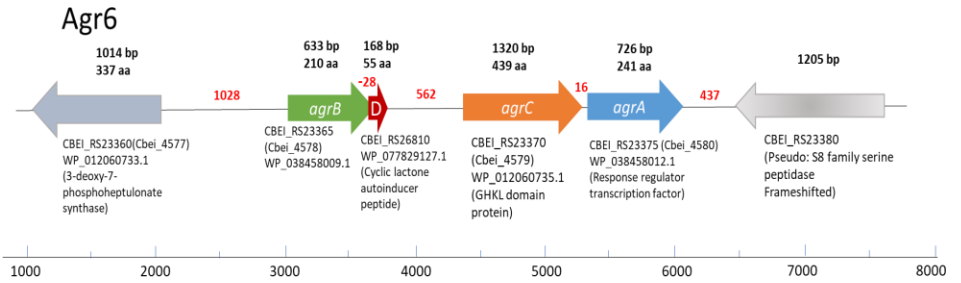
Whilst some systems occurred more commonly throughout strains, there were also more unique systems identified in other strains. For instance, the Agr10 system appears in only one strain, *C. beijerinckii* ATCC 39058 and likewise Agr13 only appears in *C. beijerinckii* HUN142. It may be that these systems allow for strain-specific communication. Other less common systems include the Agr9 grouping found in only 8.57% (3/35) of strains, Agr12 and Agr14 found in 11.43% (4/35), Agr1 found in 14.29% (5/35) Agr15 found in 17.14% (6/35), Agr11 20% (7/35), Agr6 in 22.86% (8/35) and Agr8 found in 31.43% (11/35).

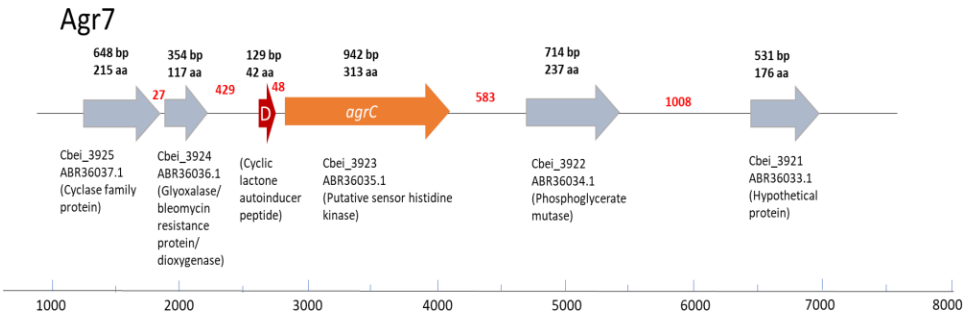

Agr1
612 bp
203 aa
399 bp
132 aa
603 bp
200 aa
120 bp
39 aa
1248 bp
415 aa
765 bp
254 aa
774 bp
952
384
-40
-24
-10
525
agrB
D
agrC
agrA
CBEI_RS01500
(Cbei_0271)
WP_011967621.1
(TetR/AcrR family
transcriptional
regulator)
CBEI_RS26205
WP_049767627.1
(Hypothetical
Protein)
CBEI_RS26420
(Cbei_0269)
WP_011967620.1
CBEI_RS26850
WP_085953352.1
(Cyclic lactone
autoinducer
peptide)
CBEI_RS01490
(Cbei_0268)
WP_011967619.1
(GHKL domain
protein)
CBEI_RS01485
(Cbei_0267)
WP_038458170.1
(Response regulator
transcription factor)
CBEI_RS26415
(Pseudo: TolC
family protein
Frameshifted)
1000
2000
3000
4000
5000
6000
7000
8000

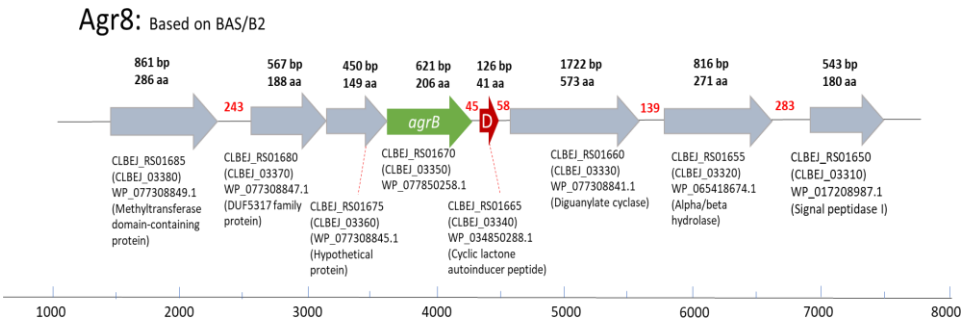

Agr2
Agr3
441 bp
146 aa
561 bp
186 aa
558 bp
185 aa
138 bp
45 aa
1263 bp
420 aa
558 bp
185 aa
126 bp
41 aa
891 bp
296 aa
1395 bp
464 aa
40
586
56
35
286
47
338
agrB
D
agrC
agrB
D
CBEI_RS03595
(Cbei_0656)
WP_011968004.1
(dUTP
diphosphatase)
CBEI_RS03600
(Cbei_0657)
WP_011968005.1
(Class I SAM-
dependent
methyltransferase)
CBEI_RS03605
(Cbei_0658)
WP_011968006.1
CBEI_RS03610
(Cbei_0659)
WP_011968007.1
(GHKL domain
protein)
CBEI_RS03615
(Cbei_0660)
WP_011968008.1
CBEI_RS03620
(Cbei_0661)
WP_011968009.1
(Sulfide/dihydroorotate
dehydrogenase-like
FAD/NAD-binding
protein)
CBEI_RS03625
(Cbei_0662)
WP_011968010.1
(NADPH-dependent
glutamate synthase)
1000
2000
3000
4000
5000
6000
7000
8000

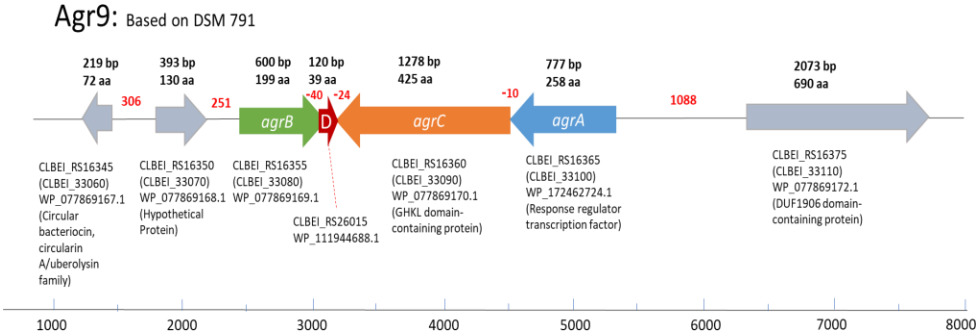

Agr4
525 bp
174 aa
402 bp
133 aa
684 bp
227 aa
468 bp
155 aa
147 bp
48 aa
1833 bp
610 aa
609 bp
202 aa
201 bp
66 aa
344
408
194
33
83
136
444
agrB
D
CBEI_RS16325
(Cbei_3173)
WP_012059356.1
(TetR/AcrR family
transcriptional
regulator)
CBEI_RS16320
(Cbei_3172)
WP_012059355.1
(Helix-turn-helix
domain-containing
protein)
CBEI_RS16315
(Cbei_3171)
WP_012059354.1
CBEI_RS16310
(Cbei_3170)
WP_012059353.1
(Hypothetical
protein)
CBEI_RS26650
WP_077836881.1
CBEI_RS16305
(cbei_3169)
WP_012059352.1
(PAS domain S-box
protein)
CBEI_RS16300
(Cbei_3168)
WP_012059351.1
(Histidine
phosphatase
family protein)
CBEI_RS16295
(Cbei_3166)
WP_012059349.1
(YvrJ family
protein)
1000
2000
3000
4000
5000
6000
7000
8000


Agr5
936 bp
311 aa
603 bp
200 aa
120 bp
39 aa
1251 bp
416 aa
768 bp
255 aa
1146 bp
381 aa
1374 bp
457 aa
297
-40
-24
-6
327
328
agrB
D
agrC
agrA
CBEI_RS20295 (Cbei_3964)
WP_012060122.1
(CHAP domain-containing
protein)
(identical to an annotated
bacteriocin BCN5 in other C. bei
strains e.g. OOM40148.1 in 4J9)
CBEI_RS20300
(Cbei_3965)
WP_01206012
3.1
CBEI_RS2
6925
WP_085
953365.1
CBEI_RS20305
(Cbei_3966)
WP_012060124.1
(GHKL domain
protein)
CBEI_RS20310
(Cbei_3967)
WP_012060125.1
(Response
regulator
transcription
factor)
CBEI_RS20315
(Cbei_3968)
WP_038457864.1
(Tetratricopeptide
repeat protein)
Cbei_3969
(CBEI_RS20320)
WP_012060127.1
(FAD-binding
oxidoreductase)
1000
2000
3000
4000
5000
6000
7000
8000

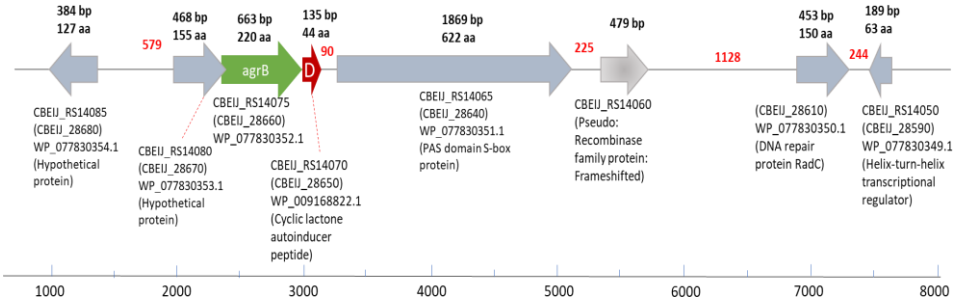

Agr6
1014 bp
337 aa
1028
633 bp
210 aa
168 bp
55 aa
-28
562
1320 bp
439 aa
16
726 bp
241 aa
437
1205 bp
agrB
agrC
agrA
CBEI_RS23360(Cbei_4577)
WP_012060733.1
(3-deoxy-7-phosphoheptulonate synthase)
CBEI_RS23365
(Cbei_4578)
WP_038458009.1
CBEI_RS26810
WP_077829127.1
(Cyclic lactone autoinducer peptide)
CBEI_RS23370
(Cbei_4579)
WP_012060735.1
(GHKL domain protein)
CBEI_RS23375 (Cbei_4580)
WP_038458012.1
(Response regulator transcription factor)
CBEI_RS23380
(Pseudo: S8 family serine peptidase Frameshifted)
1000
2000
3000
4000
5000
6000
7000
8000


Agr7
648 bp
215 aa
27
354 bp
117 aa
429
129 bp
42 aa
48
942 bp
313 aa
583
714 bp
237 aa
1008
531 bp
176 aa
agrC
Cbei_3925
ABR36037.1
(Cyclase family protein)
Cbei_3924
ABR36036.1
(Glyoxalase/ bleomycin resistance protein/ dioxygenase)
(Cyclic lactone autoinducer peptide)
Cbei_3923
ABR36035.1
(Putative sensor histidine kinase)
Cbei_3922
ABR36034.1
(Phosphoglycerate mutase)
Cbei_3921
ABR36033.1
(Hypothetical protein)
1000
2000
3000
4000
5000
6000
7000
8000

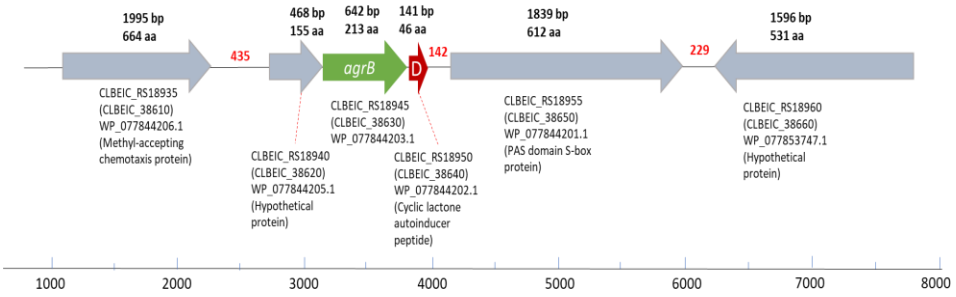

Agr8: Based on BAS/B2
861 bp
286 aa
243
567 bp
188 aa
450 bp
149 aa
621 bp
206 aa
45
126 bp
41 aa
58
1722 bp
573 aa
139
816 bp
271 aa
283
543 bp
180 aa
agrB
CLBEJ_RS01685
(CLBEJ_03380)
WP_077308849.1
(Methyltransferase domain-containing protein)
CLBEJ_RS01680
(CLBEJ_03370)
WP_077308847.1
(DUF5317 family protein)
CLBEJ_RS01675
(CLBEJ_03360)
(WP_077308845.1
(Hypothetical protein)
CLBEJ_RS01670
(CLBEJ_03350)
WP_077850258.1
CLBEJ_RS01665
(CLBEJ_03340)
WP_034850288.1
(Cyclic lactone autoinducer peptide)
CLBEJ_RS01660
(CLBEJ_03330)
WP_077308841.1
(Diguanylate cyclase)
CLBEJ_RS01655
(CLBEJ_03320)
WP_065418674.1
(Alpha/beta hydrolase)
CLBEJ_RS01650
(CLBEJ_03310)
WP_017208987.1
(Signal peptidase I)
1000
2000
3000
4000
5000
6000
7000
8000


Agr9: Based on DSM 791
219 bp
72 aa
306
393 bp
130 aa
251
600 bp
199 aa
120 bp
39 aa
-40
-24
1278 bp
425 aa
-10
777 bp
258 aa
1088
2073 bp
690 aa
agrB
agrC
agrA
CLBEI_RS16345
(CLBEI_33060)
WP_077869167.1
(Circular bacteriocin, circularin A/uberolysin family)
CLBEI_RS16350
(CLBEI_33070)
WP_077869168.1
(Hypothetical Protein)
CLBEI_RS16355
(CLBEI_33080)
WP_077869169.1
CLBEI_RS26015
WP_111944688.1
CLBEI_RS16360
(CLBEI_33090)
WP_077869170.1
(GHKL domain-containing protein)
CLBEI_RS16365
(CLBEI_33100)
WP_172462724.1
(Response regulator transcription factor)
CLBEI_RS16375
(CLBEI_33110)
WP_077869172.1
(DUF1906 domain-containing protein)
1000
2000
3000
4000
5000
6000
7000
8000

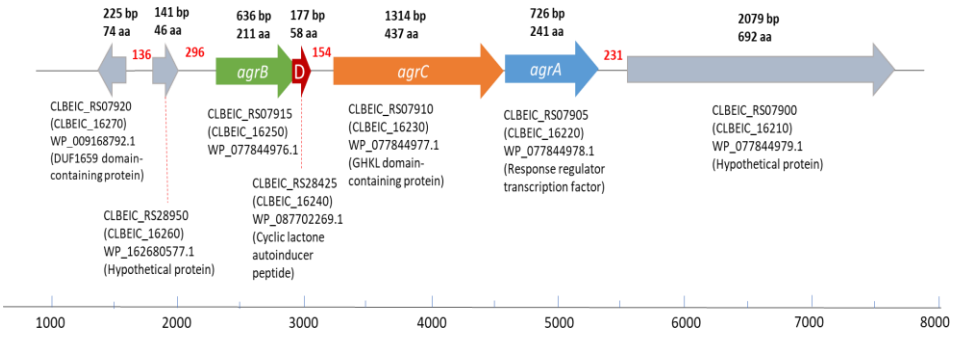
Agr10: Based on ATCC 39058



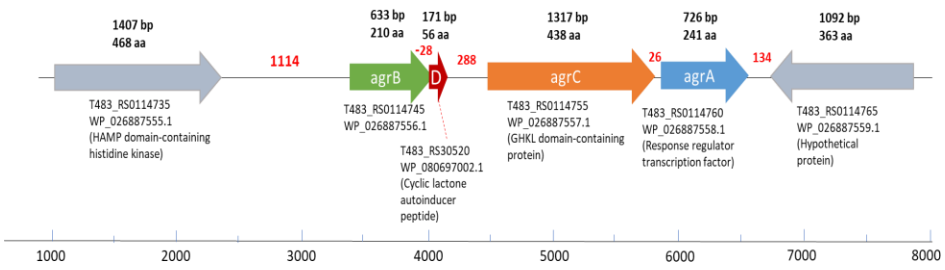
Agr11: Based on NRRL B-528



Agr12: Based on NRRL B-528



Agr13 based on HUN142



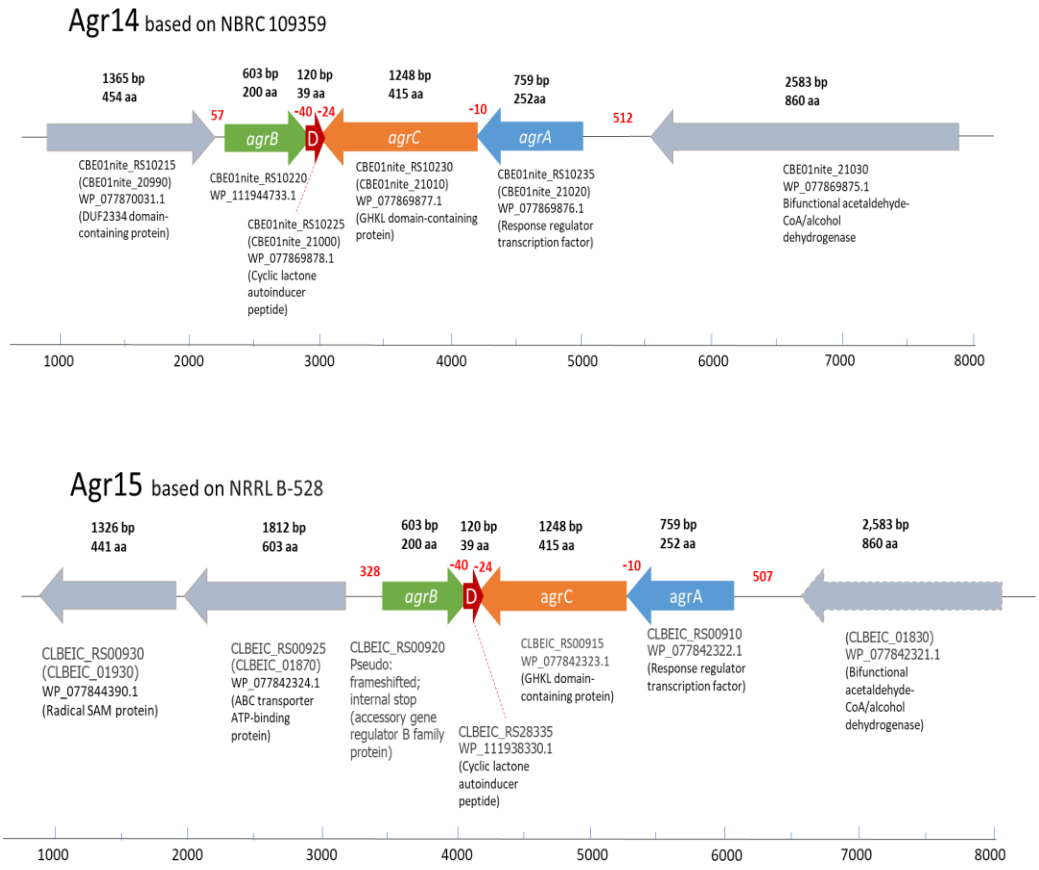


Figure 3.1 schematic representation of *agr* loci identified in genomes of *C. beijerinckii* strains.

The *agr* loci identified from individual strains were categorised into the groups Agr1-15 based on similar neighboring genes. Where possible, the schematic diagrams were based on the type strain or industrial strains**.** Agr1-7 are based on *C. beijerinckii* NCIMB 8052 (GenBank: CP000721.1). Agr8 is based on *C. beijerinckii* BAS/B2 (GenBank: NZ_LZZD01000003.1). Agr9 is based on *C. beijerinckii* DSM 791 (GenBank: NZ_LZZC01000061.1). Agr10 is based on C. beijerinckii ATCC 39058 (GenBank: LZZF01000042.1). Agr11, Agr12 and Agr15 are based on *C. beijerinckii* NRRL B-528 (GenBank: LZZK01000046.1, LZZK01000012 and LZZK01000001 respectively). Agr13 is based on *C. beijerinckii* HUN142 (GenBank: NZ_JHXK01000032). Agr14 is based on NBRC 109359 (GenBank: BKAK01000032.1) Black numbers indicate the gene nucleotide or amino acid size. Red numbers indicate the nucleotide base pairs between genes.

3.2.2 Analysis of the AgrD Homologues and Respective AIP

The AgrBD homologue pairs identified from the same loci were respectively labelled, for instance B1:D1. The full AgrD sequence of all the homologues identified were aligned to one another using Clustal Omega to identify common residues (Madeira *et al.*, 2019). **Figure 3.2** shows key residues were identified within the AgrD-precursor peptides.

In all of the AgrD-precursor homologues, a conserved cysteine residue, sometimes replaced with a serine, could be identified which was situated 10 AAs upstream from the highly conserved proline of the C-terminus, identified in blue by **Figure 3.2,** here referred to as P^{+6} according to its position relative to the C-terminal end of the predicted AIP sequence. An exception to this rule occurred with the homologues D1 and D2/29, where the conserved cysteine is 11 AAs upstream of the proline, and D6 which did not appear to have a proline (P^{+6}) in the corresponding position. This conserved positioning of the residues along with the use of the online software SignalP-5.0 (Almagro Armenteros *et al.*, 2019), were used to predict the positioning of the AIP cleavage site as the conserved cysteine residue in most of the putative AgrD homologues. It was therefore assumed the conserved cysteine formed the N-terminal end of the central AIP.

Assuming a ring-structure is formed, in most cases the AIP-ring was predicted to be 5 AAs in length based on the principle that the C-terminal AA of the AIP is usually located the same distance, 6 AAs

upstream of the conserved proline (P^{+6}), as found in putative and confirmed AIPs of other *Clostridium* species (Steiner *et al.*, 2012). Using this rule, it was found that the 5-mer cysteine-based thiolactone AIP-rings were the most common predicted structure. This AIP-type was predicted to occur in 7 of the 12 Agr groupings; Agr1, 4, 5, 8, 9, 10 and 11. Some homologues were instead predicted to have 6-mer AIP-ring variants, which were manually predicted based on the positioning and alignment of the conserved cysteine and proline residues in the peptides. There were only two Agr groups which could potentially be categorised into 6-mer cysteine-based AIP-rings. These were Agr12 with AIP-ring sequence CSMLGA and, Agr6 with AIP-ring sequences CFSGGL or CFSGGI. These putative 6-mer AIP were only present in 34.3% (12/35) of *C. beijerinckii* strains analysed.

For the homologues where the conserved cysteine was replaced with a serine at the corresponding position, it is likely the serine residue instead acts as the AIP cleavage site to form a serine-based 5-mer central AIP structure (see **Figure 3.2**) however, this could not be confirmed by SignalP-5.0 analysis. Three of the Agr groups in this study, Agr2, 3 and 7, were predicted to have 5-mer serine-based AIP making up components of the signal molecules in their systems. Interestingly, all strains analysed had both variants of a cysteine or serine based central AIP identified, showing potential intraspecific diversity of the signalling peptide ring structure.

This study identified a diverse pool of putative AIP sequence variations across the different *C. beijerinckii* strains. In total, 41 different AIP

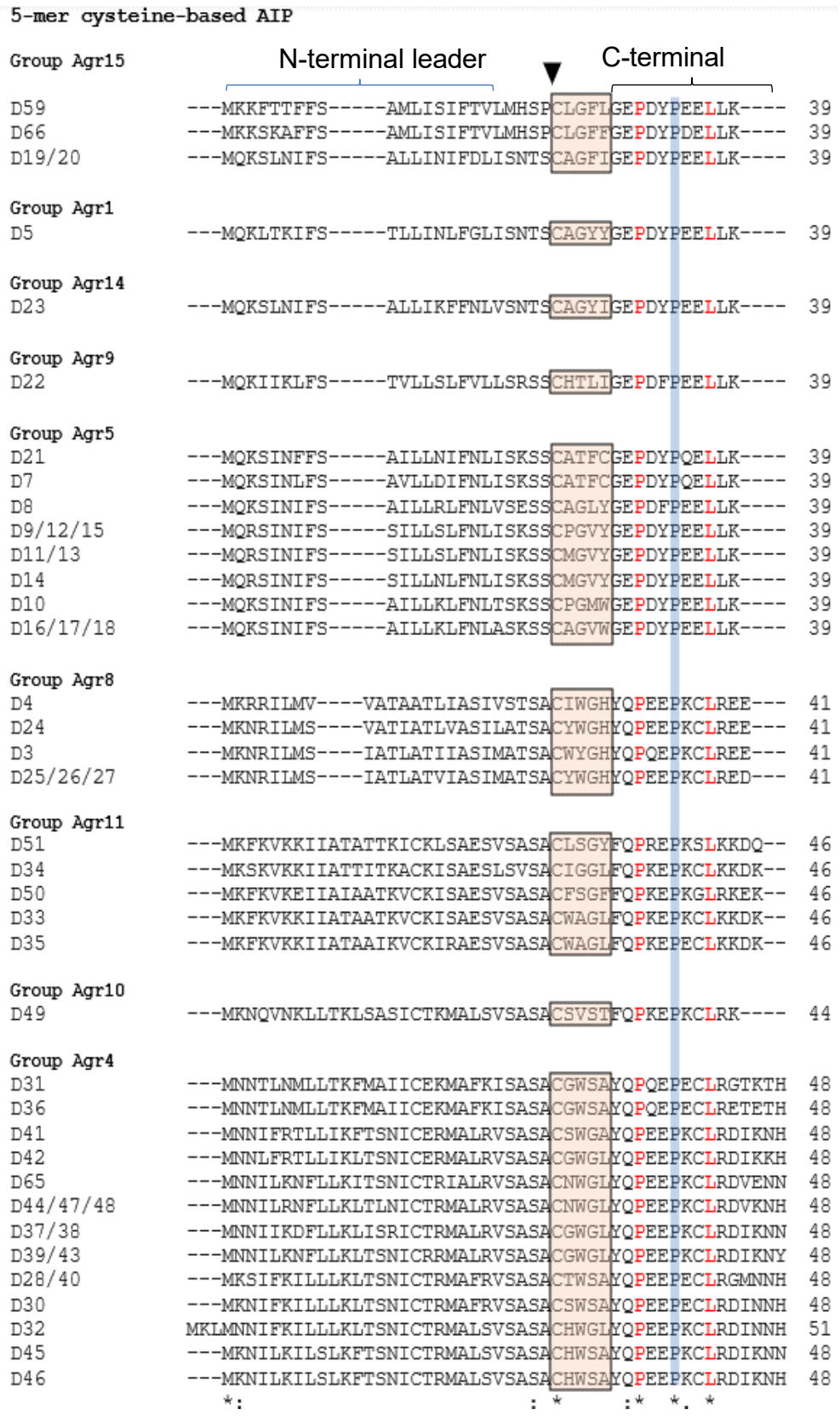
sequences were identified across all systems showing *C. beijerinckii* has potentially high intraspecific diversity in the putative AIPs. **Table 3.1** summarises the predicted AIPs, including an arbitrary 4 AA N-terminal tail, for each Agr grouping within this study. With a few exceptions, most of the AIPs organised into the same Agr group tended to have very similar AA sequences to one another and only had limited variation. For example, the Agr2 group, present in all *C. beijerinckii* strains analysed, is made up of systems which all, with the exception of strain WB53 with a putative AIP AA sequence SAGSV, had an identical serine-based AIP variant with AA sequence 'SAATA'. This provides a possible platform for conserved and direct communication between all investigated *C. beijerinckii* strains via these Agr2-type systems. Meanwhile, the Agr5-type systems were also highly present however, there were multiple and comparatively diverse AIP variants identified for this group. The AIP variants within this group were highly similar (**Table 3.1** and **Table 3.2**) which suggests they could cross-activate corresponding AgrC group 5 receptors in other strains with reduced sensitivity. On the other hand, comparison of AIPs categorised into different Agr groups tend to demonstrate higher diversity between AA sequences. The highly divergent AIPs from different Agr groups may instead act to cross-inhibit non-cognate AgrC receptors, however experimental evidence would be needed to support this idea. These observations were more easily applied with respect to the arbitrary 4 AA N-terminal tails (**Table 3.1**), of which the AA sequence was usually very similar for AIPs of the same group and dissimilar between different groups.

It was also evident that some of the different Agr groups did appear to share similarity of AIP AA sequences. For instance, similar AIP AA sequences were predicted from systems in the Agr4, Agr10 and Agr11 groups. Likewise, the putative AIPs including arbitrary tails 'SNTSCAGYY', 'SNTSCAGYI' and 'SNTSCAGFI' from Agr1, Agr14 and Agr15 respectively were highly similar.

For the AgrD precursor peptides categorised into the same Agr groups, the C-terminal portion tended to have highly similar AA sequences and therefore the same overall charge. For example, all AgrD peptides categorised into Agr5 group likely have an overall negative charge due to common glutamate (E) and aspartate (D) AA residues present. Alternatively, AgrD peptides from Agr11 likely have an overall positive charge due to the higher prevalence of arginine (R) and lysine (K) AA residues. This demonstrates that the overall charge of the AgrD C-terminal region of the AgrD peptide is only conserved within the same Agr group but not across the different identified groups in *C. beijerinckii*.

Alongside the highly conserved proline residue, indicated in **Figure 3.2**, the C-terminal regions of identified AgrD peptides also have other conserved AA residues. For AgrD peptides predicted to produce a cysteine-based AIP, there were also conserved proline and leucine residues, both shown in red by **Figure 3.2,** here referred to as P^{+3} and L^{+9} according to their positions relative to the C-terminal end of the predicted AIP sequence. The AgrD homologues from groups Agr2, Agr3 and Agr7, predicted to generate 5-mer serine-based AIP's

included the highly conserved proline (P^{+6}) however, did not show the key proline (P^{+3}) and leucine (L^{+9}) residues, instead they appear to be replaced by glutamate and methionine residues respectively. The replacement of a usually conserved proline, a neutral non-polar AA, with glutamate, a negatively charged polar AA, likely has an effect on the overall properties of the C-terminal portion. Likewise, not all AgrD peptides predicted to encode 6-mer cysteine-based AIPs had these conserved residues, in particular the AgrD 6 homologue from group 12 appeared to lack the highly conserved proline (P^{+6}) altogether.


5-mer cysteine-based AIP
N-terminal leader
C-terminal
Group Agr15
D59 ---MKKFTTFFS-----AMLISIFTVLMHSPCLGFLGEPDYPEELLK---- 39
D66 ---MKKSKAFFS-----AMLISIFTVLMHSPCLGFFGEPDYPDELLK---- 39
D19/20 ---MQKSLNIFS-----ALLINIFDLISNTSCAGFIGEPDYPEELLK---- 39
Group Agr1
D5 ---MQKLTKIFS-----TLLINLFGLISNTSCAGYYGEPDYPEELLK---- 39
Group Agr14
D23 ---MQKSLNIFS-----ALLIKFFNLVSNTSCAGYIGEPDYPEELLK---- 39
Group Agr9
D22 ---MQKIIKLFS-----TVLLSLFVLLSRSSCHTLIGEPDFPEELLK---- 39
Group Agr5
D21 ---MQKSINFFS-----AILLNIFNLISKSSCATFCGEPDYPQELLK---- 39
D7 ---MQKSINLFS-----AVLLDIFNLISKSSCATFCGEPDYPQELLK---- 39
D8 ---MQKSINIFS-----AILLRLFNLVSESSCAGLYGEPDFPEELLK---- 39
D9/12/15 ---MQRSINIFS-----SILLSLFNLISKSSCPGVYGEPDYPEELLK---- 39
D11/13 ---MQRSINIFS-----SILLSLFNLISKSSCMGVYGEPDYPEELLK---- 39
D14 ---MQRSINIFS-----SILLNLFNLISKSSCMGVYGEPDYPEELLK---- 39
D10 ---MQKSINIFS-----AILLKLFNLTSKSSCPGMWGEPDYPEELLK---- 39
D16/17/18 ---MQKSINIFS-----AILLKLFNLASKSSCAGVWGEPDYPEELLK---- 39
Group Agr8
D4 ---MKRRILMV----VATAATLIASIVSTSACIWGHYQPEEPKCLREE--- 41
D24 ---MKNRILMS----VATIATLVASILATSACYWGHYQPEEPKCLREE--- 41
D3 ---MKNRILMS----IATLATIIASIMATSACWYGHYQPQEPKCLREE--- 41
D25/26/27 ---MKNRILMS----IATLATVIASIMATSACYWGHYQPEEPKCLRED--- 41
Group Agr11
D51 ---MKFKVKKIIATATTKICKLSAESVSASACLSGYFQPREPKSLKKDQ-- 46
D34 ---MKSKVKKIIATTITKACKISAESLSVSACIGGLFQPKEPKCLKKDK-- 46
D50 ---MKFKVKEIIAIAATKVCKISAESVSASACFSGFFQPKEPKGLRKEK-- 46
D33 ---MKFKVKKIIATAATKVCKISAESVSASACWAGLFQPKEPKCLKKDK-- 46
D35 ---MKFKVKKIIATAAIKVCKIRAESVSASACWAGLFQPKEPECLKKDK-- 46
Group Agr10
D49 ---MKNQVNKLLTKLSASICTKMALSVSASACSVSTFQPKEPKCLRK---- 44
Group Agr4
D31 ---MNNTLNMLLTKFMAIICEKMAFKISASACGWSAYQPQEPECLRGTKTH 48
D36 ---MNNTLNMLLTKFMAIICEKMAFKISASACGWSAYQPQEPECLRETETH 48
D41 ---MNNIFRTLLIKFTSNICERMALRVSASACSWGAYQPEEPKCLRDIKNH 48
D42 ---MNNLFRTLLIKLTSNICERMALRVSASACGWGLYQPEEPKCLRDIKKH 48
D65 ---MNNILKNFLLKITSNICTRIALRVSASACNWGLYQPEEPKCLRDVENN 48
D44/47/48 ---MNNILRNFLLKLTLNICTRMALRVSASACNWGLYQPEEPKCLRDVKNH 48
D37/38 ---MNNIIKDFLLKLISRICTRMALRVSASACGWGLYQPEEPKCLRDIKNN 48
D39/43 ---MNNILKNFLLKLTSNICRRMALRVSASACGWGLYQPEEPKCLRDIKNY 48
D28/40 ---MKSIFKILLLKLTSNICTRMAFRVSASACTWSAYQPEEPECLRGMNNH 48
D30 ---MKNIFKILLLKLTSNICTRMAFRVSASACSWSAYQPEEPECLRDINNH 48
D32 MKLMNNIFKILLLKLTSNICTRMALSVSASACHWGLYQPEEPKCLRDINNH 51
D45 ---MKNILKILSLKFTSNICTRMALSVSASACHWSAYQPEEPKCLRDIKNN 48
D46 ---MKNILKILSLKFTSNICTRMALSVSASACHWSAYQPEEPKCLRDIKNH 48

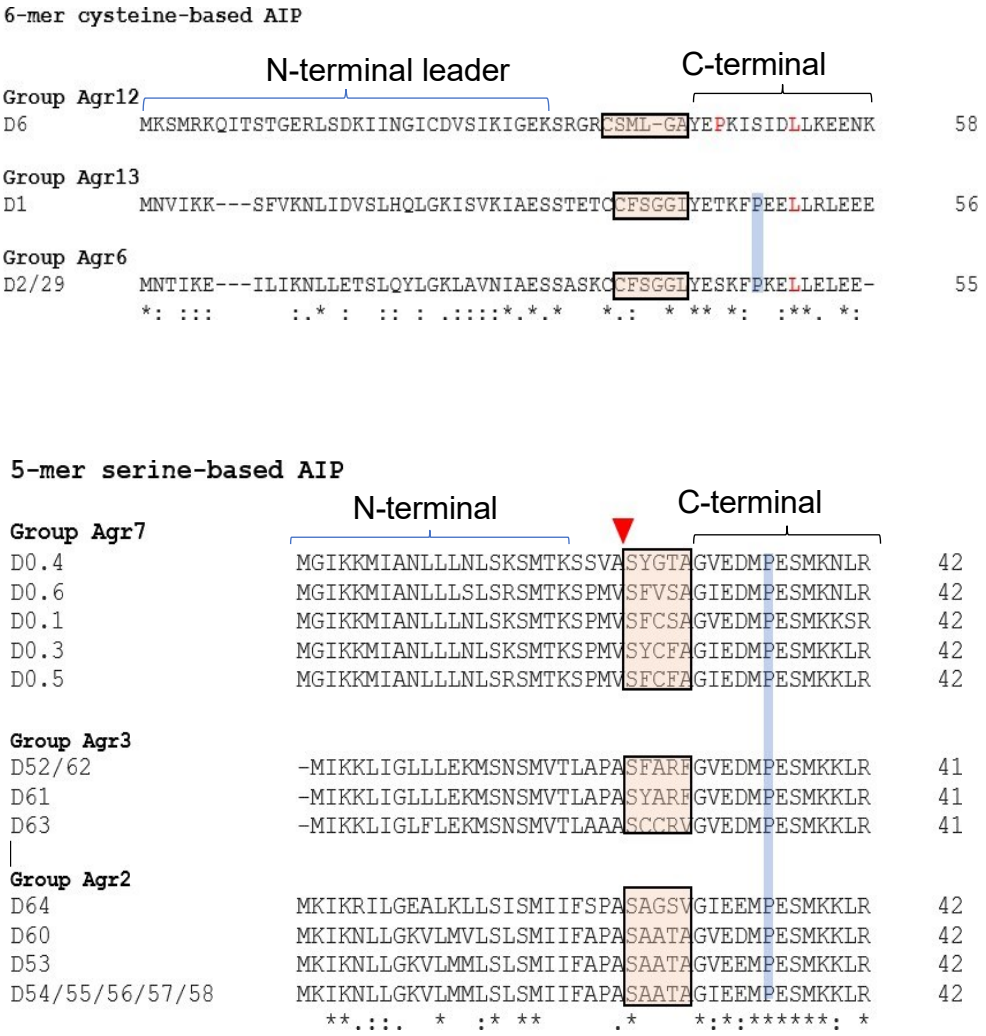


Figure 3.2 Omega multiple sequence alignment of putative AgrD-precursor sequences from *C. beijerinckii* strains analysed in this study.

AgrD sequences were grouped according to surrounding genomic arrangements and then subsequently categorised into a) 5-mer cysteine-based, b) 6-mer cysteine-based and c) 5-mer serine based predicted AIP-rings. On the left, Identities are given to individual homologues based on which AgrB homologue they pair with (e.g. D1 with B1, D2/29 with B2 and B29). The size of the peptide in amino acids is indicated at the far right in black. An asterisk (*) indicates a fully conserved residue, a colon (:) indicates conservation of AAs with similar properties and a period (.) indicates conservation of AAs with weakly similar properties. A highly conserved proline is shaded in blue whilst other common residues are shown in red. The proposed autoinducing peptide (AIP) cleavage site, predicted using SignalP-5.0 (Almagro Armenteros et al., 2019), is indicated by a black arrow. Manually predicted AIP cleavage sites are indicated by a red arrow. Predicted AIP size, excluding tails, has been highlighted in orange immediately after the proposed cleavage point.

Table 3.1 AIP sequence predictions

The AIP size and sequence was predicted using the online software SignalP-5.0 (Almagro Armenteros et al., 2019) and manually observing the sequence alignment data using the criteria employed by (Steiner et al., 2012). An arbitrary 4-AA N-terminal tail was included at the N-terminal of each AIP-ring to allow for comparison of putative tail sequences. In green are the cysteine/serine residues predicted to be required for AIP ring formation.

Agr group	AIP amino acid sequences
5-mer cysteine-based AIP	
1	SNTSCAGYY
4	SASACGWGL
	SASACGWSA
	SASACHWGL
	SASACHWSA
	SASACNWGL
	SASACSWGA
	SASACSWSA
	SASACTWSA
5	SKSSCAGVW
	SESSCAGLY
	SKSSCATFC
	SKSSCMGVY
	SKSSCPGVY
	SKSSCPGMW
8	ATSACYWGH

	ATSACWYGH
	STSACIWGH
9	SRSSCHTLI
10	SASACSVST
11	SASACWAGL
	SASACLSGY
	SVSACIGGL
	SASACFSGF
14	SNTSCAGYI
15	SNTSCAGFI
	MHSPCLGFF
	MHSPCLGFL
6-mer cysteine-based AIP	
6	ASKCCFSGGL
12	SRGRCSMLGA
13	TETCCFSGGI
5-mer serine-based AIP	
2	FAPASAATA
	FSPASAGSV
3	LAAASCCRV
	LAPASFARF
	LAPASYARF

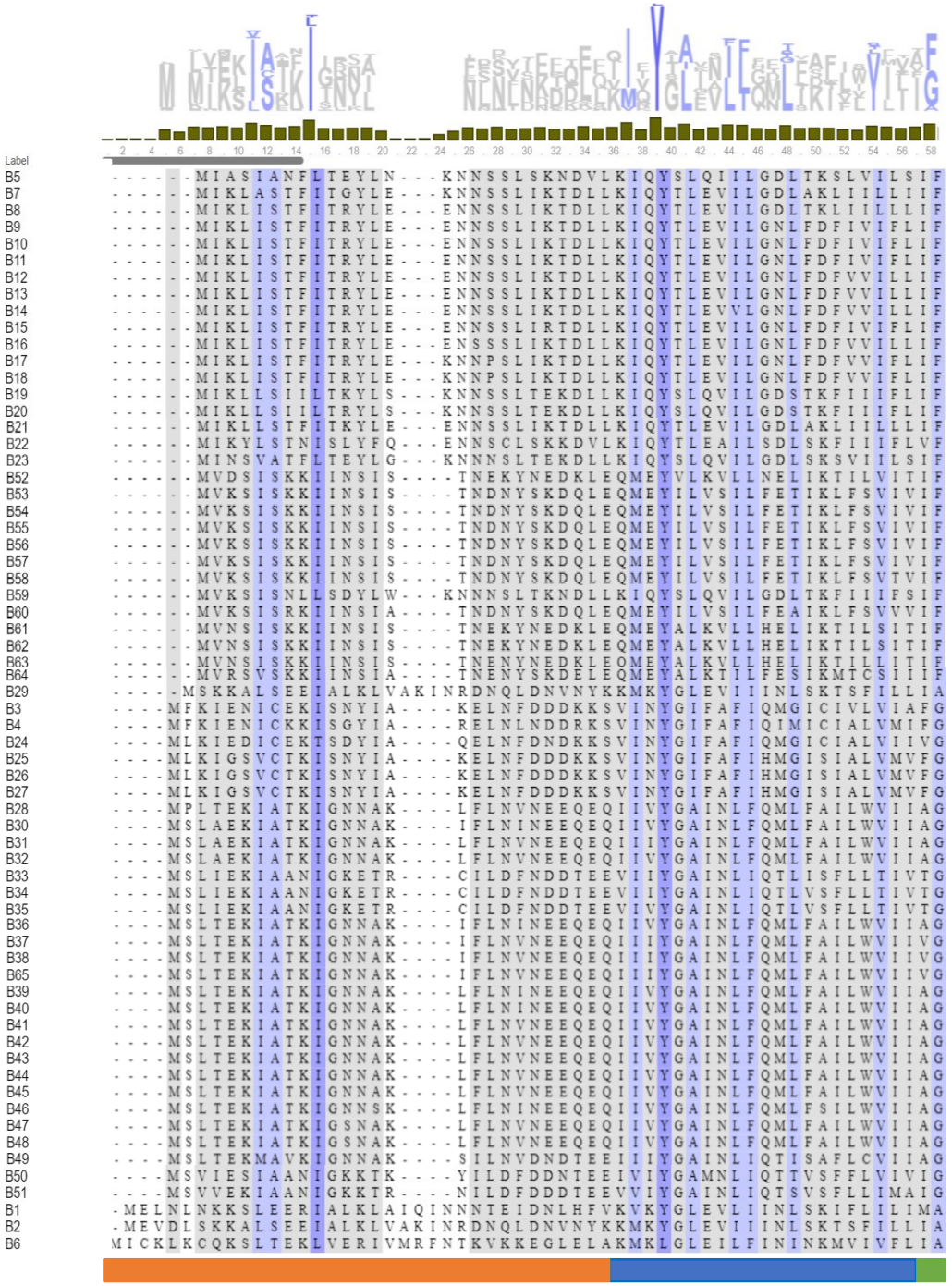
7	SPMVSFCFA
	SPMVSFVSA
	SPMVSYCFA
	SPMVSFCSA
	SSVASYGTA

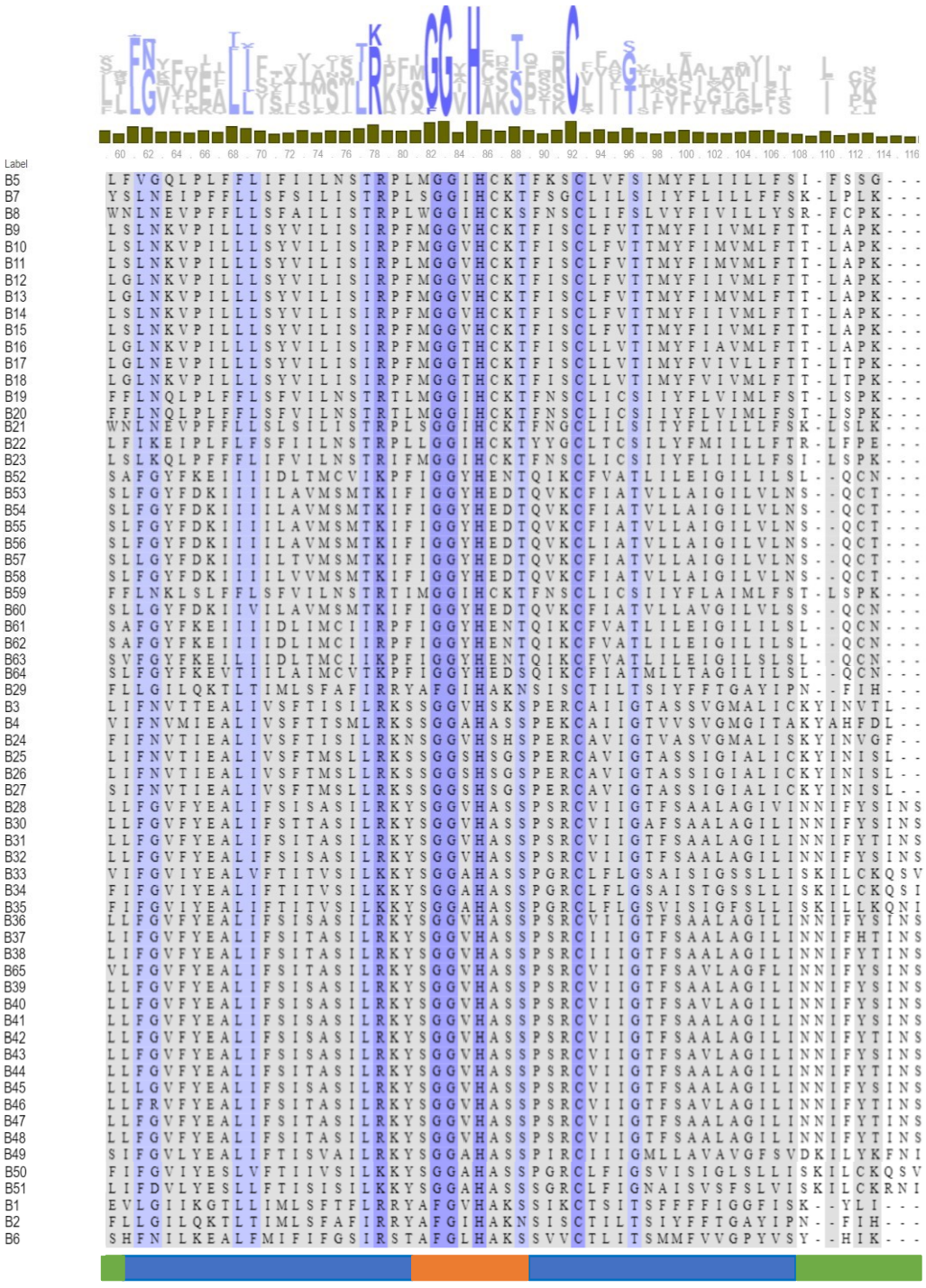
Table 3.2 *C. beijerinckii* predicted AIP sequences by group and by strain.

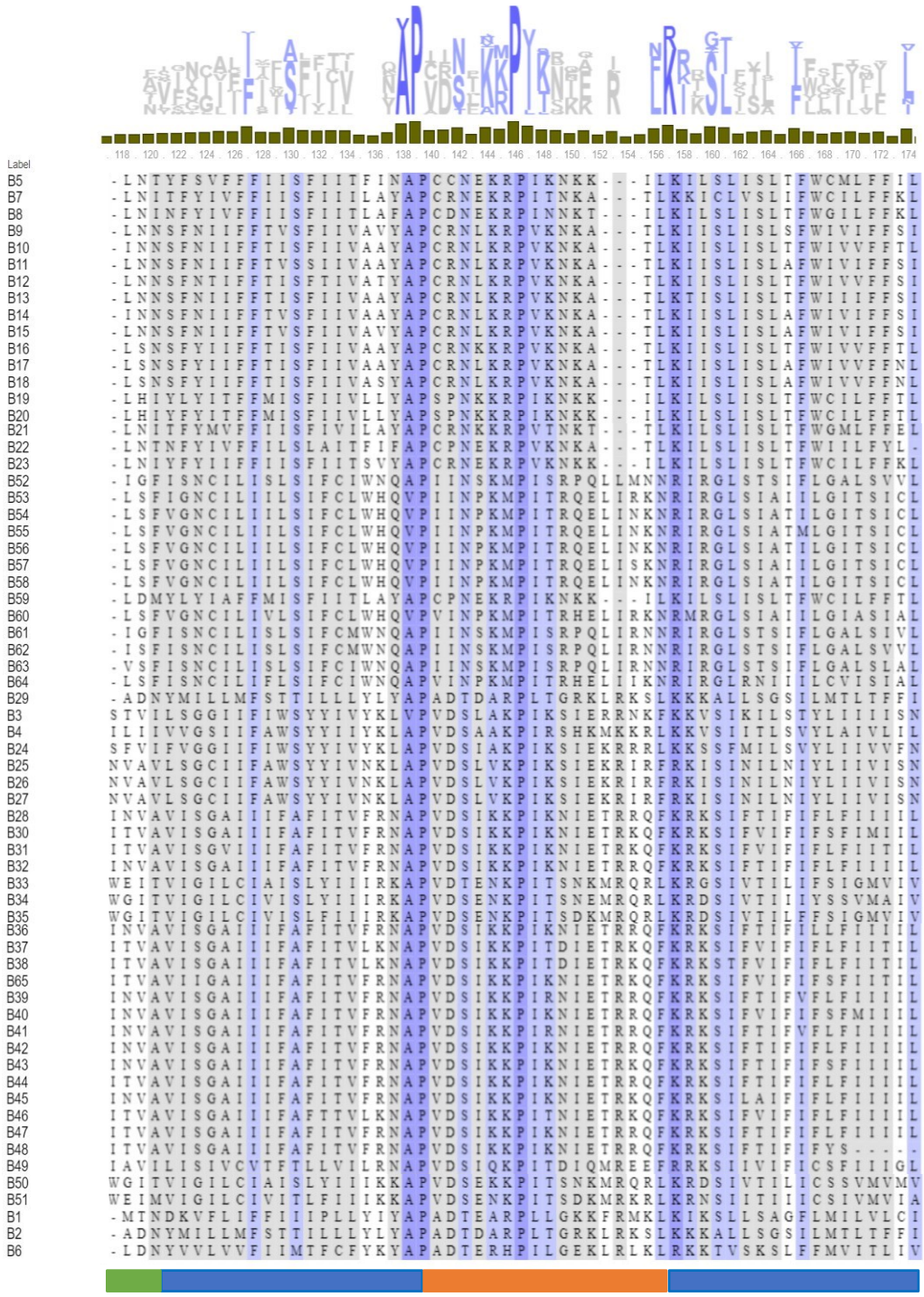
Clostridium beijerinckii strain	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4J9	CAGYY	SAATA	SFARF	CGWGL	CAGVW	CFSGGL	SYCFA								
DJ032	CAGYY	SAATA	SFARF	CGWGL	CAGVW	CFSGGL	SYCFA								
NCIMB 8052	CAGYY	SAATA	SFARF	CGWGL	CAGVW	CFSGGL	SYCFA								
NRRL B-591	CAGYY	SAATA	SFARF	CGWGL	CAGVW	CFSGGL	SYCFA								
SA-1 ATCC 35702	CAGYY	SAATA	SFARF	CGWGL	CAGVW	CFSGGL	SYCFA								
ATCC 39058		SAATA		CTWSA	CAGVW	CFSGGL	SYCFA			CSVST					
BAS/B2		SAATA	SYARF	CGWSA	CAGVW		SYCFA	CYWGH							
BAS/B3/I/124		SAATA	SYARF	CGWSA	CAGVW		SYCFA	CYWGH							
BGS1		SAATA		CHWGL	CAGVW						CLSGY				
BIOML-A11		SAATA		CGWGL	CAGLY			CIWGH							
Br21		SAATA		CGWGL	CAGLY			CIWGH							
DJ033		SAATA	SFARF	CNWGL	CAGVW	CFSGGL	SFCFA								
DJ051		SAATA	SCCRV	CNWGL	CATFC		SYGTA				CIGGL	CSMLGA		CAGYI	
DJ076		SAATA		CGWGL	CAGVW		SFCFA								
DJ077		SAATA		CGWGL	CAGVW		SFCFA								
DJ078		SAATA		CNWGL	CMGVY		SFCFA								
DJ091		SAATA		CGWGL	CAGVW		SFCFA								
DSM 53		SAATA		CSWSA	CPGVY		SFCSA								CLGFL
DSM 791		SAATA		CNWGL	CAGVW		SFCFA	CYWGH	CHTLI					CAGYI	
G117		SAATA		CSWGA	CPGVY										CAGFI
HUN142		SAATA		CHWSA	CATFC			CWYGH			CWAGL		CFSGGI		
NBRC 109359		SAATA		CNWGL	CAGVW		SFCFA	CYWGH	CHTLI					CAGYI	
NCIMB 14988		SAATA		CHWSA	CPGMW		SYGTA								CAGFI
NCP 260		SAATA	SYARF	CGWSA	CAGVW		SYCFA	CYWGH							
NCTC13035		SAATA		CNWGL	CAGVW		SFCFA	CYWGH	CHTLI					CAGYI	
NRRL B-528		SAATA	SCCRV	CGWGL	CATFC		SFVSA				CWAGL	CSMLGA			CLGFF
NRRL B-593		SAATA	SCCRV	CGWGL	CATFC		SFVSA				CWAGL	CSMLGA			CLGFF
NRRL B-596		SAATA		CGWSA	CMGVY		SFCFA				CFSGF				
NRRL B-598		SAATA	SFARF	CGWSA	CPGVY		SFCFA	CYWGH			CFSGF				
WB		SAATA		CSWGA	CPGVY										CAGFI
WB01_NA02		SAATA	SCCRV	CGWGL	CMGVY										
WB53		SAGSV						CYWGH							
Clostridium diolis DSM 15410		SAATA			CAGVW		SFCFA								
Clostridium diolis NJP7		SAATA		CTWSA	CAGIY							CSMLGA			
Clostridium diolis WST		SAATA			CAGVW	CFSGGL									
	14.29%	100%	40%	91.43%	97.14%	22.86%	71.43%	31.43%	8.57%	2.86%	20%	11.43%	2.86%	11.43%	17.14%

3.2.3 AgrB Analysis

All of the AgrB homologues identified in this study were subject to multiple sequence alignment using Clustal Omega (Sievers *et al.*, 2011). To analyse the topology of the AgrB protein homologues identified, the general outer-, inner- and trans- membrane sections of the AgrB proteins were predicted using TOPCONs (Bernsel *et al.*, 2009). This generated 6 general transmembrane regions in the AgrB homologues in *C. beijerinckii,* see **Figure 3.3**. Both the N- and C-terminals are predicted to be located on the inner cell membrane, on the cytoplasmic side. The predicted inner membrane and transmembrane segments tended to have more highly conserved residues. Interestingly, charged residues appeared to accumulate close to cytoplasmic membrane regions, these may be sites of protein-peptide interaction with AgrD. Several AA residues were identified to be completely conserved in the putative AgrB sequences from *C. beijerinckii*, these included G83, H85, C92, P139 and P146. Other residues which were highly conserved were I15, Y39, R/K78, G82, A138, R/K157 and I193. The residues R/K78, G/F82, G83, H85 and C92 formed the highly conserved motif RXXXGGXHXXXXXXC, with R sometimes replaced by a K. This motif was present in all AgrB identified. Downstream of this conserved motif there were also two completely conserved proline residues, P139/146.


Label
B5
B7
B8
B9
B10
B11
B12
B13
B14
B15
B16
B17
B18
B19
B20
B21
B22
B23
B52
B53
B54
B55
B56
B57
B58
B59
B60
B61
B62
B63
B64
B29
B3
B4
B24
B25
B26
B27
B28
B30
B31
B32
B33
B34
B35
B36
B37
B38
B65
B39
B40
B41
B42
B43
B44
B45
B46
B47
B48
B49
B50
B51
B1
B2
B6


Label
60 62 64 66 68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98 100 102 104 106 108 110 112 114 116
B5 LFVGQLPLFFLIFIILNSTRPLMGGIHCKTFKSCLVFSIMYFLIILLFSI-FSSG---
B7 YSLNEIPFFLLSFSILISTRPLSGGIHCKTFSGCLILSIIYFLILLFFSK-LPLK---
B8 WNLNEVPFFLLSFAILISTRPLWGGIHCKSFNSCLIFSLVYFIVILLYSR-FCPK---
B9 LSLNKVPILLLSYVILISIRPFMGGVHCKTFISCLFVTTMYFIIVMLFTT-LAPK---
B10 LSLNKVPILLLSYVILISIRPFMGGVHCKTFISCLFVTTMYFIMVMLFTT-LAPK---
B11 LSLNKVPILLLSYVILISIRPLMGGVHCKTFISCLFVTTMYFIMVMLFTT-LAPK---
B12 LGLNKVPILLLSYVILISIRPFMGGVHCKTFISCLFVTTMYFIIVMLFTT-LAPK---
B13 LGLNKVPILLLSYVILISIRPFMGGVHCKTFISCLFVTTMYFIMVMLFTT-LAPK---
B14 LSLNKVPILLLSYVILISIRPFMGGVHCKTFISCLFVTTMYFIIVMLFTT-LAPK---
B15 LSLNKVPILLLSYVILISIRPFMGGVHCKTFISCLFVTTMYFIIVMLFTT-LAPK---
B16 LGLNKVPILLLSYVILISIRPFMGGTHCKTFISCLLVTIMYFIAVMLFTT-LAPK---
B17 LGLNEVPILLLSYVILISIRPFMGGTHCKTFISCLLVTIMYFVIVLLFTT-LTPK---
B18 LGLNKVPILLLSYVILISIRPFMGGTHCKTFISCLLVTIMYFVIVMLFTT-LTPK---
B19 FFLNQLPLFFLSFVILNSTRTLMGGIHCKTFNSCLICSIIYFLVIMLFST-LSPK---
B20 FFLNQLPLFFLSFVILNSTRTLMGGIHCKTFNSCLICSIIYFLVIMLFST-LSPK---
B21 WNLNEVPFFLLSLSILISTRPLSGGIHCKTFNGCLILSITYFLILLLFSK-LSLK---
B22 LFIKEIPLFLFSFIILNSTRPLLGGIHCKTYYGCLTCSILYFMIILLFTR-LFPE---
B23 LSLKQLPFFFLIFVILNSTRIFMGGIHCKTFNSCLICSIIYFLIILLFSI-LSPK---
B52 SAFGYFKEIIIIDLTMCVIKPFIGGYHENTQIKCFVATLILEIGILILSL--QCN---
B53 SLFGYFDKIIIILAVMSMTKIFIGGYHEDTQVKCFIATVLLAIGILVLNS--QCT---
B54 SLFGYFDKIIIILAVMSMTKIFIGGYHEDTQVKCFIATVLLAIGILVLNS--QCT---
B55 SLFGYFDKIIIILAVMSMTKIFIGGYHEDTQVKCFIATVLLAIGILVLNS--QCT---
B56 SLFGYFDKIIIILAVMSMTKIFIGGYHEDTQVKCLIATVLLAIGILVLNS--QCT---
B57 SLLGYFDKIIIILTVMSMTKIFIGGYHEDTQVKCFIATVLLAIGILVLNS--QCT---
B58 SLFGYFDKIIIILVVMSMTKIFIGGYHEDTQVKCFIATVLLAIGILVLNS--QCT---
B59 FFLNKLSLFFLSFVILNSTRTIMGGIHCKTFNSCLICSIIYFLAIMLFST-LSPK---
B60 SLLGYFDKIIVILAVMSMTKIFIGGYHEDTQVKCFIATVLLAVGILVLSS--QCN---
B61 SAFGYFKEIIIIDLIMCIIRPFIGGYHENTQIKCFVATLILEIGILILSL--QCN---
B62 SAFGYFKEIIIIDLIMCIIRPFIGGYHENTQIKCFVATLILEIGILILSL--QCN---
B63 SVFGYFKEILIIDLTMCIIKPFIGGYHENTQIKCFVATLILEIGILSLSL--QCN---
B64 SLFGYFKEVTIILAIMCVTKPFIGGYHEDSQIKCFIATMLLTAGILILSL--QCN---
B29 FLLGILQKTLTIMLSFAFIRRYAFGIHAKNSISCTILTSIYFFTGAYIPN--FIH---
B3 LIFNVTTEALIVSFTISILRKSSGGVHSKSPERCAIIGTASSVGMALICKYINVTL--
B4 VIFNVMIEALIVSFTTSMLRKSSGGAHASSPEKCAIIGTVVSVGMGITAKYAHFDL--
B24 FIFNVTIEALIVSFTISILRKNSGGVHSHSPERCAVIGTVASVGMALISKYINVGF--
B25 LIFNVTIEALIVSFTMSLLRKSSGGSHSGSPERCAVIGTASSIGIALICKYINISL--
B26 LIFNVTIEALIVSFTMSLLRKSSGGSHSGSPERCAVIGTASSIGIALICKYINISL--
B27 SIFNVTIEALIVSFTMSLLRKSSGGSHSGSPERCAVIGTASSIGIALICKYINISL--
B28 LLFGVFYEALIFSISASILRKYSGGVHASSPSRCVIIGTFSAALAGIVINNIFYSINS
B30 LLFGVFYEALIFSTTASILRKYSGGVHASSPSRCVIIGAFSAALAGILINNIFYSINS
B31 LLFGVFYEALIFSITASILRKYSGGVHASSPSRCVIIGTFSAALAGILINNIFYTINS
B32 LLFGVFYEALIFSISASILRKYSGGVHASSPSRCVIIGTFSAALAGILINNIFYSINS
B33 VIFGVIYEALVFTITVSILKKYSGGAHASSPGRCLFLGSAISIGSSLLISKILCKQSV
B34 FIFGVIYEALIFTITVSILKKYSGGAHASSPGRCLFLGSAISTGSSLLISKILCKQSI
B35 FIFGVIYEALIFTITVSILKKYSGGAHASSPGRCLFLGSVISIGFSLLISKILLKQNI
B36 LLFGVFYEALIFSISASILRKYSGGVHASSPSRCVIIGTFSAALAGILINNIFYSINS
B37 LIFGVFYEALIFSITASILRKYSGGVHASSPSRCIIIGTFSAALAGILINNIFHTINS
B38 LIFGVFYEALIFSITASILRKYSGGVHASSPSRCIIIGTFSAALAGILINNIFYTINS
B65 VLFGVFYEALIFSITASILRKYSGGVHASSPSRCVIIGTFSAVLAGFLINNIFYSINS
B39 LLFGVFYEALIFSISASILRKYSGGVHASSPSRCVIIGTFSAALAGILINNIFYSINS
B40 LLFGVFYEALIFSISASILRKYSGGVHASSPSRCVIIGTFSAVLAGILINNIFYSINS
B41 LLFGVFYEALIFSISASILRKYSGGVHASSPSRCVIIGTFSAALAGILINNIFYSINS
B42 LLFGVFYEALIFSISASILRKYSGGVHASSPSRCVIIGTFSAALAGILINNIFYTINS
B43 LLFGVFYEALIFSISASILRKYSGGVHASSPSRCVIIGTFSAVLAGILINNIFYSINS
B44 LLFGVFYEALIFSITASILRKYSGGVHASSPSRCVIIGTFSAALAGILINNIFYTINS
B45 LLLGVFYEALIFSISASILRKYSGGVHASSPSRCVIIGTFSAALAGILINNIFYSINS
B46 LLFRVFYEALIFSITASILRKYSGGVHASSPSRCVIIGTFSAVLAGILINNIFYTINS
B47 LLFGVFYEALIFSITASILRKYSGGVHASSPSRCVIIGTFSAALAGILINNIFYTINS
B48 LLFGVFYEALIFSITASILRKYSGGVHASSPSRCVIIGTFSAALAGILINNIFYTINS
B49 SIFGVLYEALIFTISVAILRKYSGGAHASSPIRCIIIGMLLAVAVGFSVDKILYKFNI
B50 FIFGVIYESLVFTIIVSILKKYSGGAHASSPGRCLFIGSVISIGLSLLISKILCKQSV
B51 LIFDVLYESLLFTISISILKKYSGGAHASSSGRCLFIGNAISVSFSLVISKILCKRNI
B1 EVLGIIKGTLLIMLSFTFLRRYAFGVHAKSSIKCTSITSFFFFIGGFISK--YLI---
B2 FLLGILQKTLTIMLSFAFIRRYAFGIHAKNSISCTILTSIYFFTGAYIPN--FIH---
B6 SHFNILKEALFMIFIFGSIRSTAFGLHAKSSVVCTLITSMMFVVGPYVSY--HIK---


Label
B5
B7
B8
B9
B10
B11
B12
B13
B14
B15
B16
B17
B18
B19
B20
B21
B22
B23
B52
B53
B54
B55
B56
B57
B58
B59
B60
B61
B62
B63
B64
B29
B3
B4
B24
B25
B26
B27
B28
B30
B31
B32
B33
B34
B35
B36
B37
B38
B65
B39
B40
B41
B42
B43
B44
B45
B46
B47
B48
B49
B50
B51
B1
B2
B6

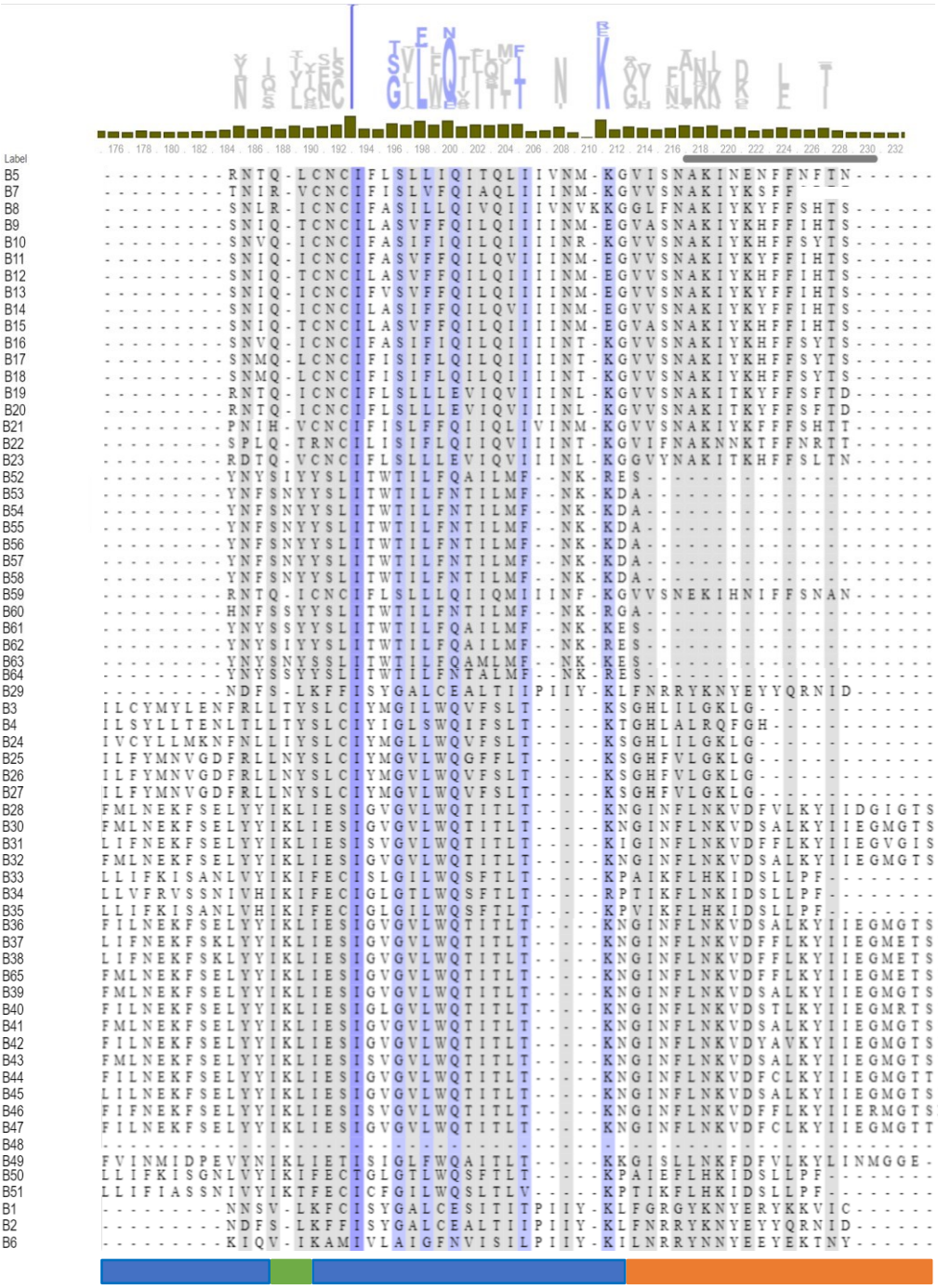


Figure 3.3 Multiple sequence alignment of all the AgrB homologues identified in *C. beijerinckii* strains within this study.

The alignment was conducted using the tool Clustal Omega with a 5x iteration (Sievers *et al.*, 2011) and then visualised using MSAviewer with PID colour scheme (Yachdav *et al.*, 2016). The size of the AA residue above the sequence data indicates the level of conservation, with larger font size showing a more highly conserved residue. Blue shading indicates the most highly conserved residues. Grey shading indicates lower levels of conservation. Along the bottom, blue bars represent where the general transmembrane regions occur, orange bars represent inner-membrane regions on the cytoplasmic side and green bars represent outer membrane regions, as predicted by TOPCONs (Bernsel *et al.*, 2009).

3.2.4 AgrBDC protein evolutionary clusters

Three maximum likelihood phylogenetic trees were constructed according to the multiple Clustal Omega alignment data generated using AgrB, AgrD and AgrC homologue protein sequences observed in this study. The trees, shown in **Figure 3.4, 3.5** and **3.6** demonstrate the divergence and evolutionary relationship between different protein homologues. Comparison of **Figures 3.4, 3.5** and **3.6** demonstrates the cognate AgrBDC, obtained from the same identified *agr* loci, which have been labelled respectively to each other (e.g. B1:D1:C1).

The phylogenetic trees demonstrate how protein homologues, obtained from the same Agr groups in **section 3.2.1**, tended to form evolutionary clusters and are therefore closely related. For instance, **Figure 3.4** shows that the AgrB homologues B33, B34, B35, B50 and B51 form a cluster and are all from systems within the Agr11 group (see **Figure 3.1**). Comparison to **Figure 3.5** and **Figure 3.6** also demonstrates how the cognate AgrD and AgrC form into similar corresponding clusters to their AgrB counterparts. The similar evolutionary clustering pattern of cognate AgrBDC homologues indicates that the associated proteins, deriving from the same Agr systems, evolved alongside each other as a unit rather than as independent units. It appears that there may have been very limited or no swapping of *agrD* or *agrC* genes between different systems, and thus signal and receptor remained linked.

The AgrC protein homologues represent the sensor histidine kinase which interact with the cognate AIP. In some cases, the AgrC identified from different Agr groups were closely phylogenetically related. This was true for AgrC homologues identified from Agr4, Agr10 and Agr11 type systems. This indicated reduced variance between the AgrC homologues found between these groups which could result in AIP cross-interaction.

One group of AgrD homologues, D0.1, D0.3, D0.4, D0.5 and D0.6, all categorised into the Agr7 group, did not appear to have any cognate AgrB, however the cognate AgrC did appear to form into a corresponding evolutionary cluster. This once again suggests the AgrD and AgrC in this orphan-type system evolved as a unit however they lost the AgrB component. The AgrDC from the Agr7 group did appear to have a closely related ancestral sequence to the respective homologues from the Agr2 group which indicates a possible origin linking to this group.

The AgrBD proteins deriving from group Agr6, are largely unrelated to the corresponding components of different groups. It is possible that these systems could have originated from another bacterial species leading to the assumption that these types of systems may have been gained through horizontal gene transfer.

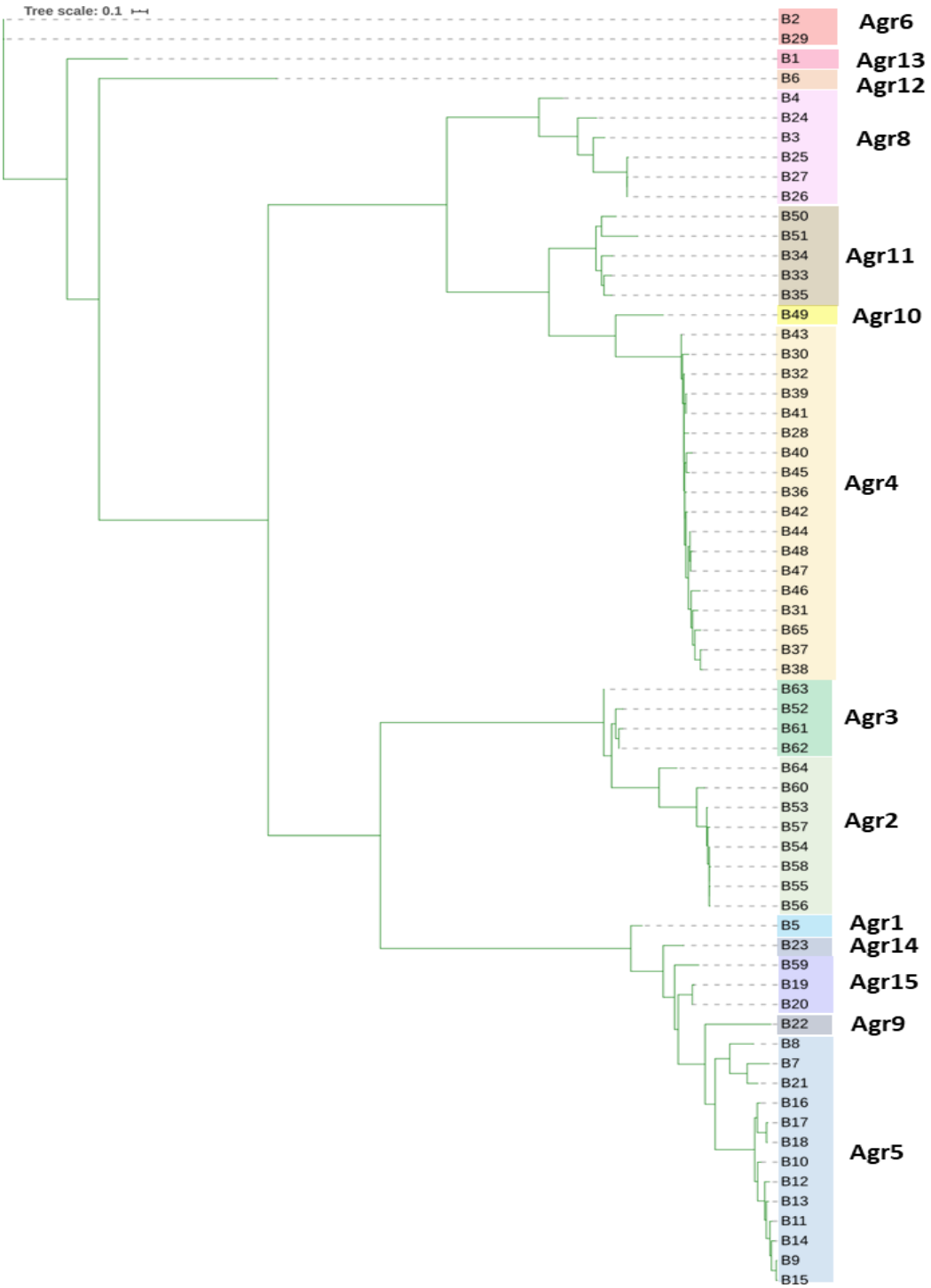


Figure 3.4 Phylogenetic tree for the different AgrB homologues identified in this study.

All AgrB homologues were identified from *C. beijerinckii* strains (taxid 1520) from the NCBI database. The tree was generated using NGPhylogeny and was based on the Clustal omega alignment data in Figure 3.3. Numbers indicate the respective AgrD and AgrC found within the same identified system (B1:D1:C1). The Agr groups for which the different homologues derived from are indicated on the right-hand side.

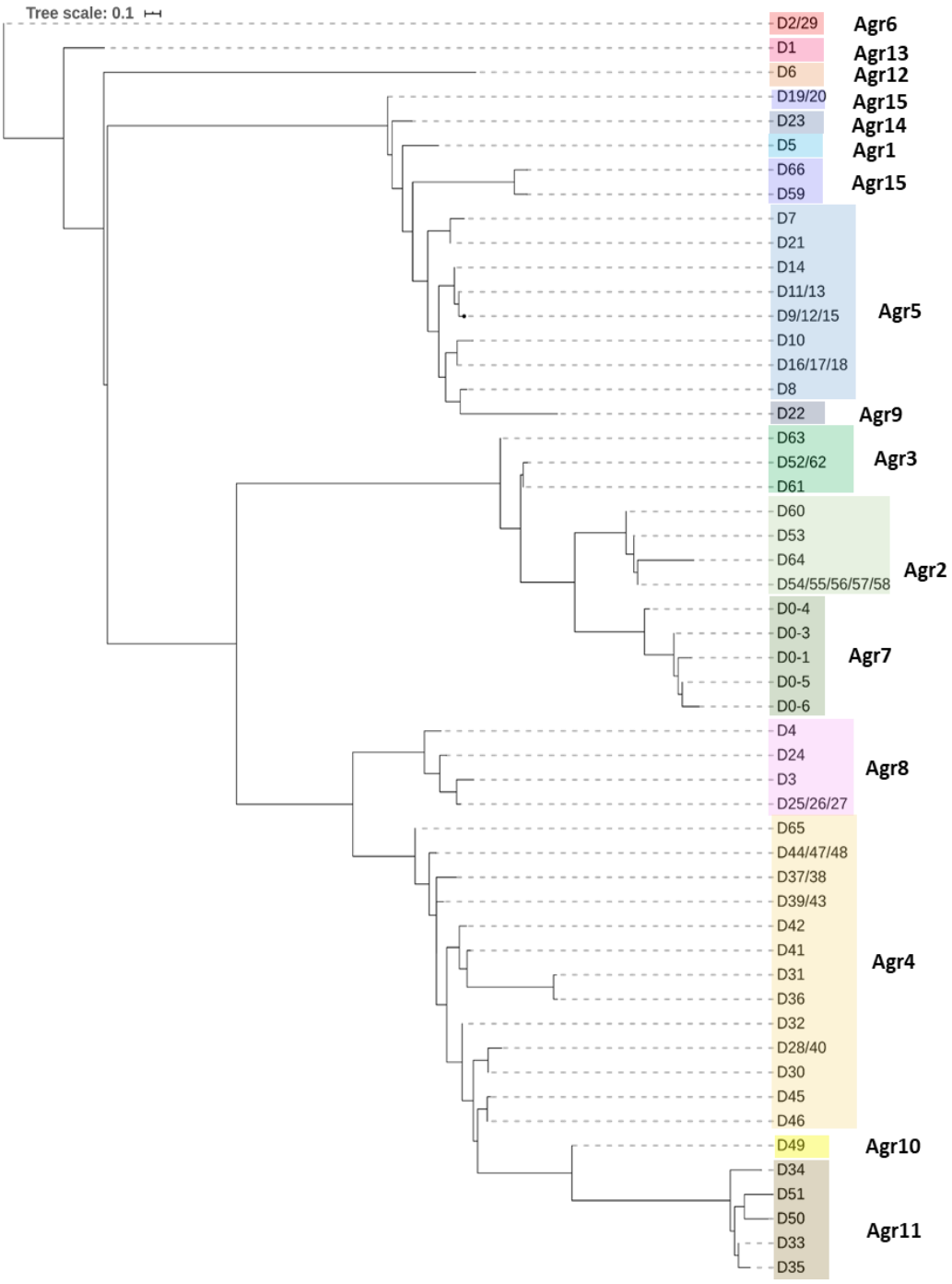


Figure 3.5 Phylogenetic tree for the different AgrD homologues identified in this study.

All AgrD homologues were identified from *C. beijerinckii* strains (taxid 1520) from the NCBI database. The tree was generated using NGPhylogeny and was based on the Clustal omega alignment data in Figure 3.2. Numbers indicate the respective AgrB and AgrC found within the same identified system (B1:D1:C1). The Agr groups for which the different homologues derived from are indicated on the right-hand side.

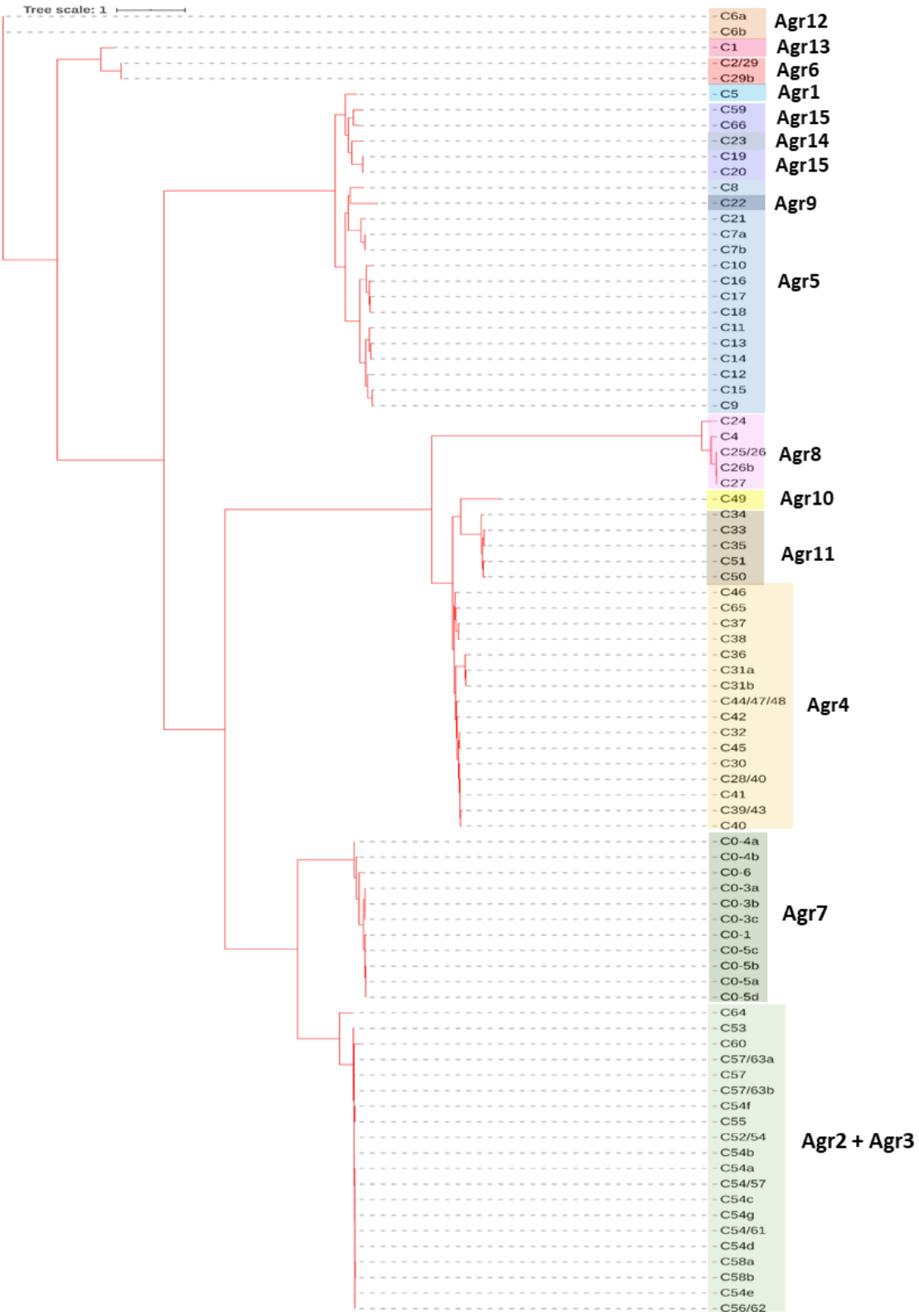


Figure 3.6 Phylogenetic tree for the AgrC homologues identified in this study.

All AgrC homologues were identified from *C. beijerinckii* strains (taxid 1520) from the NCBI database. The tree was generated using NGPhylogeny and was based on Clustal omega alignment data. Numbers indicate the respective AgrB and AgrD found within the same identified system (B1:D1:C1). The Agr groups for which the different homologues derived from are indicated on the right-hand side.

3.3 Discussion

The Agr QS system, first identified in the genus *Staphylococcus* (Novick and Geisinger, 2008), is common in Gram-positive bacteria including many *Clostridium* species. The clostridial Agr QS systems have largely been studied based on bioinformatic studies and genome annotations however, have been experimentally confirmed in a select few species such as *C. perfringens* (Li *et al.*, 2011; Vidal *et al.*, 2012; Yu *et al.*, 2017; Navarro *et al.*, 2020), group I *C. botulinum* (Cooksley *et al.*, 2010), *C. difficile* (Darkoh *et al.*, 2015) and *C. acetobutylicum* (Steiner *et al.*, 2012). This study aimed to identify putative *agr* genes in the sequenced genomes of strains belonging to the industrially important species *C. beijerinckii*. Each strain had multiple *agr* gene clusters, usually between 2 and 7 and so it is likely that each strain has more than one functional system.

3.3.1 Agr gene arrangement and flanking genes of unknown function

The *agr* loci in *S. aureus*, *L. monocytogenes* and *C. acetobutylicum* contain *agrBDCA* genes arranged respectively, occurring in the same orientation (Autret *et al.*, 2003; Novick and Geisinger, 2008; Steiner *et*

al., 2012) which is in contrast to the arrangement of *agrACDB* in *C. difficile* (Martin *et al.*, 2013). For *C. beijerinckii* strains analysed in this study, the same *agrBDCA* genomic arrangement of complete *agr* loci was identified. However, a common observation in these complete systems was that they existed in two forms. One form was agrBDCA all in the same orientation (groups 6, 12, 13) and another form was two separate operons, *agrBD* and *agrCA* genes in convergent orientation (groups 1, 5, 9, 14, 15). This observation could suggest that the two pairs, *agrBD* and *agrCA* (or *agrAC*, depending on the cluster), at least in their convergent orientation, form two independent transcriptional units where the genes from the respective pairs maintained the same orientation, a scenario which has been experimentally proven in *C. acetobutylicum* (Paredes, Rigoutsos and Papoutsakis, 2004; Steiner *et al.*, 2012).

Some systems identified only appeared to encode a partial *agr*-like system consisting of either *agrBD*, *agrDC* or *agrBDC* genes. The *agr* loci containing partial *agr*-like systems with only putative AgrDC identified were those from Agr7. These were denoted as potential orphan AgrD, since it is generally accepted in the literature that the AgrB protein is required to process AgrD into the mature AIP (Novick and Geisinger, 2008; Le and Otto, 2015). Given their high degree of conservation in the different strains it appears likely that the encoded AgrD peptide is processed by AgrB proteins encoded by other systems in the same organism.

These “partial” Agr systems have been identified in *C. beijerinckii* previously, for example Patakova *et al.* (2019) found that *C. beijerinckii* NRRL B-598 contained an *agr* loci that lacked the gene encoding the cyclic AIP (*agrD*). In the case of the *agrBD* partial systems, these have been identified before in *C. botulinum* Group I and *C. sporogenes* which appear to only encode AgrBD with putative histidine kinase genes identified nearby (Cooksley *et al.*, 2010) and also similar to the *agr1* locus found in many *C. difficile* strains (Stabler *et al.*, 2009; Martin *et al.*, 2013). Likewise, *C. perfringens* lack AgrCA encoding genes and some studies have suggested that instead the VirSR regulatory system holds their functional equivalent (Chen *et al.*, 2014; Ma, Li and McClane, 2015). It could therefore be the case that surrounding genes of unknown function could carry out the functional equivalent of the missing *agrCA* genes in the *C. beijerinckii* analysed. Several scenarios, similar to those previously proposed, are possible in *C. perfringens* where other as of yet unidentified two-component systems may sense the generated AIP, or that the AIP is sensed by AgrCA two-component systems that are part of other agr clusters in the same organism, or the AIP is produced with the purpose of cross-inhibiting/activating QS in other species. For example, some partial *agr* gene clusters identified in *C. beijerinckii* had regulatory proteins situation within close proximity which could potentially contribute to the functionality of the *agr* gene regulation. One such regulatory protein, occurring upstream of most Agr4-type systems, was a TetR/AcrR family transcriptional regulator (Accession WP_012059356.1). Proteins of this family typically repress

the expression of an adjacent operon. TetR/AcrR transcriptional regulators have been known to regulate QS pathways, for instance LuxIR regulatory proteins in Gram-negative bacteria are members of this family. It may be that the TetR/AcrR family protein identified upstream of the *agr* locus in *C. beijerinckii* strains holds a role in Agr QS and plays a similar function to that of AgrA. These regulatory proteins are known to participate in various types of regulatory networks in Gram-positive bacteria also, an example being cellular differentiation or sporulation in *Streptomyces* (Ramos *et al.*, 2005).

Genes of unknown function could also contribute to the functionality of the *agr* loci. The Agr4-type systems appeared to have an additional hypothetical protein encoding gene within the *agr* locus itself. The hypothetical gene, situated between the *agrB* and *agrD* genes, was highly conserved within the Agr4-type systems of different strains. This hypothetical gene had no characterised function however, due to its positioning within the *agr* locus, it is likely to play a role linked to this Agr system. One explanation may be that it inactivates the agr4-type systems. In *Staphylococcus*, the incorporation of a hypothetical gene or transposase into the *agr* locus can result in loss of Agr QS function which has been indicated to increase antibiotic resistance and biofilm formation (Choudhary *et al.*, 2018). It may therefore be the case that loss of a specific Agr system in *C. beijerinckii* can provide an unidentified fitness advantage.

3.3.2 AgrB AA Residues

Studies on various AgrB identified in *Staphylococcus*, *Listeria* and *Clostridium* species identified 7 key residues are completely conserved: G32, G75, H77, A78, C84, P125 and P132 (Thoendel and Horswill, 2013). This study identified residues which correlate to all of these key residues within the putative AgrB AA sequences of *C. beijerinckii.* For example, in this study G75, H77, C84, P125 and P132 were identified as G83, H85, C92, P139 and P146 respectively. Meanwhile, residues resembling G32 and A78 were present in some AgrB homologues, but not highly conserved. In *Staphylococcus*, it has been demonstrated that AgrB acts as an endopeptidase which cleaves the AgrD C-terminal tail. This specific process was found to be abolished upon mutation of residues H77 and C84 in AgrB, showing that these residues are the catalytic centre in this process (Qiu *et al.*, 2005). The same histidine and cysteine residues have previously been found to be completely conserved in AgrB of other species such as *L. monocytogenes, B. cereus, C. acetobutylicum* and *C. perfringens* (Qiu *et al.*, 2005). In this study the respective H85 and C92 residues were also completely conserved and predicted to be on or within close proximity to the hypothetical cytoplasmic membrane regions, like in *Staphylococcus* (Qiu *et al.*, 2005; Thoendel *et al.*, 2011). It is therefore likely these residues play a critical role in the cleavage of AgrD, indicating that AgrB in *C. beijerinckii* might also be a putative endopeptidase. Topological studies of AgrB, conducted using *E. coli*, have demonstrated that both

the N- and C-terminals reside inside the cell, as was predicted for *C. beijerinckii* in this study (Zhang, Lin and Ji, 2004).

3.3.3 AgrD AA Residues

The AIP region of the AgrD peptide is variable across different species however, in *S. aureus*, which produce a 7-9 AA AIP, a completely conserved cysteine residue is required for the formation of a thiolactone ring (Novick and Geisinger, 2008; Thoendel *et al.*, 2011). An equivalent cysteine residue, sometimes replaced by a serine, was found in the putative AgrD sequences for *C. beijerinckii* strains situated 10 or 11 AAs upstream of a highly conserved proline. In this study, based on SignalP-5.0-based predictions, the bond between this cysteine/serine and the preceding amino acid residue was proposed to be a likely AIP cleavage site. The same proline residue is highly conserved across a variety of bacterial species including *Staphylococcus sp., Bacillus sp., Clostridium sp*. and *Listeria sp.* as examples (Thoendel *et al.*, 2011).

In *S. aureus* the negatively charged C-terminal region is a highly conserved portion of the AgrD. It has previously been demonstrated that this portion of the AgrD peptide plays a critical role in the production of the mature AIP by mediating the peptide-protein interaction with AgrB. Key AA residues like Asp-33, Glu-43 and Leu-41 were found to be required for AgrB endopeptidase activity and AIP

production (Thoendel and Horswill, 2009). For *C. beijerinckii*, the overall charge of the AgrD C-terminal region varied between the different identified Agr groups. However, similar AA residues were identified in this study, in particular a leucine (L), resembling Leu-41 in *S. aureus,* and two proline residues. Both the highly conserved proline and leucine residues have been previously identified in the AgrD C-terminal from various other species including, but not limited to, *L. monocytogenes, B. cereus, C. acetobutylicum, C. botulinum, C. perfringens* and *C. difficile* (Thoendel *et al.*, 2011). The understanding that these highly conserved residues play an important role in AIP processing in other species, it is likely to be the case also in *C. beijerinckii*.

The bioinformatic analysis indicated that three types of central AIP likely occur in *C. beijerinckii.* Excluding arbitrary AIP tails, the most common AIP type predicted across the Agr systems identified was a 5-mer cysteine-based cyclic thiolactone AIP, variants of which were identified in all strains analysed. These finding were not unexpected, for instance, a previous study on *C. perfringens* concluded that the natural AIP is likely a serine/cysteine-based 5-mer AIP (Ma, Li and McClane, 2015). Likewise, *C. botulinum* and *C. sporogenes* were proposed to produce various 5-mer cysteine-based thiolactone AIP-rings (Cooksley *et al.*, 2010).

It was also found that all strains had the potential to produce variants of 5-mer serine-based AIP. However, as to whether these serine-based AIP would be able to form cyclic lactone structures remains unclear. Such serine-based AIPs have been experimentally identified in other

species, such as *B. halodurans*, *C. difficile* and *S. intermedius* (Ji *et al.*, 2005; Thoendel *et al.*, 2011). However, in *C. beijerinckii* strains analysed here, key leucine and proline residues, shown to be required for AIP processing, were replaced with the alternative AA residues methionine and glutamate in the C-terminus of the AgrD peptides predicted to generate these serine-based AIPs. This would likely affect how these AgrD proteins interact with the respective AgrB during the production of a mature AIP. Whilst linear AIPs have been demonstrated to induce signalling in some Agr QS systems (Ma, Li and McClane, 2015), for most Gram-positive Agr systems it is generally agreed that the AIPs contain a thiolactone or lactone ring (Novick and Geisinger, 2008).

An interesting hypothesis might be that *C. beijerinckii* strains can sense both serine-based lactone and cysteine-based thiolactone AIPs, since all strains analysed appeared to have the potential to produce both cysteine and serine-based AIP structures. This ability to sense both lactone and thiolactone AIPs has been demonstrated before in *S. intermedius,* which all produce serine-based cyclic lactone AIPs, but could still sense cysteine-based AIPs and cross-inhibit other AIPs (Ji *et al.*, 2005). A suggestion was that the ability of *S. intermedius* to sense both thiolactone and lactone AIPs indicated evolutionary memory. It was likely therefore that the cysteine-serine switch occurred during evolution due to a single nucleotide substitution (Ji *et al.*, 2005). It is not illogical to suggest this substitution may have occurred in *C. beijerinckii* evolution.

The third and less common AIP-type identified was a 6-mer cysteine-based AIP. The idea of 6-mer cysteine-based AIP has been suggested as the natural AIP structure for *C. acetobutylicum* (Steiner *et al.*, 2012). Whether *C. beijerinckii* produces all or just one of these AIP types would need to be experimentally confirmed. However, diversity in the signalling molecule is not uncommon for families of QS systems. In Gram-positive bacteria the QS families can encode for many signalling molecule variants, often with intraspecific diversity (Aframian and Eldar, 2020). The family of Agr QS systems can have more than 10 signal variations in the same species which can vary in sequence, N-terminal tail and cyclisation moiety, making up the main factors which determine activation or inhibition properties of AIPs (Novick and Geisinger, 2008; Aframian and Eldar, 2020).

3.3.4 Coevolution of AgrBDC

This study suggested that *agrBDC* genes deriving from the same systems likely coevolved together as a unit rather than exchanging between different systems. Coevolution of receptor and corresponding inducer has been demonstrated for other QS systems also. For example, a study by Lerat and Moran (2004) demonstrated how the inducer and receptor of the LuxIR QS systems coevolved as a unit and only rarely had exchange of partners. This study also identified systems which may have lost partner components. The systems of group Agr7 are termed orphan systems as they only appeared to consist of

agrDC. The same study by Lerat and Moran found that in some lineages of γ- Proteobacteria the loss or replacement of partner components had occurred in LuxIR systems. This study showed how Agr7 protein components were closely related to those from Agr2 systems, which would suggest an origin for these orphan systems.

Another interesting point from this study was the possibility that some of the *agr* loci could have been gained through horizontal gene transfer. The Agr6 group protein components were largely unrelated to the other systems identified in this study. This suggests that these *agrBD* genes were horizontally transferred from another species, however this is only a suggestion based on evolutionary relationship observed here.

3.3.5 Accumulation of multiple QS systems and AIP variants

Multiple putative *agr* loci were identified in all *C. beijerinckii* strains analysed in this study. These putative Agr systems were categorised into groups according to the genomic arrangement of neighbouring genes. It was found that a diverse pool of putative AIP AA sequence variations across different Agr groups were predicted in the *C. beijerinckii* strains, demonstrating intraspecific diversity. Some of these AIPs were highly similar to each other whereas others were highly dissimilar. The ostensible diversity between putative AIPs identified in

this study suggests that there is a selection pressure in favour of accumulating multiple QS systems with variable AIP sequences/structures. In other species it has been established that natural microbial communities contain several QS system variants (Even-Tov *et al.*, 2016). It has previously been suggested that this diversity in QS signal molecules is maintained due to beneficial social interactions (Aframian and Eldar, 2020). A study by Pollitt *et al.* (2014) found that in *S. aureus*, QS via the *agr* locus allowed a population to grow at higher cell densities when compared to *agr* mutant strain cultures, demonstrating a beneficial cooperative social interaction must have occurred through Agr QS.

Many bacteria use multiple QS systems to regulate a cooperative social response that results in the production of 'public goods', extracellular products, beneficial to the whole population, allowing the coexistence of multiple strains within a microbial community (Even-Tov *et al.*, 2016). However, multiple systems can also allow the phenomenon of social cheating to occur. This is where one subpopulation takes advantage of goods produced by other subpopulations, without directly contributing to the production of these goods themselves (Kalamara et al., 2018; Kotte et al., 2020). A study by Pollak *et al.* (2016), showed how the AIP diversity of the *B. subtilis* ComQXPA QS system is maintained by the social cheating and exploitation mechanism. They found that AIP diversity in a mixed population allowed the minority subpopulation to exploit the majority subpopulation for surfactin production. This is because the minority were not contributing to cooperation through their

QS system yet were still benefiting from surfactin production from the majority. Facultative social cheating could therefore contribute to the accumulation of multiple Agr QS systems and AIP variation in *C. beijerinckii*. This idea of facultative social cheating may also explain why some strains from this study have more unique Agr systems identified. For example, the Agr13 only appears in *C. beijerinckii* HUN142. Considering the diversity of the AIPs in these systems compared to other systems identified in variant strains it is likely that these distinct systems allow strain-specific cooperation. If this was the case, these specific strains could exploit other strains within a majority population by using up 'public goods' produced by the majority population without expending their own energy.

3.3.5.1 Cross-activation and inter-strain communication

Cross-activation has been observed in other bacterial species, for instance between the 4 types of *Staphylococcus* Agr systems. The similarity between *S. aureus* AgrI and IV type AIPs allow for such cross-reactivity between the different systems (Thoendel *et al.*, 2011).

This study identified conserved Agr groups which had putative AIP variants with highly similar or identical AA sequences and structures. In particular, the genetic loci categorised into groups Agr2 and Agr5 were highly conserved amongst the different strains and were found to have lower

degrees of both AIP and AgrC variation between systems in the same group. For instance, In the Agr2-type systems, all but one strain was found to have identical putative AIP AA sequence and structure. This would suggest that any strain encoding the *agr2* loci would be able to communicate through this system, therefore allowing inter-strain communication.

This study also showed high sequence similarity of AIPs when comparing between groups Agr4, Agr10 and Agr11 and also between Agr1, Agr14 and Agr15. Previous literature has suggested that AgrC receptors with highly similar AA sequences could potentially cross-interact with non-cognate AIPs with a reduced sensitivity (Aframian and Eldar, 2020). This is the idea of cross-activation of QS systems. This would suggest that the highly similar AIPs identified between Agr groups from this study could cross-activate non-cognate AgrC receptors, with reduced sensitivity. The idea that *C. beijerinckii agr* loci are highly responsive to each other would suggest that Agr QS systems could be important for inter-strain communication that potentially regulates important biological processes.

However, whilst cross-activation is evident for some species, it has been observed that cross-activation is more difficult to achieve compared to cross-inhibition. This is because with an activating signal there is still a certain level of AgrC specificity required for partial interaction (Olson *et al.*, 2014).

3.3.5.2 Cross-Inhibition

When multiple AIP variants are simultaneously expressed within a given microbial community, non-cognate AIPs can interfere with non-cognate receptors from different strains and species, altering their signalling behaviors (Mehta *et al.*, 2009; Yusufaly and Boedicker, 2017). Highly diverse AIP variants have been linked to the spectacle of cross-inhibition that can occur between QS systems of the same family. For example, the cross-inhibition of Agr signalling between AgrI and II types, as well as AgrI and III types due to the diversity of AIP variants observed between various *Staphylococcus* species and different *S. aureus* strains (Ji, Beavis and Novick, 1997; Novick and Geisinger, 2008; Olson *et al.*, 2014; Canovas *et al.*, 2016). Furthermore, in *S. intermedius* serine-based lactone AIPs were found to cross-inhibit Agr activation in both *S. aureus* and other *S. intermedius* strains, which had highly divergent AIP sequences (Ji *et al.*, 2005). This interference mechanism has also previously been demonstrated in *C. perfringens* where a synthetic 6-mer thiolactone AIP interfered with signalling for toxin production via Agr QS (Ma, Li and McClane, 2015). This study identified that signalling molecules categorised into different Agr groups tended to have more diverse AIP structures or AA sequences, which supported the idea that cross-inhibition could potentially occur between the different Agr groups identified in *C. beijerinckii*.

An interesting finding from this study was the close association of genes in the Agr2 and Agr3 systems to the same *agrC* gene, which would suggest that the AIPs from these two groups both interact with the same AgrC. However, comparison of the AA sequence of the AIPs from these two groups showed diversity between the signal molecules which would make interaction with the same AgrC protein unlikely. Alternatively, due to the close association of these systems, it may be that these systems potentially influence the expression of the other by cross-inhibition. The close evolutionary relationship of the AgrBD proteins from these groups, and the proximity of their genomic loci suggests they could hold a similar function and influence similar pathways. However, this is just speculation and experimental evidence is needed to support this.

Chapter 4:

Generation of *agr* mutants

4.1 Introduction

The *Clostridium* genus contains many industrially important strains with *C. beijerinckii* having been one of the most used *Clostridia* species for the production of solvents such as biofuels (Keis *et al.*, 1995; Lee *et al.*, 2009; Poehlein *et al.*, 2017; López-Linares *et al.*, 2021; Zuleta-Correa, Chinn and Bruno-Bárcena, 2022). Therefore, there is a desire to genetically manipulate this organism to improve the strain further for large-scale industrial application. However, due to the lag in efficient genetic tools available in *Clostridium* species, targeted gene manipulation has proven to be a challenge.

Established methods include the use of mobile group II introns known as the ClosTron and homologous recombination (HR) technologies such as *pyrE*-based allele exchange. The ClosTron methodology has been widely used for targeted gene disruption in many clostridia, including solventogenic types like *C. beijerinckii*, making it a good starting point for work in this organism (Heap *et al.*, 2010; Steiner *et al.*, 2012; Liew *et al.*, 2017; Little *et al.*, 2018).

HR technologies have also been used in a variety of *Clostridium* species. In the earlier days of genome editing, the use of HR was not often used on solventogenic clostridia due to low efficiency and inability to produce markerless deletions (Harris, Welker and Papoutsakis, 2002). However, HR technology has been updated to increase editing efficiency in solventogenic clostridia, such as *C. beijerinckii* NCIMB 8052, through the development of a clostridia-specific replicable

plasmid which was able to integrate into the target chromosome via a double-crossover event. The use of allelic-coupled-exchange (ACE) and the *pyrE* system involve these updated technologies (Heap *et al.*, 2012; Ehsaan *et al.*, 2016; Lee *et al.*, 2016; Little *et al.*, 2018). The updated HR technologies hold advantages over the use of ClosTron by allowing for a scarless deletion or insertion of the target gene rather than just a disruption. However, HR technologies can be time-consuming and inefficient in some species and strains (Shanmugam, Ngo and Wu, 2020).

To further unlock the capabilities of *Clostridium* species, further progression has resulted in more efficient genetic tools for target genes, particularly the CRISPR-Cas9 system. As previously described in **section 1.10.3**, the CRISPR-Cas9 system relies on the Cas9 nuclease guided to its target via sgRNA. This therefore overcomes the issue of inefficiency of HR-dependent techniques (Shanmugam, Ngo and Wu, 2020). In recent years, the use of CRISPR-Cas9 has been practiced in solventogenic clostridia including *C. beijerinckii*, for target gene knock-out and knock-in (Jiang *et al.*, 2013; Xin *et al.*, 2020). However, the apparent cytotoxicity of Cas9 nuclease due to off-target mutations require the mutated forms of Cas9 to be adopted such as nCas9 and dCas9. A study by Li *et al.,* (2016) demonstrated that with the use of the plasmid pNICKclos, efficiency of genetic engineering in *C. beijerinckii* NCIMB 8052 ranged between 18.8% to 100%. Another study by (Wang *et al.*, 2016) found CRISPR-dCas9 resulted in transcriptional inhibition with an efficiency between 65-97% for a target

gene. Meanwhile Diallo *et al.*, (2020) used the Cas9 system to create scarless target gene deletion and insertion in *C. beijerinckii* NCIMB 8052 with 75-100% efficiency.

In this chapter, the various genetic tools available for targeted gene editing were explored for use in *C. beijerinckii* NCIMB 8052. The aims were therefore as follows:

- Determine which genetic manipulation technique to use when targeting *agr* genes in *C. beijerinckii* NCIMB 8052.
- To inactivate a subset of three Agr systems, representing those most commonly found in *C. beijerinckii* strains based on the analyses undertaken in Chapter 3.

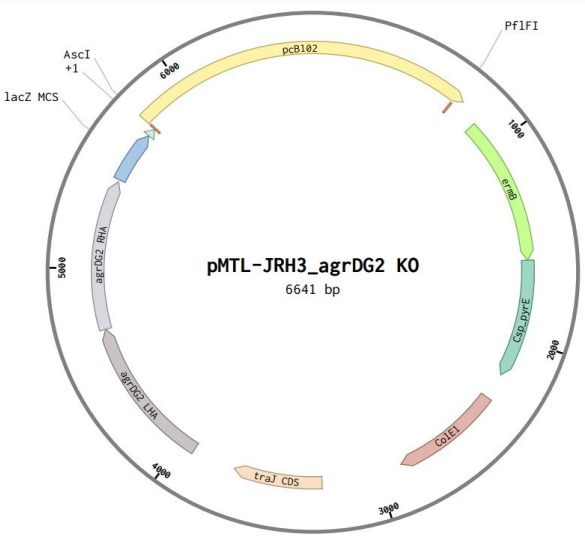
4.2 Results

4.2.1 Targeted *agrD* gene mutation via allelic exchange and ClosTron in *C. beijerinckii* NCIMB 8052

The genomes of various *C. beijerinckii* strains were analysed for *agr* gene clusters, as described in **chapter 3**. The two most conserved systems, Agr2 and Agr5, were targeted for *agrD* gene disruption in *C. beijerinckii* NCIMB 8052. The *agrD* gene was targeted as this encodes the precursor peptide for the autoinducing peptide of the *agr* system, therefore this would eliminate the signalling molecule for the target system. The use of allelic-coupled exchange, adapted from Ehsaan *et al.* (2016) and Ehsaan, Kuehne and Minton (2016), was initially adopted to create in-frame deletions of the target *agrD* genes Cbei_RS26925 *agrD5* and *agrD2*. Construction of the pMTL-JRH3 *agrD* gene KO pseudo-suicide vectors was carried out as described in **section 2.3.3.1**. These vectors contained around 750 bp upstream and downstream homology arms complementary to that of the target gene (see appendix **Figure 8.7**).

Figure 4.1 shows the plasmid map and plasmid digest using the restriction enzymes AscI and PflfI to confirm the vectors pMTL-JRH3_agrD2 KO and pMTL-JRH3_agrD5 KO used in subsequent transformations, expected band sizes were around 1.6 kb, containing the pCB102 replicon, and 5.1 kb.

A



B

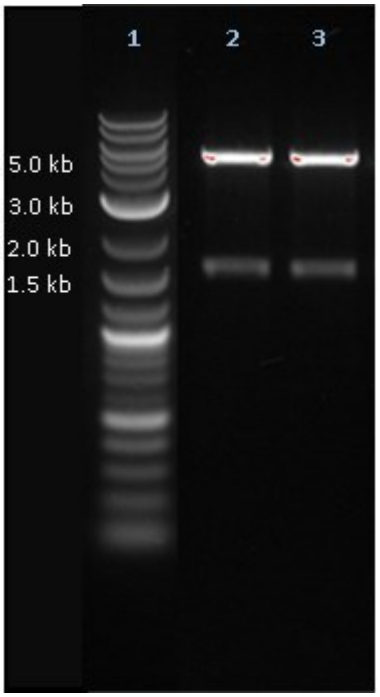


Figure 4.1 The AscI and PflfI restriction enzyme digest of pMTL-JRH3_agrD2 KO and pMTL-JRH3_agrD5 KO.

The plasmid map showing AscI and PflfI restriction sites on pMTL-JRH3_agrD2 (A).

The restriction digest run on agarose gel (B). Expected band sizes were 1.58 kb and 5.06 kb for JRH3_agrD2 KO (lane 2) and 1.58 kb and 5.11 kb for 2, pMTL-JRH3_agrD5 KO (lane 3). The size of the PCR products was estimated using a 1 % agarose gel, run for 40 minutes at 100 V, 400 mA and using a 2-log 1kb DNA ladder (lane 1).

The *pyrE* mutant of *C. beijerinckii* NCIMB 8052, previously created by a PhD student (Humphreys, 2019), was used in initial attempts to create the in-frame *agrD* mutations, this will be referred to as *JRH-pyrE*. The JRH-*pyrE* mutant was transformed with the appropriate KO vector and transformants selected for on CGM agar + 10 µg/ml erythromycin. The largest colonies were restreaked twice onto the same medium to encourage a single cross-over event with one of the vector-based homology arms. Cells with single-crossovers would produce larger colonies as the stable integration of the plasmid, containing the erythromycin resistance gene, would allow faster growth in the presence of erythromycin selection compared to cells without an integrated plasmid (Ehsaan *et al.*, 2016; Ehsaan, Kuehne and Minton, 2016). However, single cross-over isolates were either occurring in mixed culture or most often, not at all, therefore double cross-over integration was not observed at all. This is demonstrated by **Figure 4.2** and **Figure 4.3** which shows a PCR product around 2.0 kb PCR product for suspected crossover isolates, using primer pair JRH3 insert check F and JRH3 insert check R which confirmed the presence of the intact plasmid and also shows a PCR product of 1.9 kb using primer pair c-agrD5 cross LHA check_F + c-agrD5 cross RHA check_R, which indicated the chromosome was still intact. This method therefore proved to be time-consuming and inefficient at producing the desired mutation in the strain used.

Due to the inefficiency of allelic exchange in creating the desired target mutation in the strain used, it was decided that a less time-consuming

method, already widely used in previous research, would be adopted to create a disruption in the two target *agrD* genes to ascertain phenotypes. The use of ClosTron mutagenesis was adopted and the plasmids pMTL007S-E2::Cbei_rs26925 agrD5-79|80a and pMTL007S-E2::Cbei_agrD2-65|66a synthesised and sequenced by the manufacturer (ATUM, see appendix **figure 8.7**). These plasmids, targeting *agrD5* and *agrD2* respectively, alongside pMTL007S-E2::Cbe-spo0A-407a provided from a previous study (Heap *et al.*, 2010), were transformed into *C. beijerinckii* NCIMB 8052 (Humphreys, 2019), JRH strain, and selected for plasmid uptake using CGM + 750 µg/ ml spectinomycin. Confirmed transformants were then plated onto CGM + 10 µg/ml erythromycin to select for intron insertion into the genome. Unfortunately, for the JRH strain no colonies grew on the erythromycin selection plates. This suggested that the ClosTron cassette was not inserting into target genes in this version of the strain.

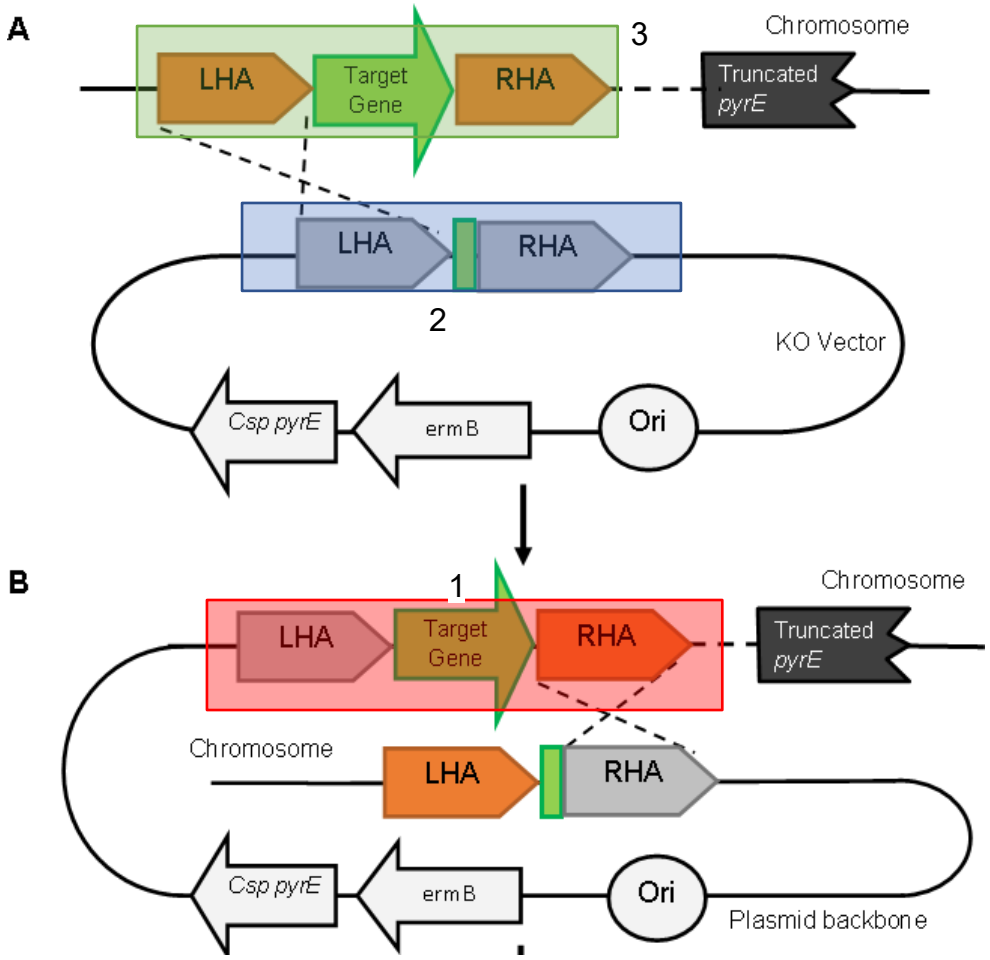


Figure 4.2 Generation of in-frame deletions using the *pyrE* system. Adapted from Ehsaan, Kuehne and Minton (2016). Non-integrated plasmid (A), single cross-over resulting in integrated plasmid into the chromosome (B).

Three different primer pairs were used to confirm or deny the presence of a single crossover mutant. The red highlight region (1) indicates primer pair JRH3 insert check_F + c-agrD5 cross RHA check_R (this screened for a single cross over isolate), blue highlighted region (2) indicates primer pair JRH3 insert check_F + JRH3 insert check_R (this screened for an intact JRH3 plasmid) and the green highlighted region (3) indicates primer pair c-agrD5 cross LHA check_F + c-agrD5 cross RHA check_R (this screened for an intact chromosome). A band present for primer pairs blue and green confirmed that there was no single-cross over isolate.

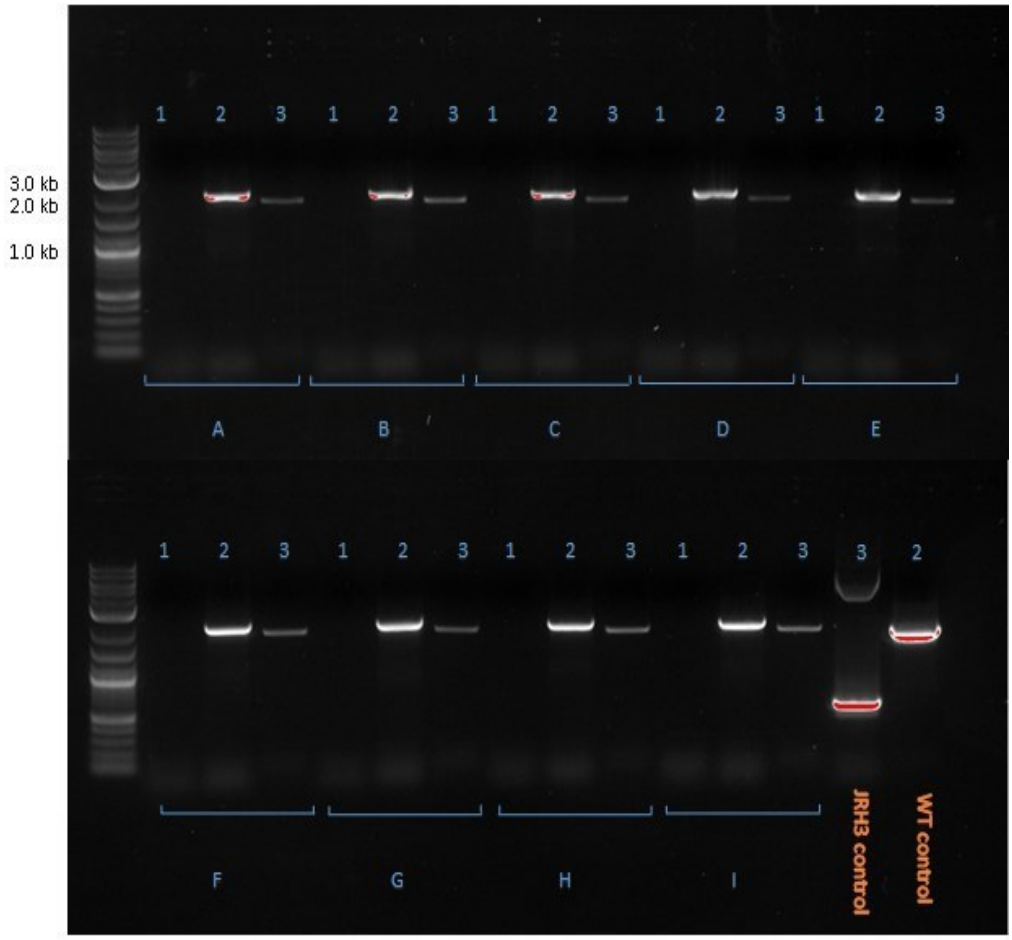


Figure 4.3 No single crossover mutants were isolated in *C. beijerinckii* NCIMB 8052 concerning the JRH3-agrD5 KO plasmid transformants after carrying out the allelic exchange technique.

In total nine individual colonies were screened, these individual isolates are indicated by letters A-I. Three different primer pairs were used to confirm or deny the presence of a single cross over mutant, these are indicated by the numbers 1-3. The 1 indicates primer pair JRH3 insert check_F + c-agrD5 cross RHA check_R (this screened for a single cross over isolate), 2 indicates primer pair NEW JRH3 insert check_F + JRH3 insert check_R (this screened for an intact JRH3 plasmid) and 3 indicates primer pair c-agrD5 cross LHA check_F + c-agrD5 cross RHA check_R (this screened for an intact chromosome). A band present for primer pairs 2 and 3 confirmed that there was no single-cross over isolate. The control reactions consisted of the plasmid JRH3 and *C. beijerinckii* NCIMB 8052 WT gDNA respectively. The size of the PCR products was estimated using a 1 % agarose gel run for 40 minutes at 100 V, 400 mA and using a 2-log 1kb DNA ladder.

4.2.2 Characterising the WT *C. beijerinckii* NCIMB 8052 (Mike Young lab)

Due to the inability to create the desired in-frame and ClosTron mutants in the JRH strains obtained from Humphreys (2019), a genetically amenable version of the NCIMB 8052 strain was chosen from the SBRC Nottingham culture collection, originally obtained from Mike Young's lab in Aberystwyth.

This strain variant, from here on referred to as *C. beijerinckii* NCIMB 8052 WT-Y, or just WT-Y, was first confirmed by 16S rRNA genes sequencing using the primers 16s forward and 16s reverse (results not shown). The sequencing result was then compared to that reported from *C. beijerinckii* strain NCIMB 8052 (GenBank accession: CP000721.1). To provide a reliable WT comparison in future studies, this strain was then grown in culture for 5 days to prepare a spore stock, as described below. Colonies were then enumerated from this spore stock and three individual colonies, WTA, WTB and WTC, were phenotypically characterised with respect to growth, granulose accumulation, spore formation and solvent profile (See Appendix **section 8.1**). This was done to generate and work with genetically homogeneous stocks and to ensure cells were displaying the organism's established phenotypic characteristics. After this preliminary phenotypic characterisation, all of the WT isolates tested showed to behave as expected for *C. beijerinckii* NCIMB 8052 (Humphreys, 2019). The isolate denoted as WTC was carried forward to use as a

master stock for all future experiments and will hereafter be referred to as WT. The genomic DNA was extracted from this WT using phenol chloroform-based extraction and sent for whole genome sequencing.

4.2.3 Using ClosTron mutagenesis for insertional *agrD* gene inactivation

The WT strain was then transformed as before with the ClosTron mutagenesis plasmids pMTL007S-E2::CBEI_agrD2-65|66a and pMTL007S-E2::CBEI_agrD5-79|80a. ClosTron mutagenesis was then carried out as described in **section 2.5.3**. A single ClosTron mutant was successfully isolated for Cbei_RS26925 *agrD5,* which was restreaked to produce multiple colonies, however no ClosTron mutants were isolated for the *agrD2* gene. **Figure 4.4** shows confirmation of the target ClosTron insertion into the *agrD5* gene region, amplified using primer pair *agrD5* KO confirmation_F and *agrD5* KO confirmation_R. The WT should produce a PCR product of 429 bp whereas those isolates with the ClosTron insertion should give a PCR product around 1.1 kb larger at 1.5 kb. Plasmid loss was tested by both patch plating onto CGM + 10 µg/ml erythromycin and CGM + 750 µg/ml spectinomycin, colonies which grew on the erythromycin but not the spectinomycin were confirmed to be ClosTon insertion mutants that had lost the plasmid.

The *Clostridium beijerinckii 8052 agrD5::CTermB* mutant was then subject to phenotypic screening. This involved analysis of growth rate, granulose formation, sporulation assays and solvent profile analysis (see appendix **8.1.3**). The isolates showed defects in sporulation and granulose formation, these phenotypes could not be genetically complemented, and this approach was discarded.

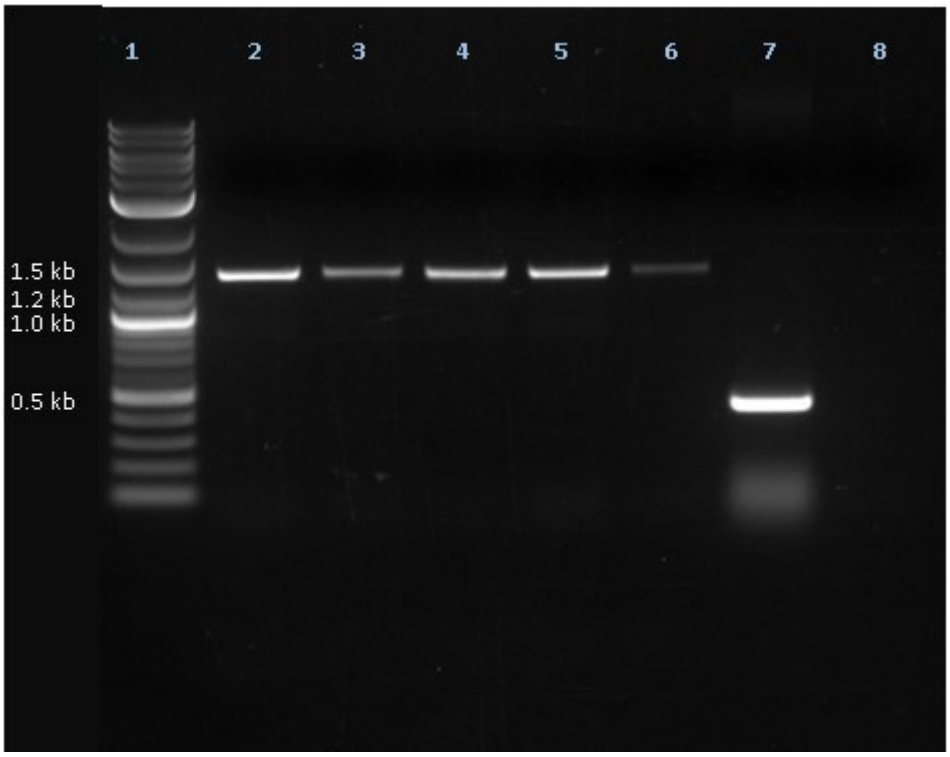


Figure 4.4 PCR confirmation of agrD5 ClosTron insertion in *C. beijerinckii* NCIMB 8052 WT.

The *agrD5* genome region in *C. beijerinckii* WT and 5 mutant strains were PCR amplified using the primer pair *agrD5* KO confirmation_F and *agrD5* KO confirmation_R. The ClosTron insertion is shown to be around 1.1 kb, with the ClosTron mutants (lanes 2-6) showing an expected increase in band size compared to the WT (lane 7). The size of the PCR products was estimated using a 1 % agarose gel run for 40 minutes at 100 V, 400 mA and using a 2-log 1kb DNA ladder (lane 1). The negative control was carried out using H2O (lane 8).

4.2.4 Creation of a *pyrE* mutant using allelic exchange in *C. beijerinckii* NCIMB 8052

Due to the ability to create ClosTron mutants in the WT strain, it was decided that allelic exchange would be attempted to create an in-frame deletion of the target genes *agrD2* and Cbei_RS26925 *agrD5*.

To carry out allelic exchange in the WT it was necessary to create a *pyrE* mutant background. This was carried out via allelic exchange by transforming the WT with the plasmid pMTL-JRH1. This plasmid contains the first 300bp and 1200bp downstream of the *pyrE* gene (Cbei_1006) from the *C. beijerinckii* NCIMB 8052 genome. Transformants were plated onto CGM + 10 µg/ml erythromycin. The largest colonies were restreaked onto CBM + 600 µg/ml 5-FOA + 2 µg/ml uracil twice to encourage the double cross-over event that would result in the truncation of the *pyrE* allele. Isolates which were resistant to 5-FOA were assumed to be *pyrE* mutants, this was confirmed by PCR amplification of the extracted gDNA using primer pair *pyrE* KO confirmation_F and *pyrE* KO confirmation_R (**Figure 4.5**) and sanger sequencing of the generated product. The expected WT PCR product should be 1.33 kb whilst the *pyrE* mutant should show a PCR product at 955 bp. Confirmed mutants were then patch plated onto CGM + 20 µ/ml uracil and CGM + 10 µg/ml erythromycin to check for plasmid loss, those which could grow on the non-selective media but not in the presence of erythromycin were confirmed to have lost the plasmid.

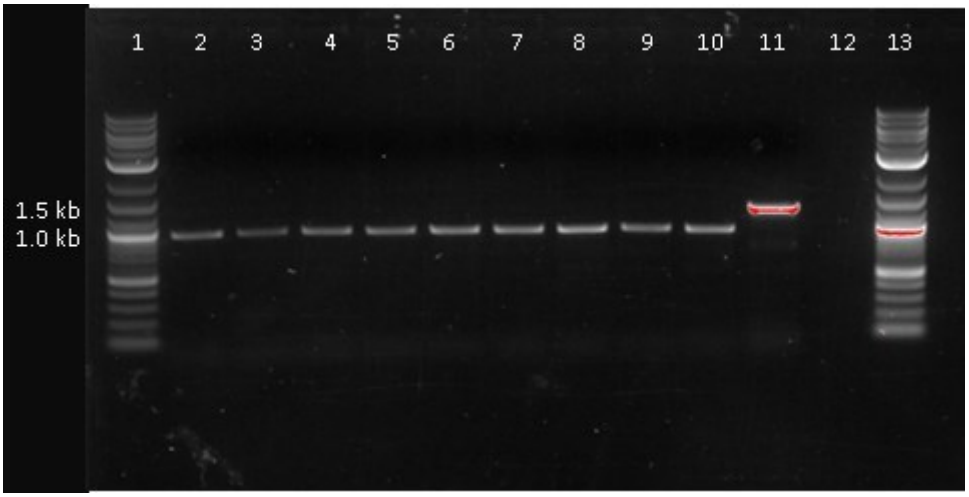


Figure 4.5 PCR confirmation of *pyrE* truncation in *C. beijerinckii* NCIMB 8052 WT.

The *pyrE* region in *C. beijerinckii* WT and 9 suspected mutant strains was PCR amplified using the primer pair *pyrE* KO confirmation_F and *pyrE* KO confirmation_R. The *pyrE* mutants (lanes 2 to 10) showed a band around 955 bp whilst the WT (lane 11) showed a larger band size around 1.33 kb. The size of the PCR products was estimated using a 1 % agarose gel run for 40 minutes at 100 V, 400 mA and using a 2-log 1kb DNA ladder (lanes 1, 13). The negative control was carried out using H2O (lane 12).

4.2.3 Phenotypic characterisation of the *C. beijerinckii* NCIMB 8052 *pyrE* mutants

Three independent *pyrE* mutant isolate colonies, denoted as *pyrE* A, *pyrE* D and *pyrE* G, were taken and fully phenotypically characterised with respect to growth kinetics, solvent profiles, granulose accumulation, and sporulation using the *C. beijerinckii* NCIMB 8052 WT as a comparative control. This was necessary to ensure that no other mutations affecting key phenotypes had spontaneously occurred.

The *pyrE* mutant isolates and the WT comparison spore stocks were enumerated and then resulting colonies used in triplicate 60 ml CBM broth cultures (+ 20 µg/ml uracil in the case of the pyrE mutant) for the main growth experiment.

Figure 4.6 shows the growth kinetics for the *pyrE* mutant isolates and the WT were highly similar. Exponential growth was observed between 0 to12 hours, with the mean growth rate calculated to be 0.444, 0.432, 0.422 and 0.453 for Δ*pyrE* A, D, G and the WT respectively between 2 to 10 hours. All cultures plateaued between an OD_{600} of 4.4 and 5.5 by 24 hours.

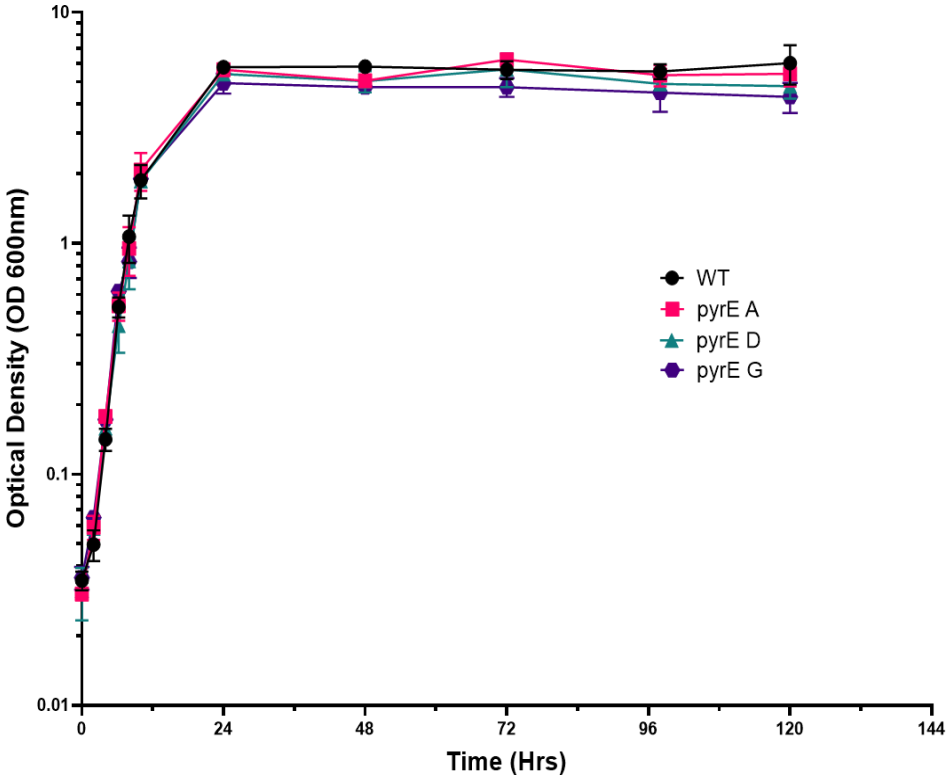


Figure 4.6 Growth of three independent *C. beijerinckii* NCIMB 8052 *pyrE* mutant isolates and the WT control.

The three isolates are indicated by *pyrE* A, D and G, these were grown in CBM broth + 20 µg/ml uracil. The WT culture was set up using CBM broth without uracil supplement. The values plotted are the means with standard error of the mean (SEM) from technical triplicates.

Figure 4.7 The solvent profile for the three *pyrE* mutant isolates alongside the WT control. Whilst the solvent profile for the isolate *pyrE* A and *pyrE* G appear to be similar to the WT, it appears that the overall butanol production in the isolate *pyrE* D was lower, with a difference of around 30 mM, than that observed in the WT. The *pyrE* mutant showing the most similar profile to the WT was the *pyrE* A isolate.

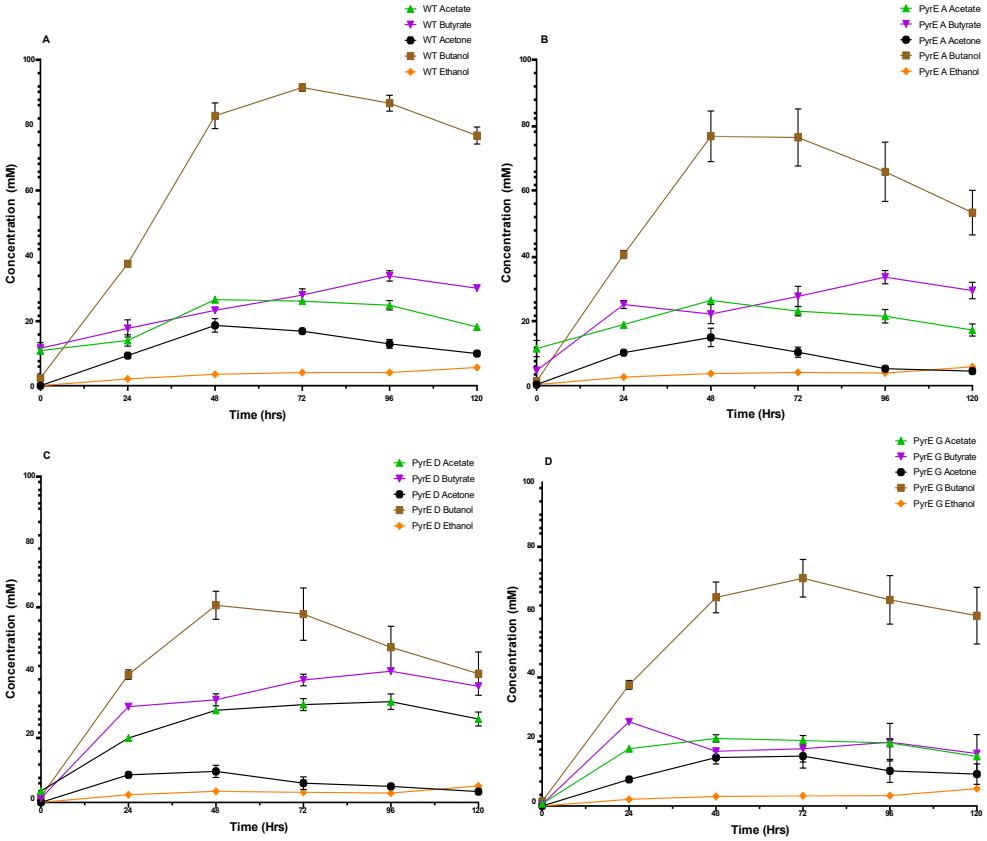


Figure 4.7 Fermentation product profiles for three *independent C. beijerinckii* NCIMB 8052 *pyrE* mutant isolates and the WT control.

Samples were grown in CBM broth + 20 µg/ml uracil in the case of the *pyrE* mutant. The WT culture was set up using CBM broth without uracil supplement. Samples were taken from each culture at 0, 12, 24, 48, 72, 96 and 120 hours. Gas chromatography was used to measure acetate, butyrate, acetone, butanol, and ethanol concentration for (A) WT, (B) *pyrE* mutant isolate A, (C) *pyrE* mutant isolate D, and (D) *pyrE* mutant isolate G. The values plotted are the means with standard error of the mean (SEM) from technical triplicates.

The accumulation of the storage polysaccharide granulose (starch) was analysed in the WT and *pyrE* mutant isolates (as described in **section 2.6.3**). **Figure 4.8** shows the *pyrE* mutants all had the same degree of granulose accumulation as the WT control, indicated by the black coloration.

The number of heat resistant spores was calculated for the three *pyrE* mutants and the WT control. After 120 hours, 100 µl of culture was heat shocked at 80°C for 10 minutes. This was then used to make a 1:10 serial dilution series and each was plated onto CGM agar. After 24 hours the CFU/ ml was calculated. **Figure 4.9** shows that after 120 hours the sporulation profiles for the *pyrE* mutant isolates tested were highly similar to that of the WT, with around 1 $X10^8$ CFU/ ml calculated for all.

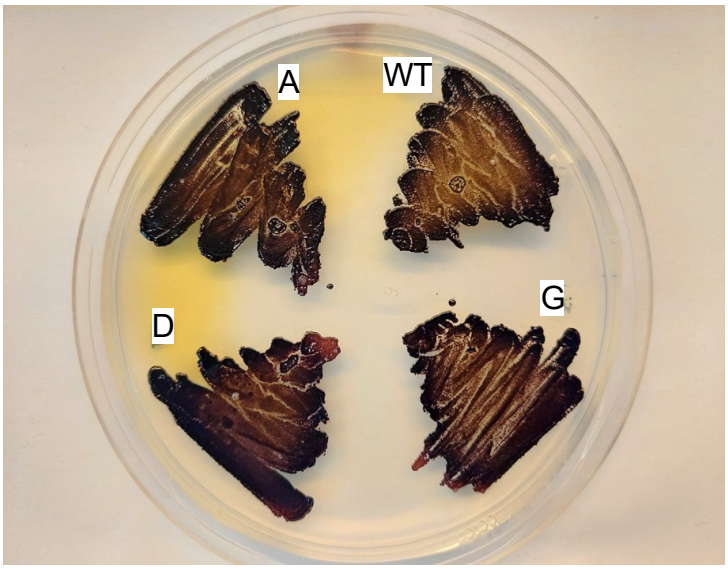


Figure 4.8 Granulose accumulation for three independent *C. beijerinckii* NCIMB 8052 *pyrE* mutant isolates and the WT control.

The three *pyrE* mutant isolates, indicated by A, D and G, and the WT control were streaked and allowed to grow for 48 hours. The accumulation of granulose was observed by dropping Lugols iodine solution onto the surface of the growth. The presence of granulose was confirmed by the black coloration observed.

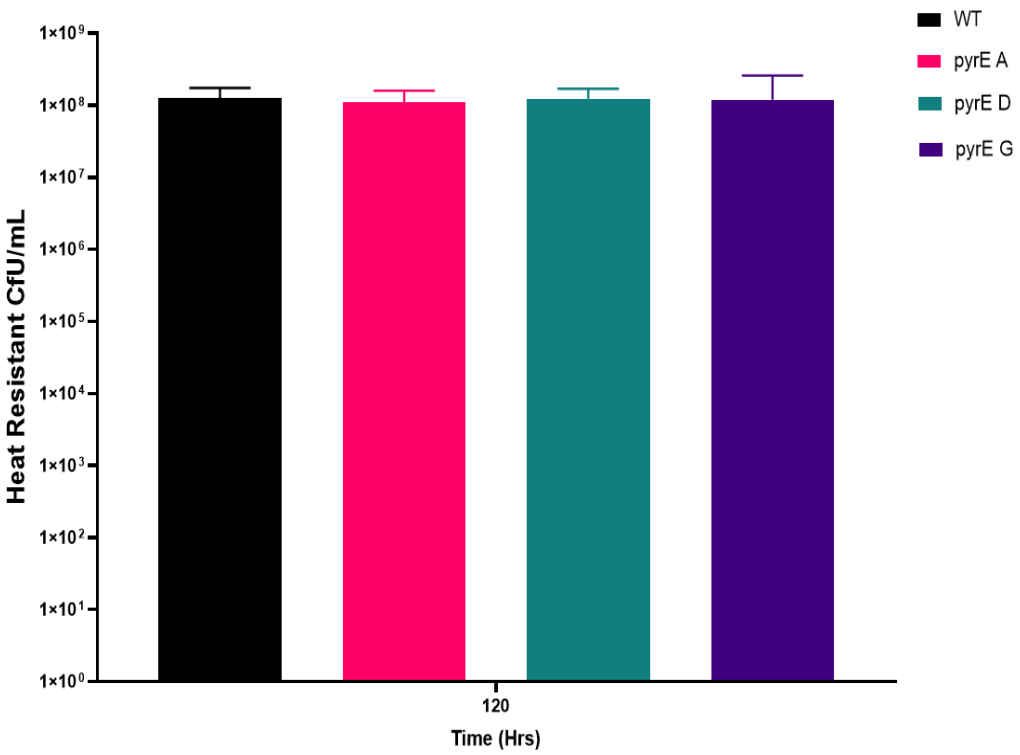


Figure 4.9 The calculated heat resistant CFU/ml for the three C*. beijerinckii* NCIMB 8052 *pyrE* mutants and WT.

Heat resistant CFU/mL were calculated after 120 hours of growth in CBM broth (6% glucose). 100 µl of culture was heat shocked at 80 °C and used to make a 1:10 serial dilution which were subsequently plated onto CGM (+ 20 µg/ml uracil). The CFU/ ml was calculated after 24 hours of growth, values plotted are the means with standard error of the mean (SEM) from technical triplicates.

Based on the undertaken phenotypic characterisation, all the Δ*pyrE* mutant isolates tested showed to behave in a similar manner to that of the *C. beijerinckii* NCIMB 8052 WT previously characterised. The isolate denoted as *pyrE* A was carried forward to use as a master stock for all future experiments. The genomic DNA was extracted from this Δ*pyrE* mutant using phenol chloroform-based extraction and sent for whole genome sequencing.

The vectors pMTL-JRH3_agrD2 KO and pMTL-JRH3_agrD5 KO were transformed into the Δ*pyrE* mutant strain and plasmid uptake selected for by plating onto CGM + 10 µg/ml erythromycin. As previously described (**section 4.2.4**), the largest colonies were restreaked twice onto the same media to encourage a single cross-over event with one of the vector-based homology arms. However, single cross-over isolates were not observed, suggesting that allelic exchange continued to be inefficient.

To increase the selective pressure of a single cross-over event, suicide vectors of these KO constructs were made by digesting the relevant vector with AscI and PflFI restriction enzymes to remove the Gram-positive pCB102 replicon. This digest product was then treated with T4 DNA polymerase following the manufacturer's instructions, to create blunt ends. The blunt end digest product was then treated with T4 DNA ligase and transformed into TOP 10 *E. coli* for storage. Plasmid transformation efficiencies in *C. beijerinckii* for pMTL 83251 and pMTL-JRH3_agrD5 KO were calculated to reach 1 x 10^5 CFU/ml per µg of DNA. However, even with this added selection pressure, no single

crossover isolates were detected. It was therefore decided to use a CRISPR based method which had been recently developed for use in *Clostridia* by another SBRC Nottingham researcher, Dr Francois Seys (unpublished).

4.2.5 Using CRISPR as an in-frame disruption for gene knockout

The CRISPR method used was based on that described by Nishida *et al.* (2016) and Dr Seys. This technique relies on a type II Cas9 nickase enzyme and an activation-induced cytidine deaminase (PmCDA1). This allows targeted single-nucleotide substitution at cytidines, within an editing window between -16 to -19 bp from the NG 3' PAM, with a thymine. This system has benefits over WT Cas9 enzymes as the toxicity of the nickase Cas9 (nCas9 D10a) is greatly reduced.

The CRISPR nCas9 plasmids, vFS77, contains an NG 3' PAM guided *Petromyzon marinus* cytidine deaminase 1 (PmCDA1) fused to a Cas9 nickase enzyme which is under transcriptional control by the *pfdxE* inducible promoter, activated by theophylline. The sgRNA for the target gene is under constitutive expression via the P_{araE} promoter.

To make the plasmid suitable for use in *C. beijerinckii* NCIMB 8052 its *catP* antibiotic resistance marker first had to be substituted to *ermB.* The construction of the CRISPR KO vectors, hereafter referred to as vLRB1 vectors, are fully described in **section 2.3.3.3**. The restriction enzyme digest of vFS77 with FseI and HpaI was visualised on an

agarose gel via electrophoresis to create two DNA fragments of 9.20 Kb and 0.71 Kb (**Figure 4.10**)**.** The 9.20 Kb band contains the linearised vFS77 backbone without the *catP* marker. The sgRNA protospacer could then be substituted in vLRB1 to target specific *agr* genes and thus creating the vLRB1-agr target vectors vLRB1-agrB2, vLRB1-agrB4 and vLRB1- agrB5. In this instance it was decided to target *agrB* genes as all the relevant Agr systems had a suitable editing window (a cytidine between -16 to -19 bp from the NG 3' PAM) that could be targeted by this CRISPR at the start of the gene, allowing gene function to be knocked out. In the case of the *agrD* genes, it was not possible to find an editing window for this CRISPR tool in the *agrD2* gene, and it was difficult to target a site in the *agrD* genes which would ensure the gene would become non-functioning.

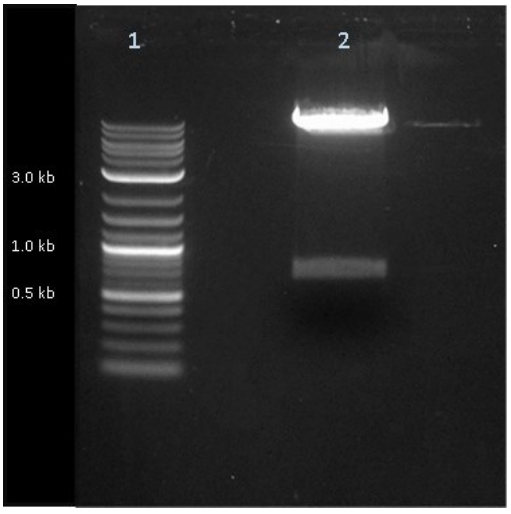


Figure 4.10 FseI and HpaI restriction enzyme digest of vFS77.

Expected band sizes were 0.71 kb and 9.20 kb (lane 2). The size of the PCR products was estimated using a 1 % agarose gel, run for 40 minutes at 100 V, 400 mA and using a 2-log 1kb DNA ladder (lane 1).

The vLRB1-*agr* target vectors were transformed into *C. beijerinckii* NCIMB 8052 Δ*pyrE*. Transformants were selected for by plating onto CGM + 10 µg/ml erythromycin + 20 µg/ml uracil and left to grow overnight. The resulting colonies were patch plated onto CGM + 10 µg/ml erythromycin + 20 µg/ml uracil + 5 mM theophylline two times to induce CRISPR Cas9 activity. Colonies which grew were screened for the target mutation through PCR amplification followed by Sanger sequencing using the primer pairs agr KO confirmation_F + agr KO confirmation_R for the relevant *agr* gene. Those which were confirmed were then subjected to plasmid loss selection by streaking onto non-selective media multiple times. Those which had lost the plasmid backbone were confirmed when they could grow on CGM + 20 µg/ml uracil but not on CGM + 10 µg/ml erythromycin + 20 µg/ml uracil. Point mutations were confirmed in genes *agrB5* (CBEI_RS20300/Cbei_3965), *agrD5* (CBEI_RS26925), *agrB4* (CBEI_RS16315/Cbei_3171), *agrD4* (CBEI_RS26650) and *agrB2* (CBEI_RS03605/Cbei_0658) (**Figure 4.11**). Double point mutations were also made for the combinations *agrB2agrB4*, *agrB2agrB5* and *agrB4agrB5*.

Mutants were sent for sanger sequencing to ensure the other *agr* loci within the same genome were indeed intact, and that off-target mutation had not affected alternate loci. The *agrB* mutants were also routinely screened to be the correct strain by using the primers 8052 LAC_F + 8052 LAC_R, which amplified a specific DNA methylase gene (Cbei_1632) only present in the *C. beijerinckii* NCIMB 8052 strain.

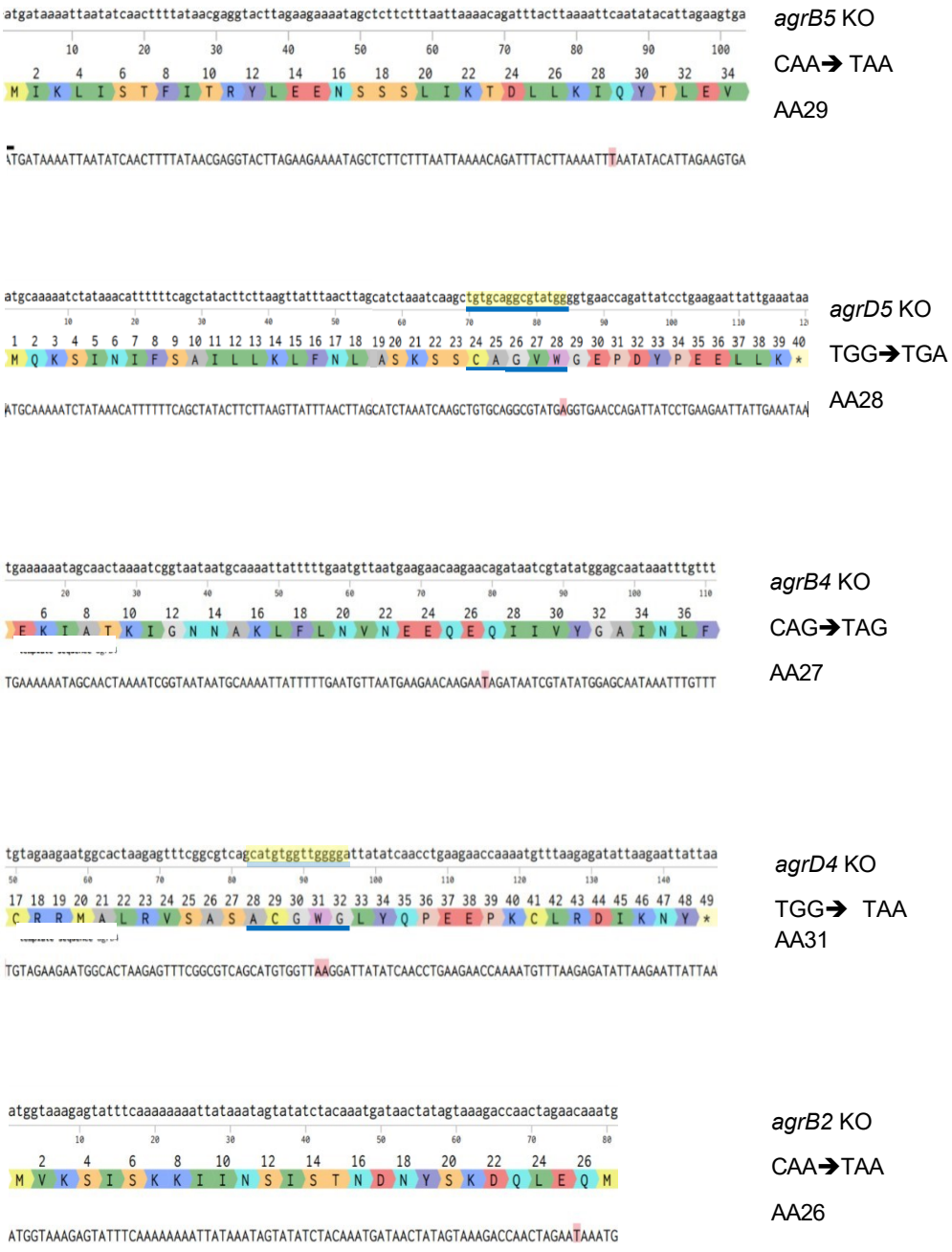


Figure 4.11 A section of DNA sequence alignment for individual agr gene mutations in *C. beijerinckii* NCIMB 8052 in comparison to the corresponding WT gene on each top row.

Mutations were confirmed in genes *agrB5* (CBEI_RS20300/Cbei_3965), agrD5 (CBEI_RS26925), *agrB4* (CBEI_RS16315/Cbei_3171), *agrD4* (CBEI_RS26650) and *agrB2* (CBEI_RS03605/Cbei_0658). Only relevant sections of these genes flanking the mutated bases are shown. Yellow highlight/blue underline indicates the predicted AIP sequences (encoding DNA and protein) in the WT, red highlight indicates the mutated base in the mutant. Codon changes and number of the amino acid (AA) affected are indicated on the right.

4.2.5.1 Δ*pyrE* repair and complementation

The obtained *agr* mutants with a truncated *pyrE* gene were transformed with the JRH4 *pyrE* repair vector and the corresponding JRH4-agr complementation vector respectively. The pMTL-JRH4 *pyrE* repair vector contain the intact *C. beijerirnckii* NCIMB 8052 *pyrE* gene and 1200 bp of its downstream region. Allelic exchange using this vector would restore the WT *pyrE* sequence and hence allow future comparison of the mutants' phenotypes with the unmutated parent strain. The *agr* complementation vectors contained the same as the pMTL-JRH4 *pyrE* repair vector and in addition also contained the intact target *agr* gene and its 200 bp upstream region which were inserted 70 bp downstream of the intact *pyrE* gene (see appendix **Figure 8.7**). Use of the latter vectors would restore *pyrE* function whilst at the same time integrating an intact copy of the respective *agr* gene immediately downstream. The resulting strain would therefore represent a genetically complemented mutant, carrying a single copy of the previously mutated *agr* gene on the chromosome, the expression of which would be driven by the native *agr* promoter for each gene.

Transformants were selected for by allowing to grow overnight on CGM+ 20 µg/ml uracil + 750 µg/ml spectinomycin. The largest colonies were then streaked directly onto CBM agar with no supplementation and left to grow overnight. This was repeated before patch plating onto CBM and CBM + 750 µg/ml spectinomycin to check for plasmid loss. Those colonies where plasmid loss was confirmed were checked for *pyrE*

gene repair and genetic complementation at the *pyrE* locus by using the primer pair *pyrE* KO confirmation_F + *pyrE* KO confirmation_R and Sanger sequencing of the resulting PCR fragment.

4.3 Discussion

As described in **chapter 3**, the genome sequences of 35 different *C. beijerinckii* strains were analysed for putative *agr* gene clusters. In this study, the aim was to create in-frame deletions in the most commonly occurring *agr* systems. Three different genetic manipulation techniques were used to attempt this, allelic exchange, ClosTron and CRISPR-Cas9 based systems (Kuehne *et al.*, 2011; Al-Hinai, Fast and Papoutsakis, 2012; Ehsaan *et al.*, 2016; Nishida *et al.*, 2016; Wang *et al.*, 2016; Little *et al.*, 2018). There have been issues with the application of genetic manipulation tools in *Clostridium* species, factors including them being Gram-positive bacteria with thick cell walls, making it difficult for cells to uptake foreign DNA (Pyne et al., 2014).

4.3.1 The use of allelic-exchange

The two most commonly conserved groups, denoted as Agr2 and Agr5, were initially targeted for *agrD* gene KO using allelic exchange (Ehsaan *et al.*, 2016; Ehsaan, Kuehne and Minton, 2016). Due to the lower efficiency DNA transfer in many *Clostridium* species, the use of suicide vectors is limiting and therefore 'pseudo-suicide' vectors have been adopted in this method (Kuan Ng *et al.*, 2013). These pseudo-suicide vectors are shuttle vectors carrying an antibiotic resistant marker and two replication origins, one for *E. coli* and one for the target *Clostridium,* which is replication deficient, meaning it has a reduced replication efficiency. This ensures a growth disadvantage in the presence of the antibiotic marker compared to cells in which the

plasmid has integrated. Therefore, cells in which the plasmid becomes integrated via homologous recombination grow larger compared to cells harboring a non-integrated plasmid (Ehsaan *et al.*, 2016; Ehsaan, Kuehne and Minton, 2016). In this study, the pseudo suicide vectors containing a Gram-positive replication origin of pCB102 were pMTL- JRH3_agrD2 KO and pMTL-JRH3_agrD5 KO, which were constructed to target *agrD2* and Cbei_RS26925 *agrD5* respectively.

Allelic exchange has been successfully used in both pathogenic and solventogenic *Clostridium* species, Ehsaan *et al.* (2016), demonstrated the use of both the *pyrE*-based and *codA*-based pseudo suicide systems to isolate in-frame deletion mutants in the solventogenic *C. acetobutylicum* ATCC 824. Another study by Foulquier *et al.* (2019) used *upp*-based suicide vectors to both delete and edit genes in *C. acetobutylicum* and *C. saccharobutylicum* using allelic exchange. Whilst Little *et al.* (2018) used allelic exchange to produce targeted point mutation in the genome of *C. beijerinckii* NCIMB 14988. However, the *pyrE*-based system did not show to be efficient in *C. beijerinckii* NCIMB 8052 for producing the target gene mutations in this study. It was demonstrated that allelic exchange could be used in this strain to produce the *pyrE* KO strain (**section 4.2.4**). This was likely due to the high selection pressure applied in this process, this is through the use of 5-FOA, which is toxic to cells with an intact *pyrE* gene, and the dependence of the *pyrE* mutant on uracil supplementation (Kuan Ng *et al.*, 2013). In this process as described in **sections 4.2.1** and **4.2.5**, no isolate colonies with an integrated plasmid at the target *agr* gene sites were detected, suggesting homologous recombination in this strain was

not efficient and requires strong selection pressure for enrichment and detection, therefore screening for desired mutations became time-consuming.

4.3.2 The ClosTron system

The ClosTron system involves the use of a targeted group II intron and a retro-transposition-activated marker to insert DNA into a specific site in the target genome resulting in gene inactivation (Heap *et al.*, 2007). The ClosTron is credited for being highly efficient and rapid in *Clostridium* species, being widespread in the research community its use has been demonstrated in industrially important species including *C. acetobutylicum*, in which Steiner *et al.* (2012) used ClosTron technology to create *agrB*, *agrC* and *agrA* mutants. Further use has been in *C. beijerinckii* NCIMB 14988 (Little *et al.*, 2018) and NCIMB 8052 (Humphreys, 2019), as well as pathogenic species such as *Clostridium difficile* (Barketi-Klai *et al.*, 2011; Dingle, Mulvey and Armstrong, 2011). However, in this study the use of ClosTron technology did not prove to be highly efficient with only a single colony isolate showing successful intron integration. Further processing of this isolate demonstrated degeneration had occurred when subculturing was carried out during genetic manipulation. It has been noted previously that the multiple restreaking of suspect colonies during ClosTron mutagenesis increases the risk of spontaneous degeneration (Steiner et al., 2012). The presumably degenerate isolate of the *agrD5* ClosTron mutant isolated in this study was characterised as having

statistically reduced sporulation and a loss of granulose production compared to the WT, phenotypes associated with degeneration in this strain (Humphreys, 2019), although it showed slightly increased solvent production. The granulose production could not be chemically complemented when the isolate was streaked within proximity to the WT, where the WT AIP would be expected to diffuse to the mutant to complement the phenotype, suggesting the *agrD5* mutation was not responsible for this phenotype (see appendix **Figure 8.5**).

Whilst ClosTron technology is a popular choice to rapidly create target mutations, the issue of polar effects due to the intron insertion is undesirable and other techniques, such as CRISPR technology are now increasingly popular and easier to use in a range of bacterial species (Peters *et al.*, 2015; Banno *et al.*, 2018; Shanmugam, Ngo and Wu, 2020).

4.3.3 CRISPR

The CRISPR-Cas9 system has previously been developed for use as a genome editing tool in *Clostridium* species (Xu *et al.*, 2015; Li *et al.*, 2016; Diallo, Kint, *et al.*, 2020; Feng *et al.*, 2020). Due to the toxicity of the native Cas9 nuclease, which catalyses a double-stranded DNA break, a mutant version called Cas9-nickase is often used, which produces a single-stranded nick and more easily repaired in prokaryotes. This variant of Cas9 was first applied to *Clostridium* in a study by Xu *et al.* (2015) on *Clostridium cellulolyticum* and has since been adopted for use in solventogenic clostridia such as *C.*

acetobutylicum ATCC 824 and *C. beijerinckii* NCIMB 8052 (Li *et al.*, 2016; Diallo, Hocq, *et al.*, 2020a).

The application of CRISPR-Cas9 requires DNA cleavage followed by homologous recombination. The apparent reduced efficiency of homologous recombination demonstrated in this strain of *C. beijerinckii* suggested the application of the general CRISPR-Cas9 systems may be difficult. Therefore, a more recent tool based on single nucleotide editing was used. In this study the CRISPR-nCas9-AID system used to successfully produce targeted mutations was described by Nishida *et al.* (2016) and Banno *et al.* (2018). Here an activation-induced cytidine deaminase (AID) is fused to the nCas9 nuclease which modifies the deoxycytidine of the variable region of the immunoglobulin locus and results in cytosine-to-thymine substitution. The system therefore avoids double-stranded breaks and repair via homologous recombination using donor DNA. This approach was further developed by Dr Seys (SBRC Nottingham, unpublished) for use in *Clostridium* species and recently published for its use in *C. autoethanogenum* (Seys *et al.*, 2023). In this study the system proved to be highly efficient and extremely rapid in producing point mutations which resulted in an early stop codon in the target *agr* genes *agrB5* (CBEI_RS20300), *agrD5* (CBEI_RS26925), *agrB4* (CBEI_RS16315), *agrD4* (CBEI_RS26650) and *agrB2* (CBEI_RS03605). The *agrB* genes could easily be complemented by inserting a copy of the relevant gene and its native promoter into the *pyrE* locus of the genome, again facilitated by

the strong selective pressure associated with restoration of uracil prototrophy.

Chapter 5:

Phenotypic characterisation of *agr* mutants

5.1 Introduction

As previously discussed, *C. beijerinckii* is an anaerobic, endospore forming bacterium which gains energy through the fermentative process of converting sugars into acids such as butyrate and acetate. These are then further converted into solvents including acetone, butanol, and ethanol. Previous studies have demonstrated a transcriptional link between the sporulation regulator, Spo0A, and solvent production in solventogenic clostridia (Harris et al., 2002; Ravagnani et al., 2000). There is also evidence that QS systems hold a regulatory role in solventogenic clostridia, particularly in the shift from acidogenesis to solventogenesis and sporulation. Various studies on *C. autoethanogenum*, *C. saccharoperbutylacetonicum, C. acetobutylicum* ATCC 824*, C. beijerinckii* NRRL B-598 and *C. beijerinckii* NCIMB 8052 have demonstrated the presence of Agr systems and in some cases, such as *C. autoethanogenum*, *C. acetobutylicum* and *C. saccharoperbutylacetonicum,* their involvement in these processes (Kolek et al., 2017; Kotte et al., 2020; Kosaka *et al.*, 2007; Piatek et al., 2022; Steiner et al., 2012). However, whilst Agr systems have been identified in *C. beijerinckii* NCIMB 8052, no studies have conducted in-depth experimental analysis to investigate their regulatory roles and a potential direct or indirect role in the regulation of solvent metabolism and initiation of sporulation. Understanding the regulatory elements of

processes affecting solvent formation and sporulation, in *C. beijerinckii* is important as it can better our understanding of this industrially relevant organism. The industrial desire of increasing the efficiency of solvent formation in this organism would greatly benefit from the fundamental understanding of how QS, which will occur at the high cell volumes in batch cultures, can be manipulated to make for a more successful fermentation process. There are also further applications of *Clostridium* species which go beyond solvent production, such as the potential use as probiotics or for cancer therapies. For instance, the inactive spores of *C. Beijerinckii* and other saccharolytic clostridia may be used as gene delivery vectors for cancer treatment (Fox *et al.*, 1996; Lemmon *et al.*, 1997). The overarching aims of this chapter were therefore to:

- Phenotypically characterise the *agrB KO* mutants created (chapter 4) using CRISPR-cas9 with respect to growth characteristics, granulose accumulation, sporulation, and solvent formation.
- Identify *agrB* mutants which show significant phenotypic differences compared to the WT for RNA sequencing (chapter 6).

5.2 Results

5.2.1 Phenotypic characterisation of single *agrB* KO mutants

5.2.1.1 Growth

Each *C. beijerinckii* NCIMB 8052 *agrB* mutant and its genetic complement were grown in culture for 5 days to produce spore stocks. From these spore stocks, vegetative cultures were obtained for phenotypic characterisation, as described in **Section 2.6**, which focused on growth characteristics, sporulation, and solvent analysis across several time points.

Initially two independent isolates were characterised in triplicate for each mutant and complement strain. The isolate samples showing consistent behavior, with regards to granulose accumulation and colony morphology, were carried forward and from then on only one mutant isolate and its complement derivative were continued for in- depth analysis. This was done due to the higher likelihood of species degeneration and therefore reduced the chance of isolating a degenerate mutant. For the isolates carried forward into further experiments, a summary of the growth curve analysis is represented by

Figure 5.1A. The mean growth rate was calculated for each sample, between 2-10 hours, using the formula described in **section 2.6.2**. The WT mean growth rate was calculated as 0.40 h^{-1}, like that measured in **section 4.2.2.1**. Similarly, the *agrB2*, *agrB4* and *agrB5* mutants had average growth rates of 0.46 h^{-1}, 0.46 h^{-1} and 0.43 h^{-1} respectively. Meanwhile, the *agrB2*, *agrB4* and *agrB5* complement strains had the respective calculated growth rates of 0.40 h^{-1}, 0.34 h^{-1} and 0.43 h^{-1}.

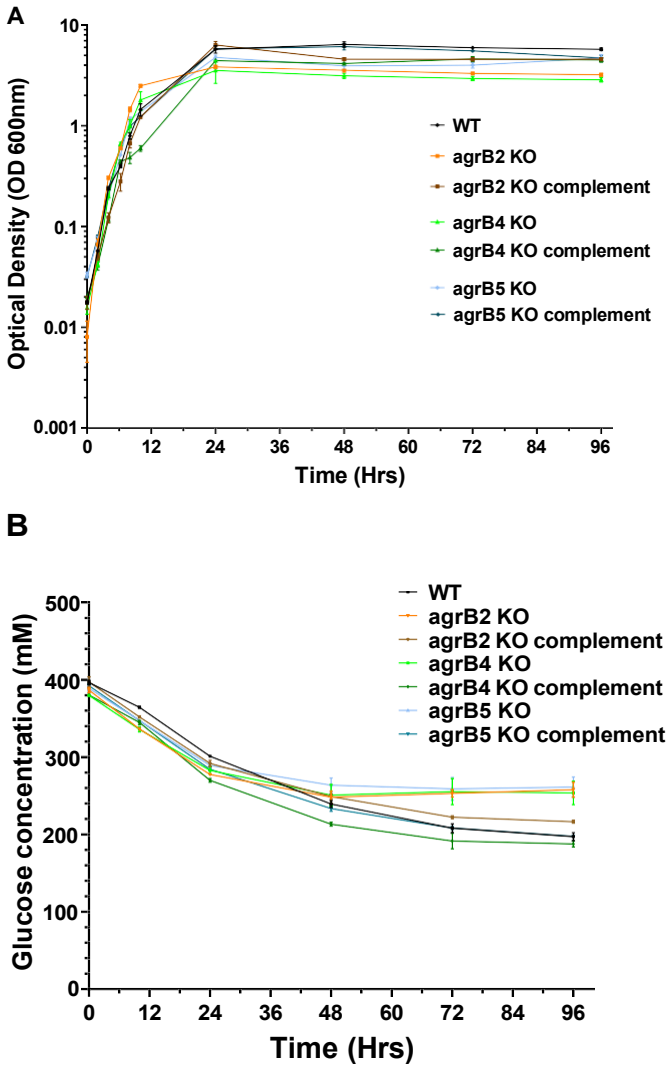


Figure 5.1 Growth curve and glucose consumption for three different *C. beijerinckii* NCIMB *agrB* single mutants with their genetic complements and comparative WT.

The growth curves for the *C. beijerinckii* NCIMB 8052 *agrB2 (*CBEI_RS03605/Cbei_0658*), agrB4* (CBEI_RS16315/Cbei_3171) and *agrB5* (CBEI_RS20300/Cbei_3965) mutants alongside their complemented strains and the WT, are shown by graph A. High performance liquid chromatography was used to measure glucose concentration (graph B). The values plotted are the means with standard error of the mean (SEM) from 3 technical replicates.

Whilst the growth rates of each *agrB* mutant was calculated to be very similar to that of the WT comparative control, there did appear to be a consistent difference in the OD_{600} at which each strain plateaued, after 24 hours. All *agrB* mutants tended to reach a final OD_{600} lower than that of the WT, in this case *agrB2* reached 3.8, *agrB4* reached 3.5 and *agrB5* reached 4.8, all comparatively lower to the WT which reached an OD_{600} of 5.8 after 24 hours. These final OD_{600} values were seen to be higher in each of the complemented strains.

Figure 5.1B demonstrates the glucose consumption for each strain under investigation. The *agrB* mutants all showed an overall lower glucose consumption between 0-96 hours of growth, when compared to the WT, with consumption slowing considerably after 24 hours. The *agrB2*, *agrB4* and *agrB5* consumed 128.7 mM, 126.0 mM and 127.7 mM respectively, whereas the WT consumed 199.1 mM of glucose, this was statistically significant ($P<0.05$) according to a T-test. This reduction in glucose consumption was genetically complemented in the *agrB2*, *agrB4* and *agrB5* complemented strains with these consuming 180.7 mM, 192.6 mM and 194.8 mM respectively. After 72 hours, neither WT nor mutants showed significant consumption of external glucose.

5.2.1.2 granulose assay

Granulose accumulation was analysed in all *agrB* mutants and their complemented strains alongside the WT comparison (**Figure 5.2**). In an initial granulose assay using all the mutant isolates generated through CRISPR, two of the *agrB* mutants were visibly different to the other clones and started to show characteristics of degeneration which were the typical changes in colony morphology and no granulose accumulation as described by Humphreys *et al*, (2023) and thus were excluded from further analysis. The *agrB* mutants, confirmed not to be degenerate strains, appeared to have a similar level of granulose accumulation to that of the WT, indicated by the dark coloration after the addition of Lugol's solution.

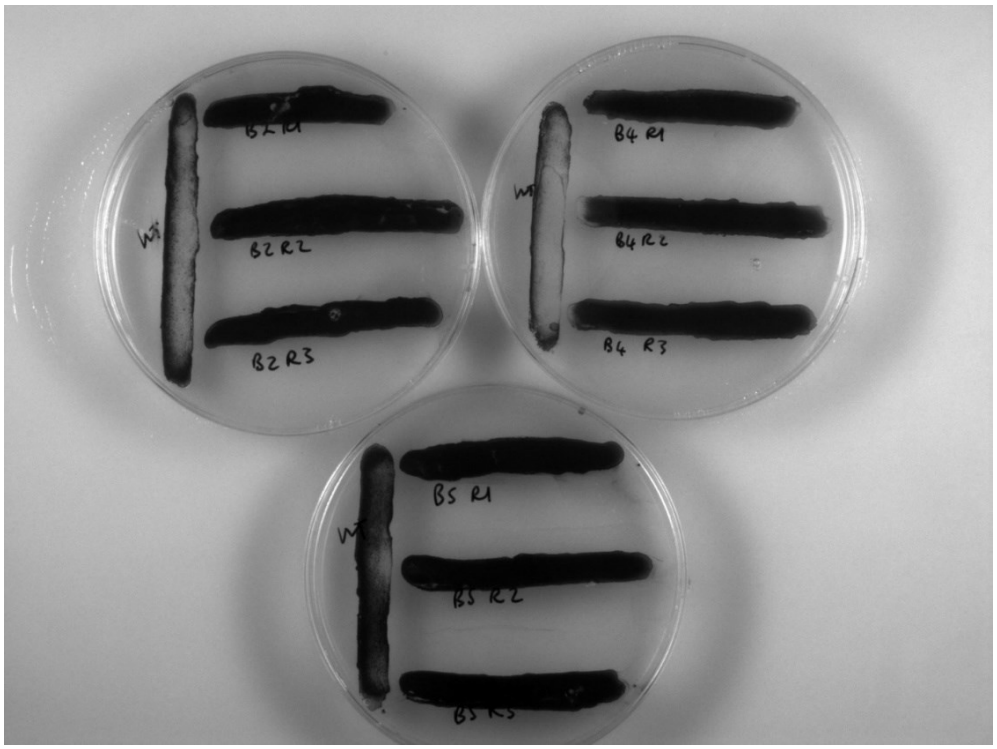


Figure 5.2 Granulose assay for the *C. beijerinckii* NCIMB 8052 *agrB2*, *agrB4* and *agrB5* mutants.

The WT and *agrB* mutants were heavily streaked onto CBM lacking $CaCO_3$ and allowed to grow for 48 hours. The accumulation of granulose was observed by dropping Lugols iodine solution onto the surface of the growth. The presence of granulose was confirmed by the black coloration observed.

5.2.1.3 Sporulation

The number of heat resistant spores across the period of growth for each *agrB* mutant and complement strain was determined as described in **section 2.6.4**. Sporulation was quantified at 0 hours, 24 hours, and 120 hours (**Figure 5.3**). The WT reached around 10^8 CFU/ml by 24 hours growth, in comparison the *agrB2* and *agrB4* mutants were observed to have a 5-log and 4-log reduction in heat resistant spore formation after 24 hours respectively. The *agrB5* mutant had significantly reduced spore formation after 24 hours, by 1-log, but not at 120 hours. These sporulation profiles could be genetically complemented to similar CFU/ml found in that of the WT comparison, demonstrating that this reduction was linked to the respective *agrB* gene inactivation.

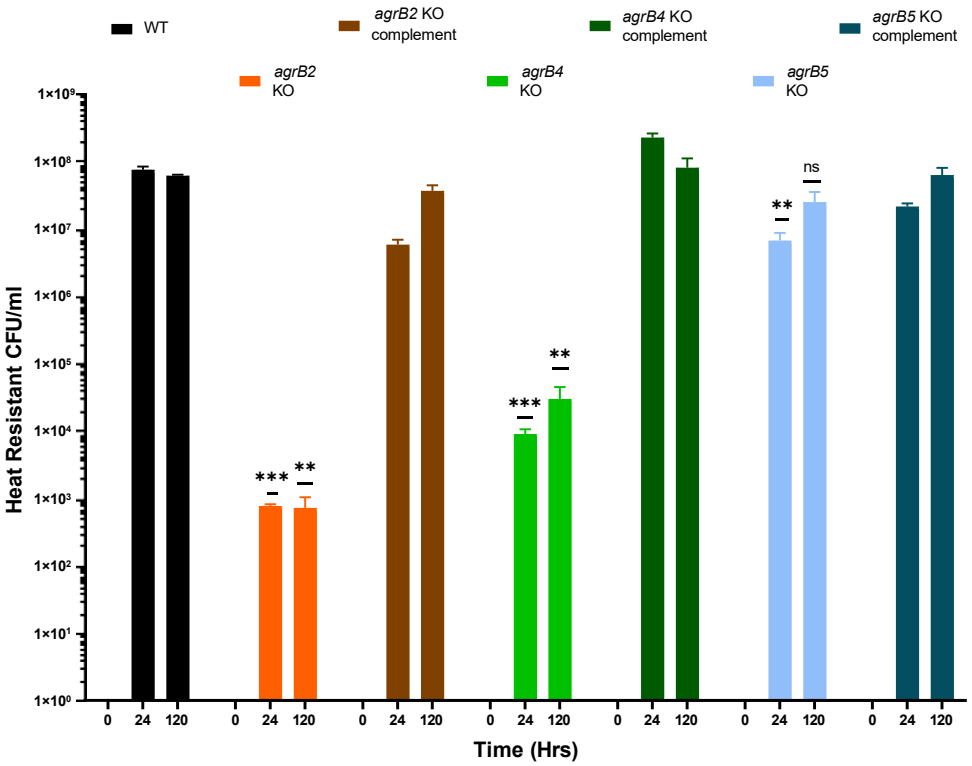


Figure 5.3 Heat resistant endospore formation in the *C. beijerinckii* NCIMB 8052 WT, *agrB* mutants and associated complement strains.

Heat resistant CFU/ml were calculated after 0, 24 and 120 hours of growth in CBM broth (6% glucose). Heat-resistant CFU were assumed to represent viable endospores. The values plotted are the means with standard error of the mean (SEM) from a minimum of 3 technical replicates. The statistical difference between the WT and *agrB* mutants was analysed using the T-test with P value ≤0.05. The degree of statistical significance is shown as <0.033 (*), <0.002(**), <0.001(***).

5.2.1.4 Solvent profiling

The *agrB* mutants and their complemented strains were tested for acid and solvent formation alongside the WT comparison using GC analysis for samples taken after 0, 10, 24, 48 and 72 hours (see **section 2.6).** Of particular interest was the difference in butanol production, shown in **Figure 5.4A.** Between 48-72 hours, where butanol concentrations tended to reach their highest, reaching 86 mM in the case of the WT, there was a 48-56 mM reduction in peak butanol formation observed in the mutants *agrB2*, *agrB4* and *agrB5* when compared to the WT. This reduction in butanol formation could partially be genetically complemented to WT concentrations, however not fully. The butyrate concentrations, shown in **Figure 5.4B** were significantly higher in the *agrB* mutants after 24 hours of growth compared to the WT concentration. In the WT strain, the butyrate concentration is shown to peak at 15.8 mM at 24 hours and then drop to 7.1 mM by 48 hours growth. In the *agrB* mutants this drop was not observed, and butyrate concentrations increased throughout, peaking at

49.2 mM, 34.6 mM and 34.5 mM in *agrB2*, *agrB4* and *agrB5* mutants respectively. The difference in butyrate was restored to levels similar to that of the WT in both the *agrB4* and *agrB5* complement strains and also partially in the *agrB2* complement. Ethanol production, shown in **Figure 5.5 C,** was observed to decrease in comparison to the WT in both the *agrB2* and *agrB5* mutants, however not in the *agrB4* mutant. This decrease in ethanol concentration was not genetically

complemented and so this difference cannot be subject to the *agrB* gene deletion in these strains. The acetone concentrations, visualised by **Figure 5.5B,** were also shown to have decreased in the *agrB* mutants. The acetone concentrations tended to reach their highest between 24-48 hours of growth. For the WT strain this peak concentration was a mean of 21.3 mM. In comparison, the *agrB2*, *agrB4* and *agrB5* mutants had peak values of 4.3 mM, 6.0 mM and 5.3 mM respectively, showing a statistically significant decrease between 15-17 mM. This reduction could be genetically complemented to higher concentrations similar to the WT in both the *agrB4* and *agrB5* complement strains however, for the *agrB2* complement strain this was only partial. The acetate concentration in the WT strain (**Figure 5.5A**) peaked at 24 hours around 20 mM and was maintained at this level. Very similar trends were seen for the *agrB2* and *agrB5* KO mutants. When the *agrB2* was genetically complemented, this showed an overproduction of acetate. The *agrB4* mutant did show a statistically significant increase in production of acetate after 24 hours in comparison to the WT with concentrations reaching 27.3 mM. This was shown to reduce to similar levels of the WT in the *agrB4* complement strain.

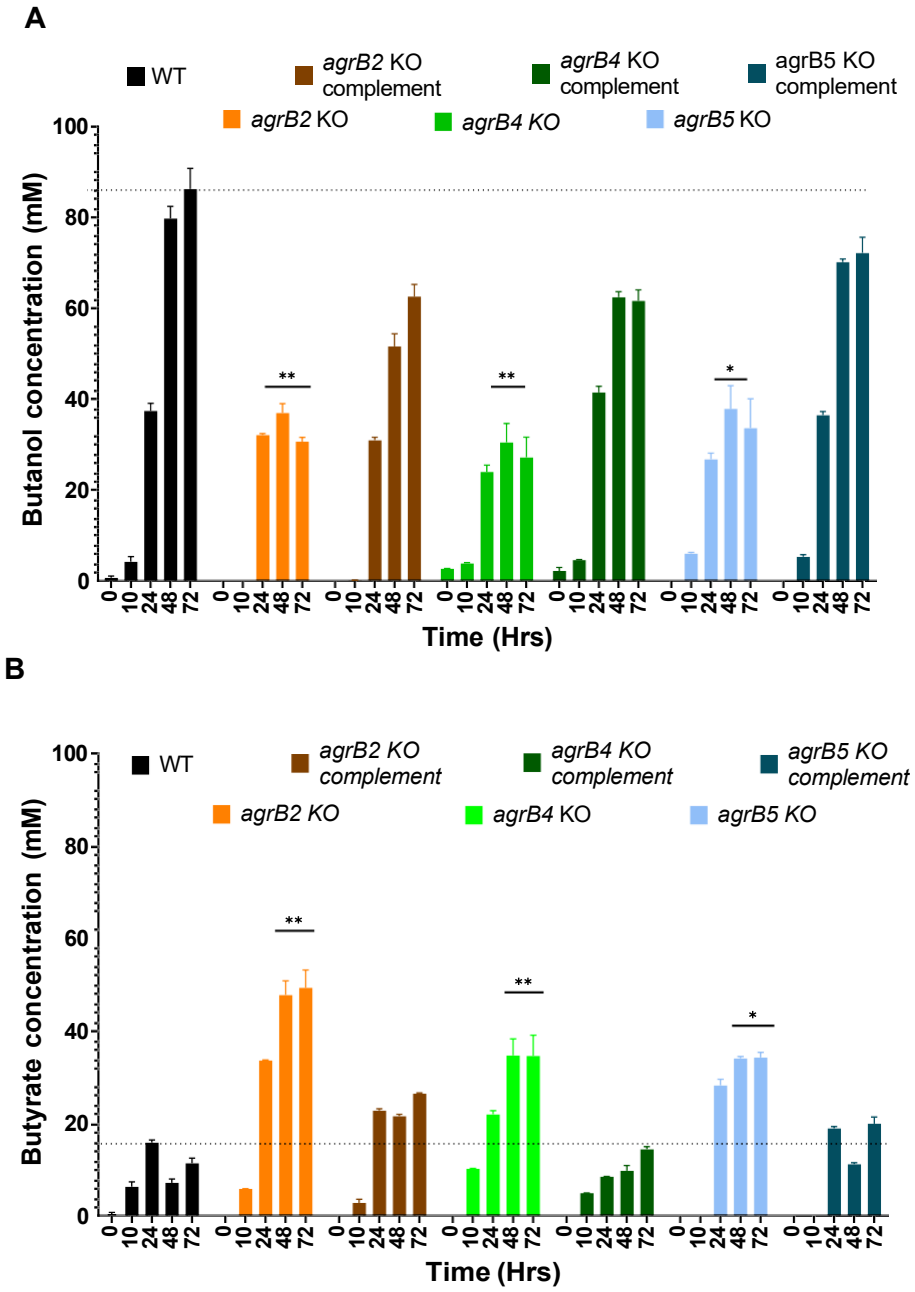


Figure 5.4 Butanol and butyrate analysis in the *C. beijerinckii* NCIMB 8052 *agrB* mutants and their complemented strains alongside the WT strain.

Gas chromatography was used to measure butanol (A) and butyrate (B) concentration for the *C. beijerinckii* NCIMB 8052 *agrB2* (CBEI_RS03605/Cbei_0658), *agrB4* (CBEI_RS16315/Cbei_3171) and *agrB5* (CBEI_RS20300/Cbei_3965) mutants alongside their complemented strains and the WT. The values plotted are the means with standard error of the mean (SEM) from a minimum of 3 technical replicates. Significant differences were calculated using the T-test with statistical significance compared to the WT shown as <0.033 (*), <0.002(**), <0.001(***). The horizontal black dotted line indicates the peak concentration reached in the WT.

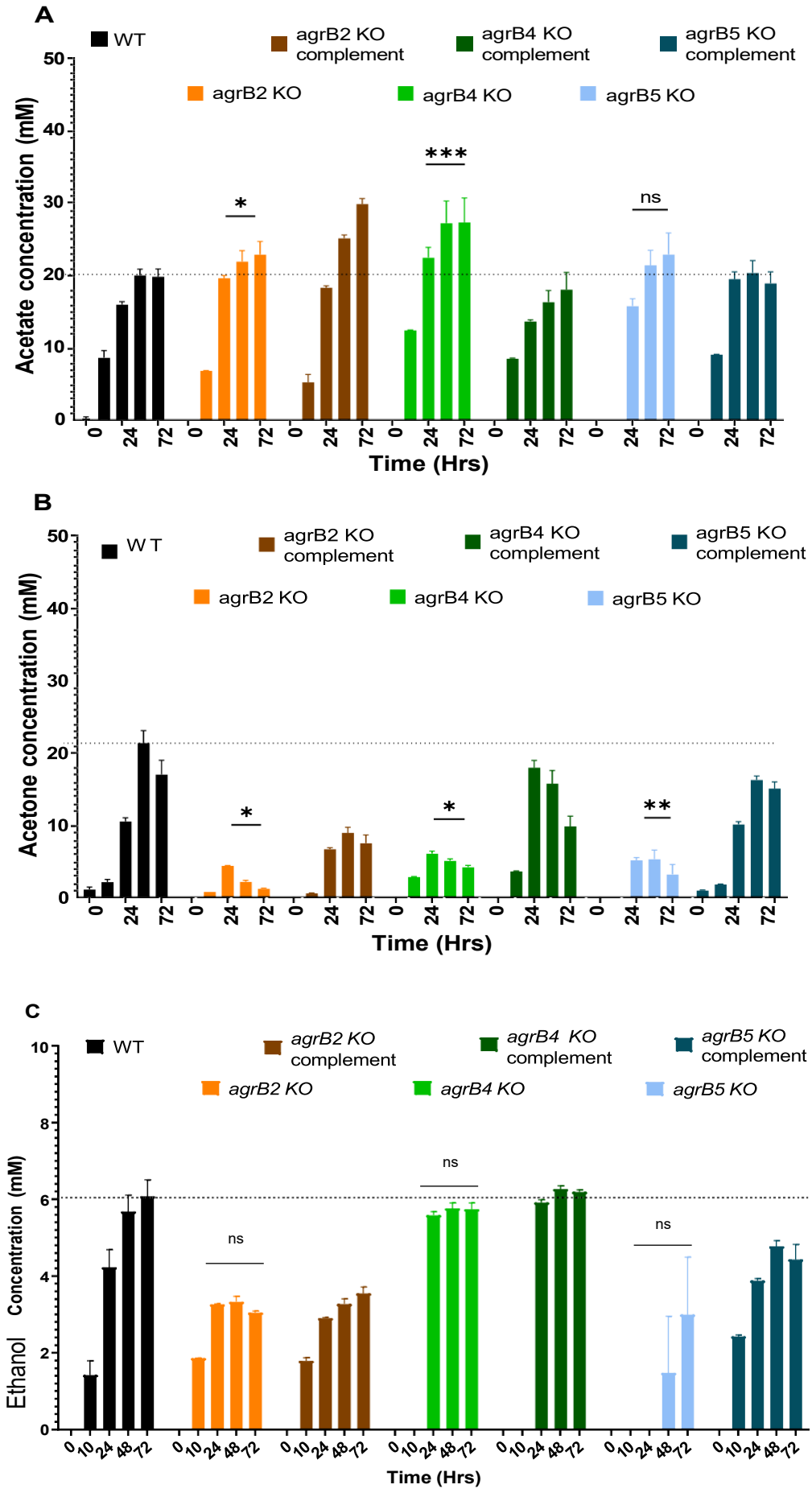

A
WT
agrB2 KO complement
agrB4 KO complement
agrB5 KO complement
agrB2 KO
agrB4 KO
agrB5 KO
Acetate concentration (mM)
Time (Hrs)
B
Acetone concentration (mM)
Time (Hrs)
C
Ethanol Concentration (mM)
Time (Hrs)

Figure 5.5 Acetate, acetone and ethanol analysis in the C. *beijerinckii* NCIMB 8052 *agrB* mutants and their complemented strains alongside the WT strain.

Gas chromatography was used to measure acetate (A), acetone (B), and ethanol (C) concentration for the *C. beijerinckii* NCIMB 8052 *agrB2* (CBEI_RS03605/Cbei_0658), *agrB4* (CBEI_RS16315/Cbei_3171) and *agrB5* (CBEI_RS20300/Cbei_3965) mutants alongside their complemented strains and the WT. The values plotted are the means with standard error of the mean (SEM) from a minimum of 3 technical replicates with statistical significance compared to the WT shown as <0.033 (*), <0.002(**), <0.001(***). The T-test was used to determine significant differences. The horizontal black dotted line indicates the peak concentration reached in the WT.

5.2.2 Double *agr* KO mutants

Due to the significant reduction of endospore formation and solvent formation observed in the single *agrB* mutants, it was of interest to observe if any *agrB* mutant combinations would result in further change. Combinations of the *C. beijerinckii* NCIMB 8052 Δ*agrB2*Δ*agrB4,* Δ*agrB2*Δ*agrB5* and Δ*agrB4*Δ*agrB5* mutants were generated.

The ΔagrB2ΔpyrE and ΔagrB4ΔpyrE mutants were first transformed with the LRB1-agrB2, LRB1-agrB4 and/or LRB1-agrB5 vectors respectively. These transformants were then processed as previously described to create a CRISPR-ncas9 based mutation in the agrB2agrB4, agrB2agrB5 and agrB4agrB5 mutants (refer to **section 2.5.4** and **Chapter 4**). These strains were first sent for sanger sequencing to check for the presence of both target mutations in the

genomes and then genetically complemented as per the other *agrB* mutant strains, using the *pyrE* locus to create chromosomal complementation. The complemented strains were also sent for sanger sequencing to confirm the genetic mutations were still present in the target *agrB* genes and that complementation had occurred at the *pyrE* locus.

The mean growth rate was calculated for each mutant and WT between 2-10 hours, using the formula described in **section 2.6.2**. The WT mean growth rate was 0.40h^{-1} on average between the different growth curve analyses. The double *agrB2B4*, *agrB2B5* and *agrB4B5* KO mutants had similar growth rates at 0.45 h^{-1}, 0.42 h^{-1} and 0.33 h^{-1} respectively (**Figures 5.6, 5.7 and 5.8**). By 24 hours of growth, all of the cultures appeared to plateau, having entered stationary phase. Like the single Δ*agrB* mutants, the double *agrB* mutants tended to reach a lower final OD600 nm when compared to the WT strain grown under the same conditions. The *agrB2B4* mutant for instance reached 2.92 OD_{600} whereas the WT under the same conditions reached 4.95 OD_{600} (**Figure 5.6A**). Glucose consumption was also observed to be reduced in the double KO mutants. The largest reduction being observed in the agrB2B4 double mutant, which consumed 79 mM less glucose compared to the WT. This was complemented to similar consumption levels to the WT when using *agrB2* or *agrB4* genes independently (**Figure 5.6B**).

Heat resistant endospore formation was measured as before. The *agrB2B4* mutant was observed to have the largest decrease in heat

resistant endospore formation, with a 7-log reduction. This was partially genetically complemented independently by either the *agrB2* gene or the *agrB4* gene at the *pyrE* locus, showing a slight increase in endospore formation in these isolates (**Figure 5.6C**). The acetone and butanol production was also observed to be reduce in this double KO mutant with the *agrB2B4* mutant butanol formation peaking at 20.8 mM by 48 hours compared to the WT butanol formation peaking at 72.0 mM at 72 hours growth, this was genetically complemented by the *agrB4* gene and partially by the *agrB2* gene (**Figure 5.6D**). The butyrate production in the *agrB2B4* mutant peaked at 52.3 mM by 72 hours, more so than in the respective single mutants compared to the WT (**Figure 5.6E**). However, for this double *agrB* mutant there did not appear to be a significant difference in acetate formation compared to the WT.

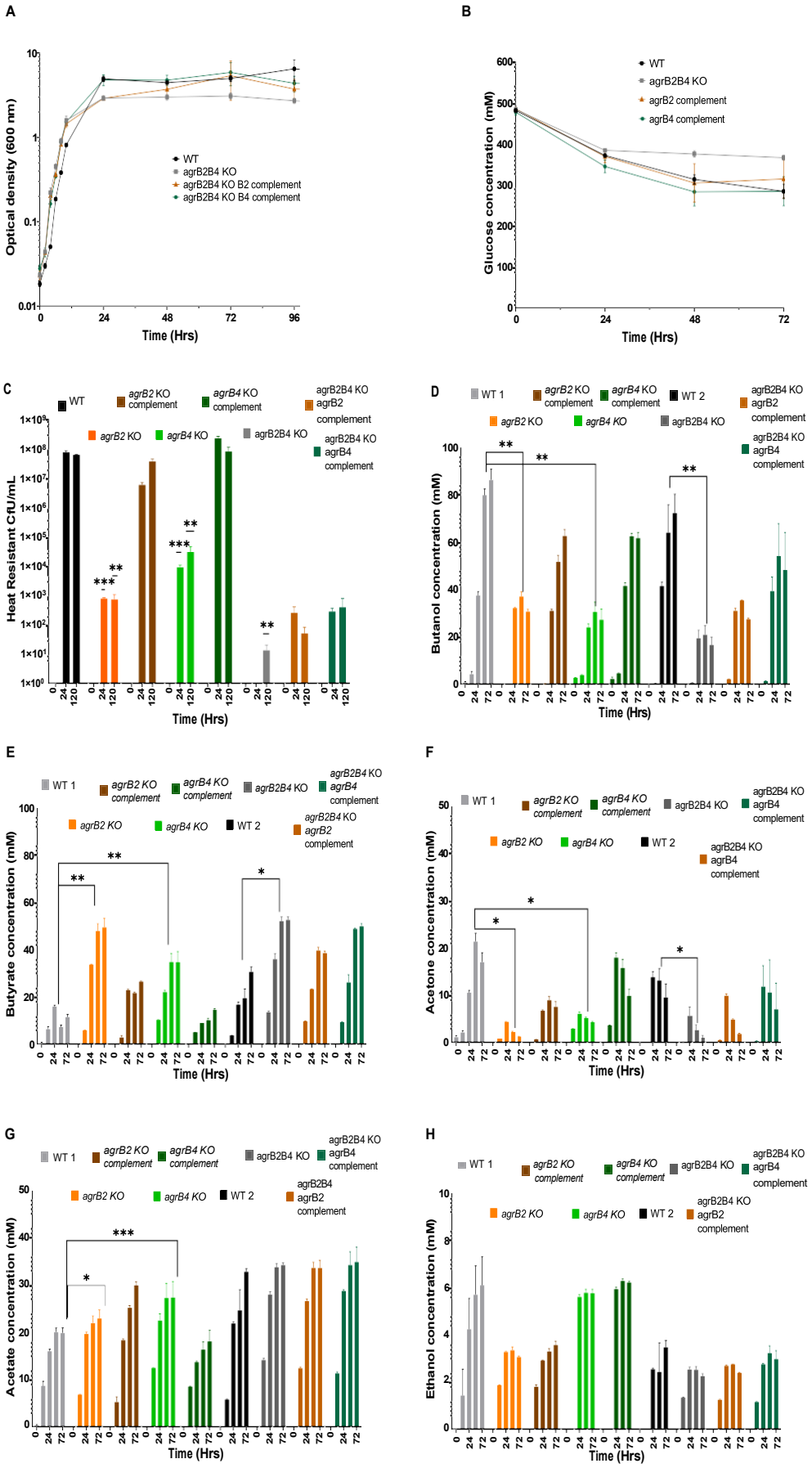

A
Optical density (600 nm)
Time (Hrs)
WT
agrB2B4 KO
agrB2B4 KO B2 complement
agrB2B4 KO B4 complement
B
Glucose concentration (mM)
Time (Hrs)
WT
agrB2B4 KO
agrB2 complement
agrB4 complement
C
Heat Resistant CfU/mL
Time (Hrs)
WT
agrB2 KO
agrB4 KO
agrB2B4 KO
agrB2B4 KO agrB2 complement
agrB2B4 KO agrB4 complement
D
Butanol concentration (mM)
Time (Hrs)
WT 1
WT 2
agrB2 KO
agrB4 KO
agrB2B4 KO
agrB2B4 KO agrB2 complement
agrB2B4 KO agrB4 complement
E
Butyrate concentration (mM)
Time (Hrs)
F
Acetone concentration (mM)
Time (Hrs)
G
Acetate concentration (mM)
Time (Hrs)
H
Ethanol concentration (mM)
Time (Hrs)

Figure 5.6 Growth curve, glucose consumption, endospore formation and solvent formation analysis for the *C. beijerinckii* NCIMB 8052 *agrB2B4* KO mutants.

The growth curve for each sample is shown by graph A. Values plotted are the mean with standard error of the mean (SEM). High performance liquid chromatography was used to measure glucose concentration (B) for the *C. beijerinckii* NCIMB 8052 *agrB2B4 (CBEI_RS03605/Cbei_0658, CBEI_RS16315/Cbei_3171)* mutant alongside the complemented strains and the WT. Heat resistant spore as CFU/ml (C) were calculated after 0, 24 and 120 hours of growth in CBM broth (6% glucose). Gas chromatography was used to measure butanol (D), butyrate (E), acetone (F), acetate (G), ethanol (H). The values plotted are the means with standard error of the mean (SEM) from a minimum of 3 technical replicates. The statistical difference between the WT and *agrB* mutants was analysed using the T-test with P value ≤0.05. The degree of statistical significance is shown <0.033 (*), <0.002(**), <0.001(***).

By 24 hours, the *agrB2B5* KO mutant growth plateaued at OD_{600} 3.79 compared to the WT isolate which remained at around OD_{600} 6.00 under the same conditions (**Figure 5.7 A**). Under these conditions, the *agrB2B5* KO mutant had a 4-log reduction in heat resistant endospore formation when compared to the WT at both 24 and 120 hours. This was genetically complemented, by the *agrB2* gene, to similar levels of the WT spore formation, however, was not genetically complemented by the *agrB5* gene (**Figure 5.7 C**). The peak butanol accumulation reduced by 55.2 mM, whilst the peak acetone accumulation reduced by 18.43 mM, in the *agrB2B5* KO mutant when compared to the WT strain grown in the same conditions, a larger reduction than that observed in the respective *agrB2* and *agrB5* single mutants **(Figure 5.7 D and F).** Accordingly, the butyrate and acetate peak production increased in the *agrB2B5* mutant by 43.7 mM and 20.2 mM respectively compared to the WT. These were partially complemented by the *agrB2* and *agrB5* genes independently (**Figure 5.7 E and G**). Whilst there was also a reduction in peak ethanol formation in this double mutant, there did not appear to be a reversal of this phenotype towards WT level in the complemented *agrB2B5* mutants (**Figure 5.7 H**).

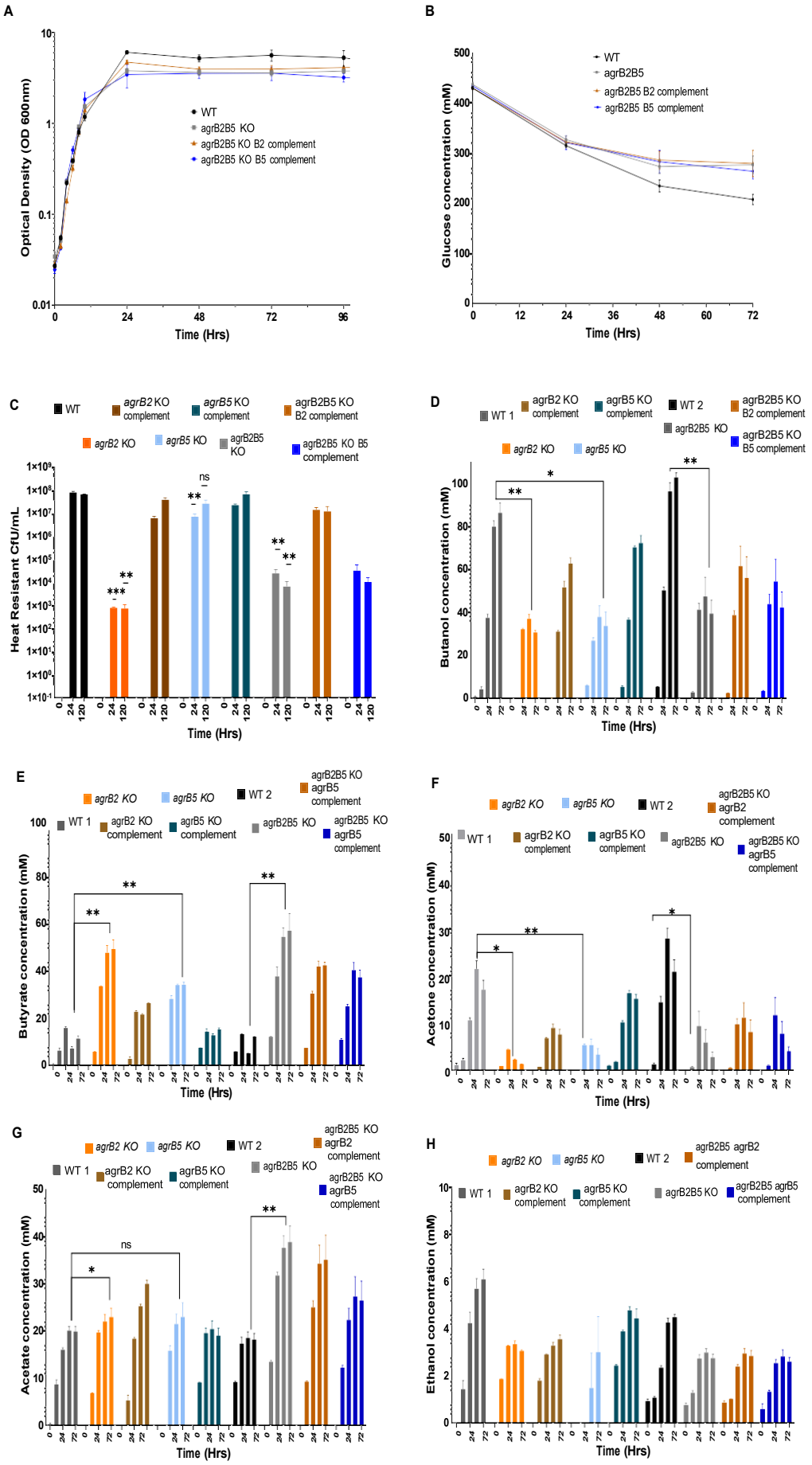

A
Optical Density (OD 600nm)
Time (Hrs)
WT
agrB2B5 KO
agrB2B5 KO B2 complement
agrB2B5 KO B5 complement
B
Glucose concentration (mM)
Time (Hrs)
WT
agrB2B5
agrB2B5 B2 complement
agrB2B5 B5 complement
C
Heat Resistant CfU/mL
Time (Hrs)
WT
agrB2 KO complement
agrB5 KO complement
agrB2B5 KO B2 complement
agrB2 KO
agrB5 KO
agrB2B5 KO
agrB2B5 KO B5 complement
D
Butanol concentration (mM)
Time (Hrs)
WT 1
agrB2 KO complement
agrB5 KO complement
WT 2
agrB2B5 KO B2 complement
agrB2B5 KO
agrB2 KO
agrB5 KO
agrB2B5 KO B5 complement
E
Butyrate concentration (mM)
Time (Hrs)
agrB2 KO
agrB5 KO
WT 2
agrB2B5 KO agrB5 complement
WT 1
agrB2 KO complement
agrB5 KO complement
agrB2B5 KO
agrB2B5 KO agrB5 complement
F
Acetone concentration (mM)
Time (Hrs)
agrB2 KO
agrB5 KO
WT 2
agrB2B5 KO agrB2 complement
WT 1
agrB2 KO complement
agrB5 KO complement
agrB2B5 KO
agrB2B5 KO agrB5 complement
G
Acetate concentration (mM)
Time (Hrs)
agrB2 KO
agrB5 KO
WT 2
agrB2B5 KO agrB2 complement
WT 1
agrB2 KO complement
agrB5 KO complement
agrB2B5 KO
agrB2B5 KO agrB5 complement
H
Ethanol concentration (mM)
Time (Hrs)
agrB2 KO
agrB5 KO
WT 2
agrB2B5 agrB2 complement
WT 1
agrB2 KO complement
agrB5 KO complement
agrB2B5 KO
agrB2B5 agrB5 complement

Figure 5.7 Growth consumption, endospore formation and butanol formation analysis for the *C. beijerinckii* NCIMB 8052 *agrB2B5* KO mutants.

Growth curves for each sample is shown by graph A. Values plotted are the mean with standard error of the mean (SEM). High performance liquid chromatography (B) was used to measure glucose concentration for the *C. beijerinckii* NCIMB 8052 *agrB2B5 (CBEI_RS03605/Cbei_0658, CBEI_RS20300/Cbei_3965)* mutant alongside the complemented strains and the WT. C, heat resistant spore CFU/ml were calculated after 0, 24 and 120 hours of growth in CBM broth (6% glucose). Gas chromatography was used to measure butanol (D), butyrate (E), acetone (F), acetate (G), ethanol (H). The values plotted are the means with standard error of the mean (SEM) from a minimum of 3 technical replicates. The statistical difference between the WT and *agrB* mutants was analysed using the T-test with P value ≤0.05. The degree of statistical significance is shown as <0.033 (*), <0.002(**), <0.001(***).

The *agrB4B5* KO mutant growth levelled at 3.3 OD600nm by 24 hours, with the comparative WT levelling at 5.4 OD600nm (**Figure 5.8A**). Under these growth conditions the *agrB4B5* KO mutant had a 5-log reduction in endospore formation compared to the WT. This was partially complemented by either the *agrB4* or *agrB5* gene independently (**Figure 5.8C**). Butanol, acetone and butyrate formation was not significantly reduced in the double mutant. Whilst acetate formation was slightly reduced by 6.4 mM by 72 hours, this was not complemented by either the *agrB4* or *agrB5* gene independently (**Figure 5.8G**). No significant difference in ethanol formation was observed.

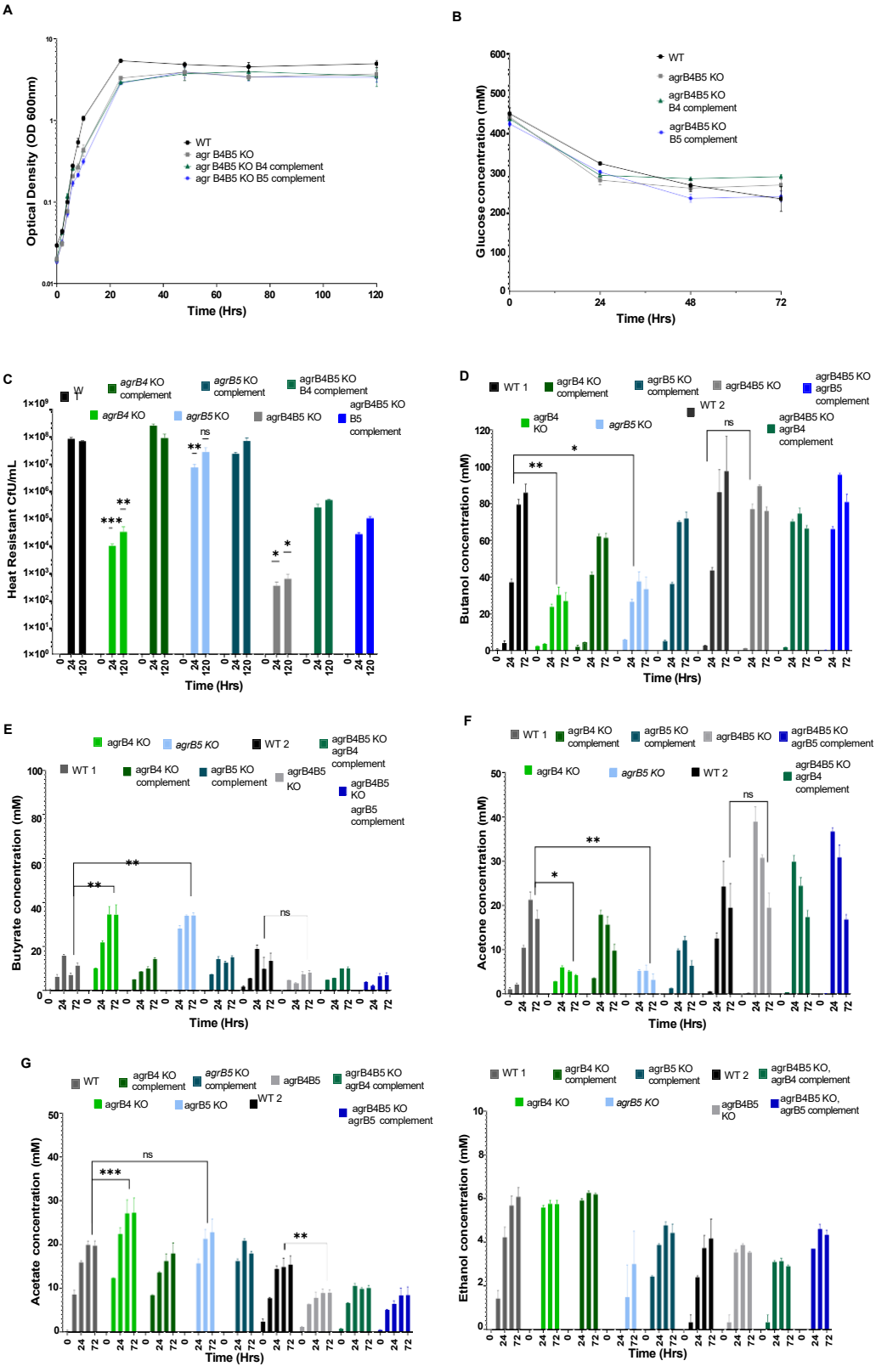

A
Optical Density (OD 600nm)
Time (Hrs)
WT
agr B4B5 KO
agr B4B5 KO B4 complement
agr B4B5 KO B5 complement
B
Glucose concentration (mM)
Time (Hrs)
WT
agrB4B5 KO
agrB4B5 KO B4 complement
agrB4B5 KO B5 complement
C
Heat Resistant CfU/mL
Time (Hrs)
W
agrB4 KO
agrB5 KO
agrB4B5 KO
agrB4 KO complement
agrB5 KO complement
agrB4B5 KO B4 complement
agrB4B5 KO B5 complement
D
Butanol concentration (mM)
Time (Hrs)
WT 1
WT 2
agrB4 KO
agrB5 KO
agrB4B5 KO
agrB4 KO complement
agrB5 KO complement
agrB4B5 KO agrB4 complement
agrB4B5 KO agrB5 complement
E
Butyrate concentration (mM)
Time (Hrs)
WT 1
WT 2
agrB4 KO
agrB5 KO
agrB4B5 KO
agrB4 KO complement
agrB5 KO complement
agrB4B5 KO agrB4 complement
agrB4B5 KO agrB5 complement
F
Acetone concentration (mM)
Time (Hrs)
WT 1
WT 2
agrB4 KO
agrB5 KO
agrB4B5 KO
agrB4 KO complement
agrB5 KO complement
agrB4B5 KO agrB4 complement
agrB4B5 KO agrB5 complement
G
Acetate concentration (mM)
Time (Hrs)
WT
WT 2
agrB4 KO
agrB5 KO
agrB4B5
agrB4 KO complement
agrB5 KO complement
agrB4B5 KO agrB4 complement
agrB4B5 KO agrB5 complement
Ethanol concentration (mM)
Time (Hrs)
WT 1
WT 2
agrB4 KO
agrB5 KO
agrB4B5 KO
agrB4 KO complement
agrB5 KO complement
agrB4B5 KO, agrB4 complement
agrB4B5 KO, agrB5 complement

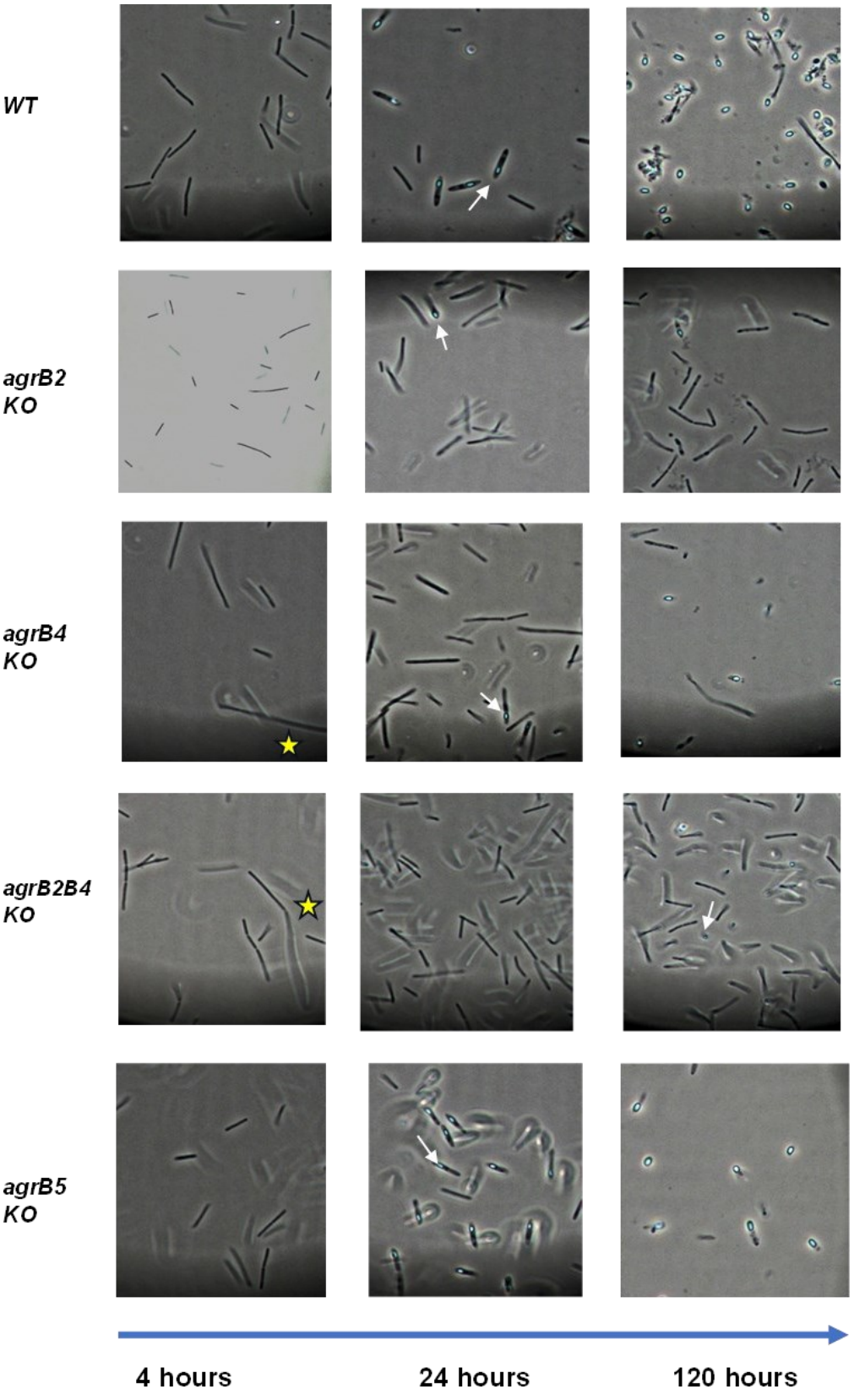
Figure 5.8 Growth curve, glucose consumption, endospore formation and butanol formation analysis for the *C. beijerinckii* NCIMB 8052 *agrB4B5* KO mutants.

Growth curves for each sample are shown by graph A. Values plotted are the mean with standard error of the mean (SEM). High performance liquid chromatography (B) was used to measure glucose concentration for the *C. beijerinckii* NCIMB 8052 *agrB4B5* (CBEI_RS16315/Cbei_3171, CBEI_RS20300/Cbei_3965) mutant alongside the complemented strains and the WT. C, heat resistant spore CFU/ml were calculated after 0, 24 and 120 hours of growth in CBM broth (6% glucose). Gas chromatography was used to measure butanol (D), butyrate (E), acetone (F), acetate (G). The values plotted are the means with standard error of the mean (SEM) from a minimum of 3 technical replicates. The statistical difference between the WT and *agrB* mutants was analysed using the T-test with P value ≤0.05. The degree of statistical significance is shown as <0.033 (*), <0.002(**), <0.001(***).

5.3 Light microscope analysis

The *agrB2*, *agrB4*, *agrB5*, *agrB2B4* mutants and, as a sporulation-negative control, a *spo0A* mutant (Humphreys, 2019) were observed under a light microscope using phase contrast, alongside the WT, to analyse cells at 4 hours, 24 hours and 120 hours. These results are shown in **Figure 5.9**. At 4 hours, during the exponential phase of growth, the cells were observed to be highly motile in all samples except the *spo0A* mutants, which had a much lower degree of motility. Whilst the *agrB2* and *agrB5* mutants were observed to be short rod-shaped cells like the WT, the *agrB4*, *agrB2B4* and *spo0A* mutants were observed to be producing short and long rod-shaped cells. By 24 hours, the WT and the *agrB5* mutant were both producing many forespores, whilst the *agrB2* and *agrB4* were producing fewer forespores. Meanwhile, the *agrB2B4* and *spo0A* mutants were still lacking the presence of forespore structures. By this point, cultures had entered the stationary phase, and it was observed that all cultures had little to no motility under microscopy, in agreement with previous work on solventogenic clostridia showing a rapid decline in cell motility after 20 hours (Long et al., 1984). By 120 hours, the *spo0A* mutant was still producing no forespores or extracellular spores. The WT and *agrB5* mutant were both producing higher quantities of these compared to the *agrB2*, *agrB4* and *agrB2B4* mutants, the latter of which was producing the fewest spores. When cell:cell ratios were observed under light microscopy the

WT showed 36% of the counted population to be free extracellular spores, with 10% of cells showing forespores. Similarly, the *agrB5* KO mutant counted population appeared as 54% extracellular spores and 8% forespores. In contrast, the *agrB2* KO mutant ratios were 3% and 5%, whilst the *agrB4* KO mutant ratios were 13% and 6% respectively.


WT
agrB2 KO
agrB4 KO
agrB2B4 KO
agrB5 KO
4 hours
24 hours
120 hours

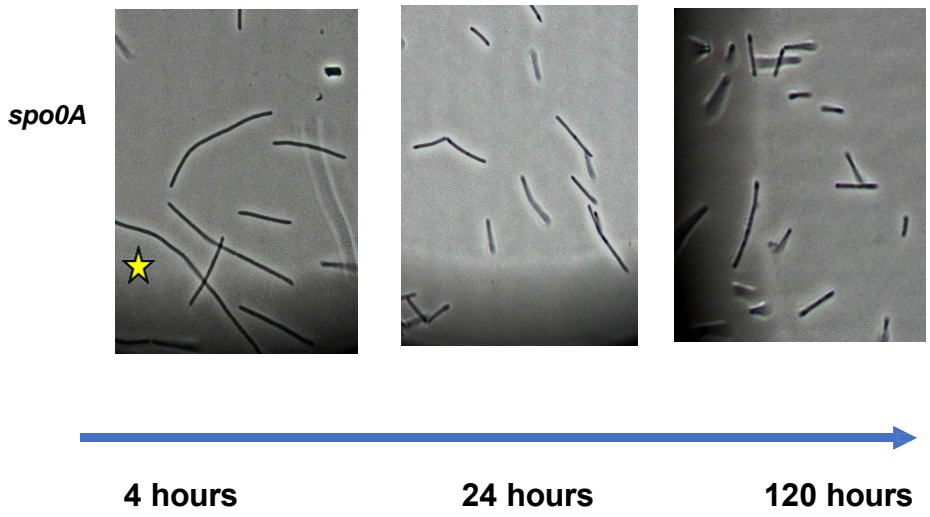


Figure 5.9 Light microscope analysis of the *C. beijerinckii* NCIMB 8052 WT, *agrB* mutants and a *spo0A* mutant using phase contrast.

Samples were observed at 4 hours, 24 hours, and 120 hours of growth. Labels on the left describe the gene(s) mutated in each sample. White arrows indicate forespores or extracellular spores when first observed in the sample. Yellow stars indicate the long-rod phenotype observed in some samples. Images taken at 1000X magnification (400x magnification on *agrB2* KO 4 hours) on a Nikon Eclipse Ci.

5.3.1 Transmission Electron Microscopy

To confirm the observation that the *agrB* mutants still produced forespores, as suggested by light microscopy, the mutants were observed under transmission electron microscopy (TEM) in the School of Life Sciences imaging facility (SLIM) at the University of Nottingham (see **section 2.6.7**). The *agrB2*, *agrB4* and *agrB5* mutants were observed to produce forespores at 24 hours (Figure 5. **B-D**). There did appear to be fewer forespores in the *agrB4* mutant in comparison to the WT. When the *agrB2B4* mutant was observed under TEM there appeared to be very few forespores observed at all across the whole sample. Exact quantification of forespores was difficult as many cells were not being cross sectioned at the required angle.

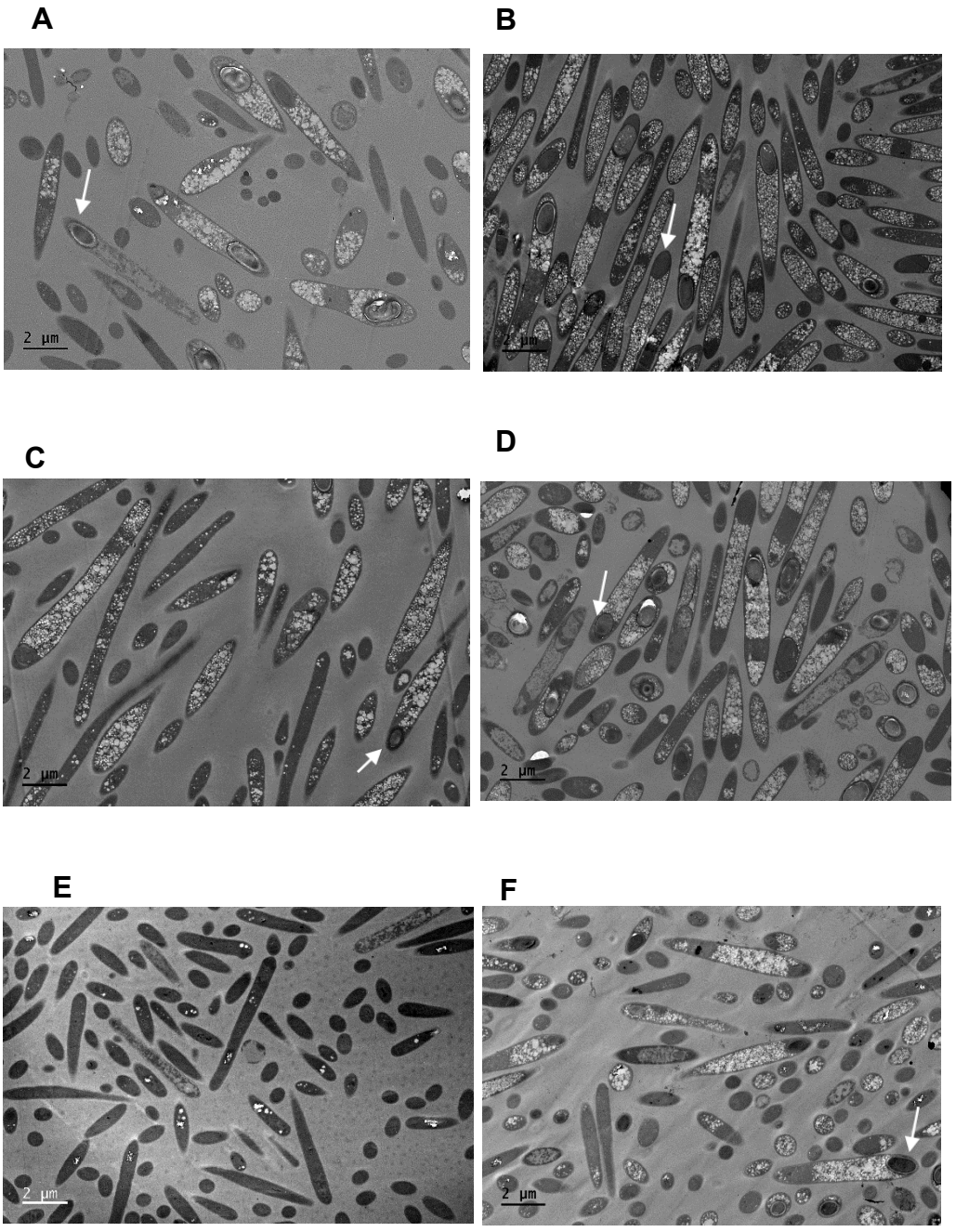


Figure 5.10 Transmission Electron Microscopy of the *C. beijerinckii* NCIMB 8052 *agrB* mutants and WT after 24 hours growth.

WT (A), *agrB2* KO (B), *agrB4* KO (C), *agrB5* KO (D), agrB2B4 KO (E), *agrB2B5* KO (F) and *agrB4B5* KO (G). White arrows indicate forespores. Yellow stars indicate long-rod shaped phenotypes. Images taken at magnification 1200X.

5.3.2 Super Resolution Microscopy

Super resolution microscopy was used to further confirm the presence of forespore structures based on DNA location within the cells and location of membrane structures. **Figure 5.11** to **Figure 5.15** show the high-resolution microscopy images for the WT, *agrB2* KO mutant, *agrB4* KO mutant, *agrB5* KO mutant and the *spo0A* mutant respectively after 24 hours growth in liquid media. The cells were stained with the DNA dye Hoechst which appears as blue and the membrane stain FM 4-64 FX which appears red. Clear evidence of forespore structures are indicated by white arrows in the super resolution microscope images, with the WT showing subterminal forespore structures **(Figure 5.11).** These forespore structures are also visible in the *agrB2* and *agrB5* mutants where the compartmentalised DNA is more prominent. The *agrB4* mutant occasionally showed interesting mother cell structures which appeared to have two forespores at opposite ends of the same cell **(Figure 5.13).** Elongated cells were observed for the *agrB*4 KO mutant and the *spo0A* mutant. Three examples are indicated in **Figure 5.13** for the *agrB4* KO mutant, reaching around 20 to 40 µm in length. **Figure 5.15**, indicates elongated mother cells observed in the *spo0A* mutant sample, here with examples measuring around 15 to 16 µm. This is larger than the average mother cell observed for the WT, examples being measured between 5 to 10 µm (**Figure 5.11**). The DNA distribution (shown in blue) also appeared to be visibly altered in the

agrB4 and *spo0A* mutants. Whilst the WT DNA appears condensed within the cell (see **Figure 5.11**), in the *agrB4* and *spo0A* mutants particularly, DNA appears to be more dispersed throughout the mother cells (**Figure 5.14** and **Figure** 5.15). However, it is also evident in the *spo0A* mutant and *agrB4* mutant that there are examples where stained DNA regions in long cells are more defined.

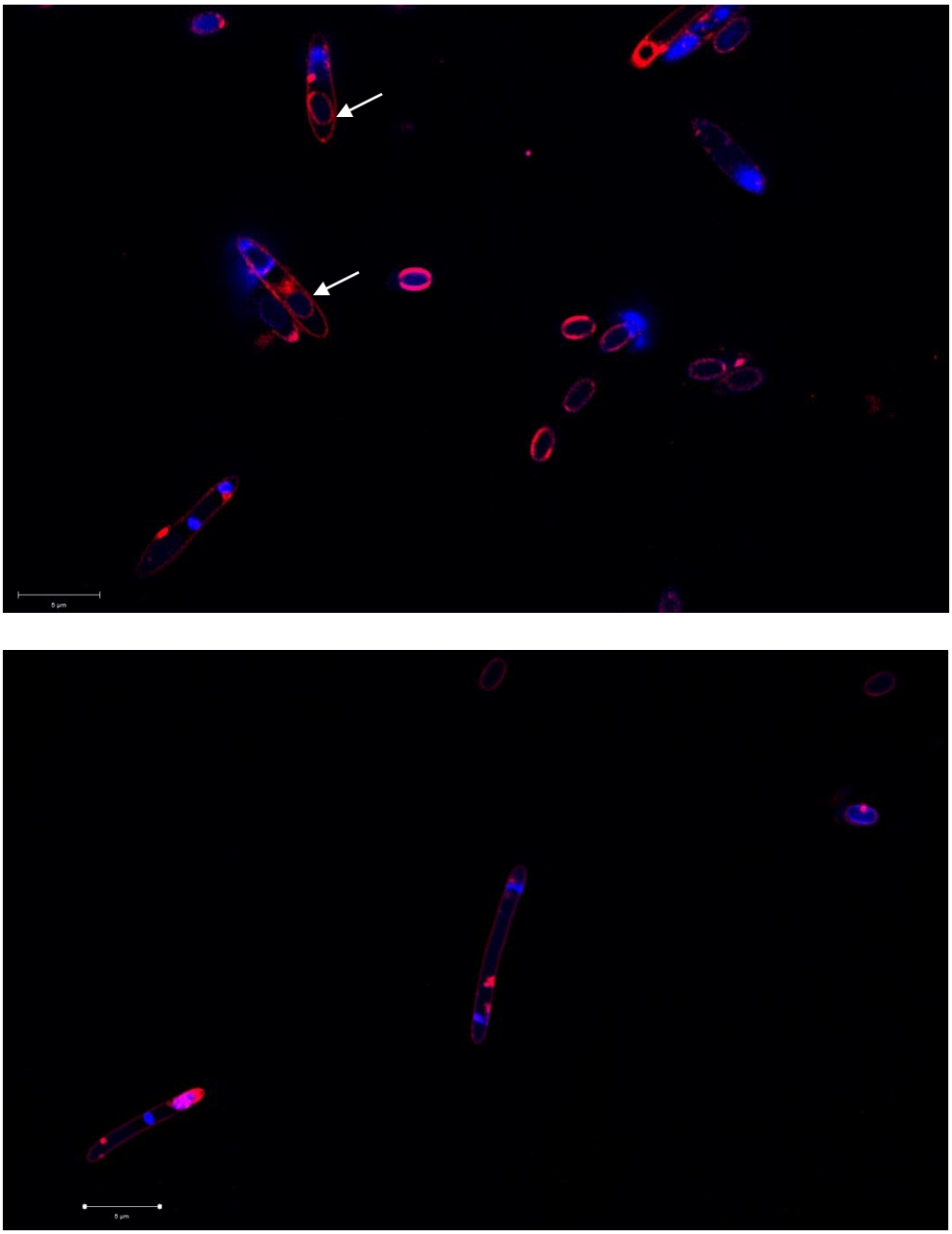


Figure 5.11 Super Resolution Microscope images of the *C. beijerinckii* NCIMB 8052 WT at 24 hours growth.

Images were taken at 1260 X magnification. A scale bar of 5 µm was used. DNA dye Hoechst appears as blue, and the membrane stain FM 4-64 FX appears red. White arrows indicate forespore structures.

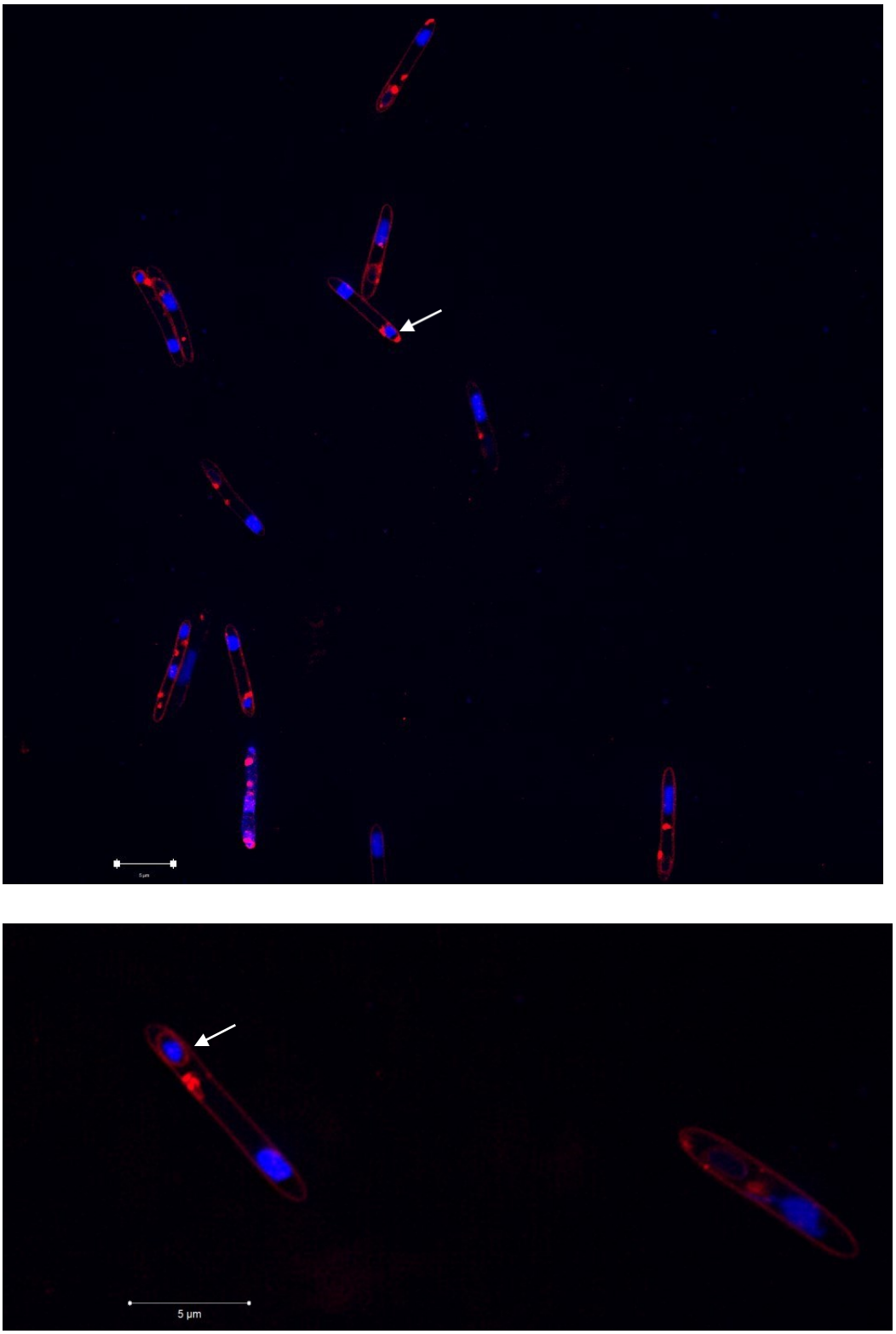


Figure 5.12 Super Resolution Microscope images of the *C. beijerinckii* NCIMB 8052 *agrB2* KO mutant at 24 hours growth.

Images were taken at 1260 X magnification. A scale bar of 5 µm was used. DNA dye Hoechst appears as blue, and the membrane stain FM 4-64 FX appears red. White arrows indicate forespore structures.

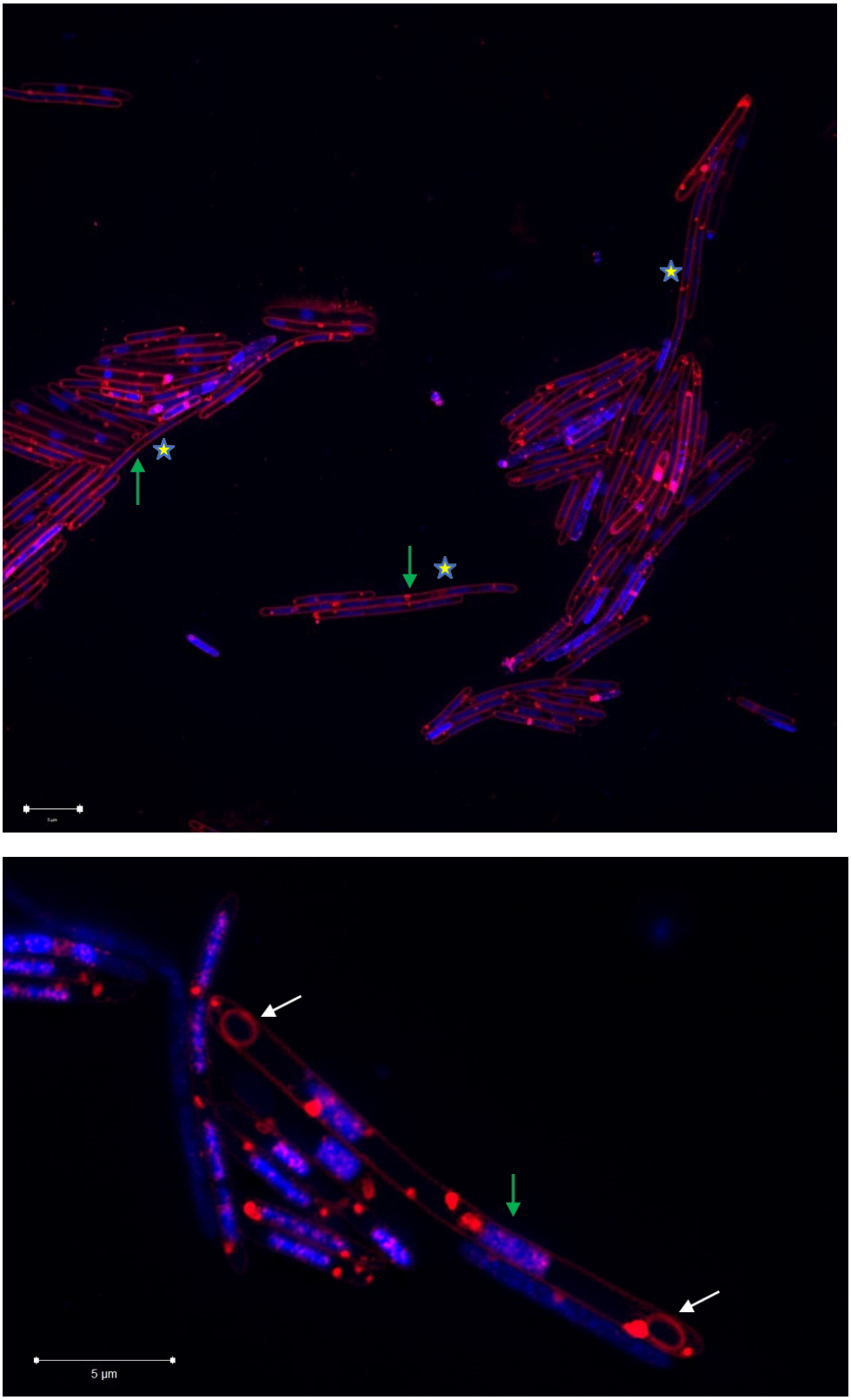


Figure 5.13 Super Resolution Microscope images of the *C. beijerinckii* NCIMB 8052 *agrB4* KO mutant at 24 hours growth.

Images were taken at 1260 X magnification. A scale bar of 5 µm was used. DNA dye Hoechst appears as blue, and the membrane stain FM 4-64 FX appears red. White arrows indicate forespore structures. Green arrows indicate elongated cells with multiple defined DNA segments. Yellow stars indicate elongated cells.

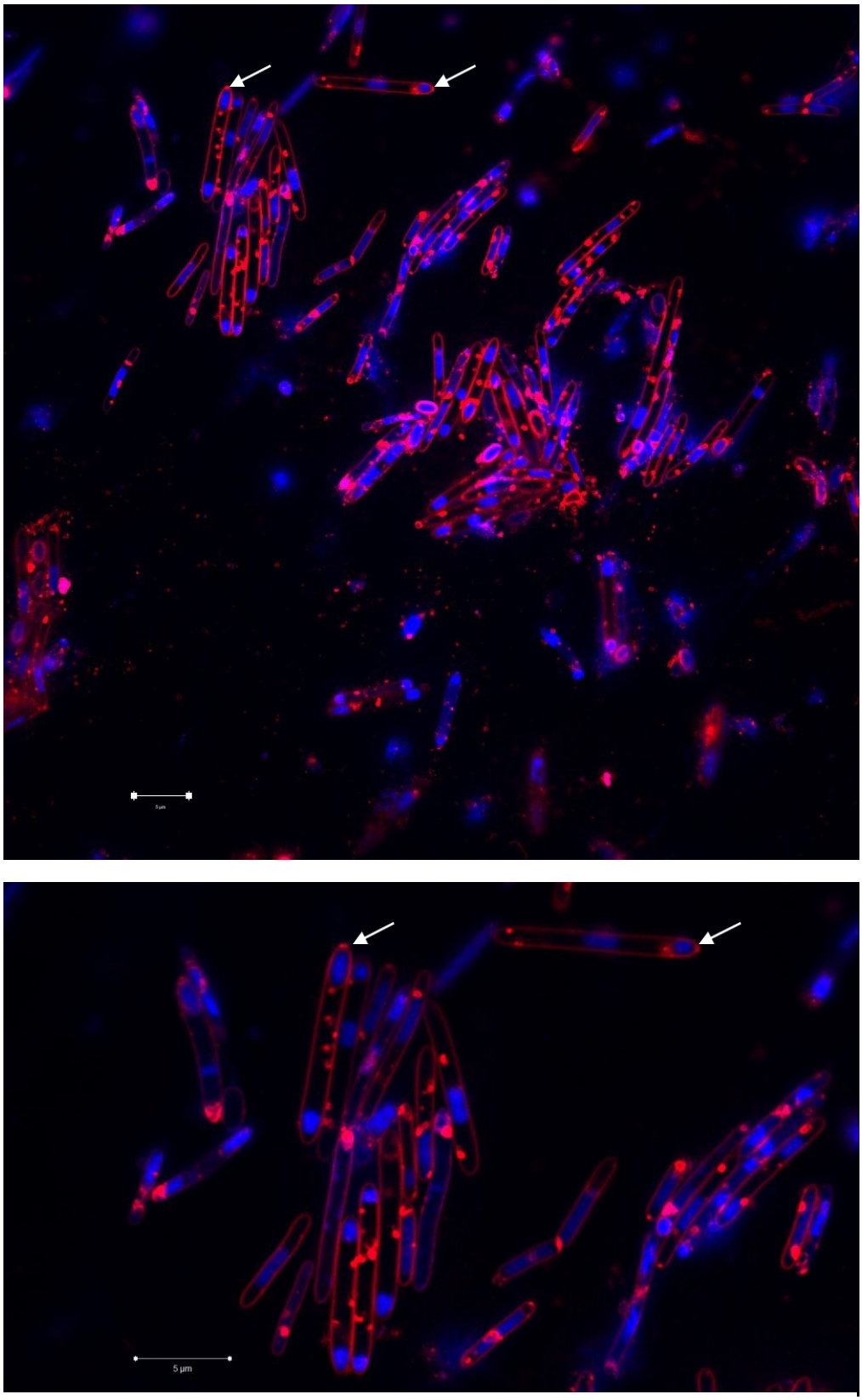


Figure 5.14 Super Resolution Microscope images of the *C. beijerinckii* NCIMB 8052 *agrB5* KO mutant at 24 hours growth.

Images were taken at 1260 X magnification. A scale bar of 5 µm was used. DNA dye Hoechst appears as blue, and the membrane stain FM 4-64 FX appears red. White arrows indicate forespore structures.

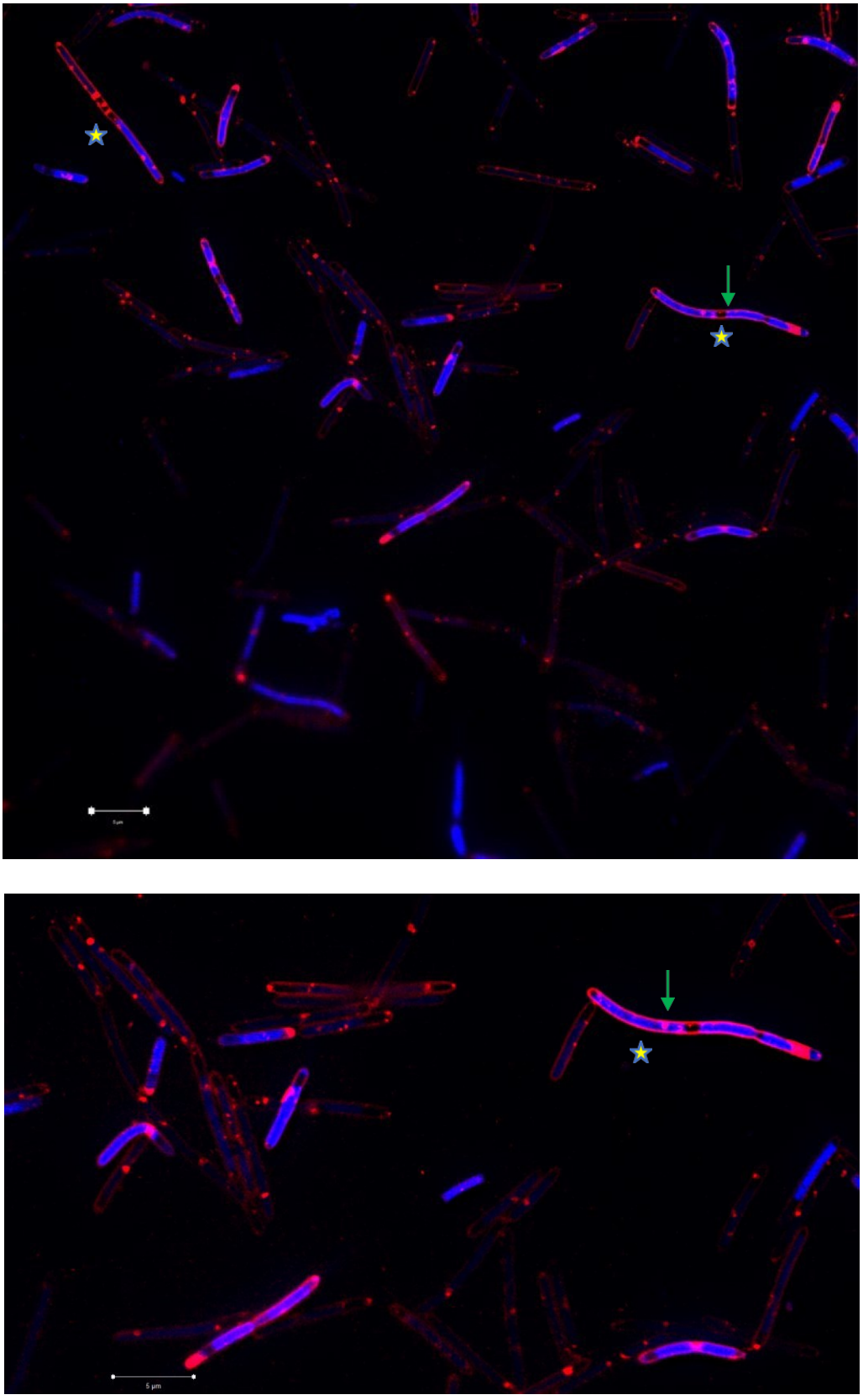


Figure 5.15 Super Resolution Microscope images of the *C. beijerinckii* NCIMB 8052 *spo0A* mutant at 24 hours growth.

Images were taken at 1260 X magnification. A scale bar of 5 µm was used. DNA dye Hoechst appears as blue, and the membrane stain FM 4-64 FX appears red. Green arrows indicate elongated cells with multiple defined DNA segments. Yellow stars indicate elongated cells.

5.4 Discussion

Studies have suggested that QS may influence metabolic pathways, sporulation and/or solvent production in the genus *Clostridium*. For example, Steiner *et al.* (2012), showed that an Agr QS system regulates granulose and sporulation production in *C. acetobutylicum*. It has also been suggested, in *C. beijerinckii* NRRL B-598, that sporulation-related genes may be regulated by Agr QS, based on indirect evidence into RNA-seq data during butanol shock (Patakova *et al.*, 2021). Furthermore, in *C. autoethanogenum*, Agr QS influences the Wood-Ljungdahl pathway (Piatek *et al.*, 2022). Therefore, it is possible that the Agr systems in *C. beijerinckii* NCIMB 8052 could influence such factors.

In this study, three independent *agr* gene clusters were analysed for their role in growth, granulose formation, sporulation, and solvent formation in *C. beijerinckii* NCIMB 8052. Specifically, this chapter investigated the phenotypic characteristics of the three *agrB* mutants created in **chapter 4**, as these were assumed to be devoid of specific AIP signals. A mutation of *agrB* is thought to prevent the production of the mature AIP signal, encoded by an associated *agrD* gene. For instance, in the species *S. aureus* and *C. acetobutylicum*, inactivation of *agrB* abolished production of the associated AIPs (Steiner *et al.*, 2012; Thoendel and Horswill, 2013), whilst allowing complementation of the resulting signalling defect with exogenously added AIP. This is in

agreement with AgrB's confirmed role in AgrD-processing and secretion (Qiu *et al.*, 2005; Thoendel and Horswill, 2009, 2013).

Solventogenic Clostridium species are prone to degeneration, and hence care has to be taken when generating mutants using procedures that are based on selection processes and isolation of single colonies (Humphreys *et al.*, 2023). During this study a granulose assay was used as a preliminary screen on all the *agrB* mutants generated to identify and eliminate potentially degenerated *agrB* mutant isolates. This revealed that some of the isolates differed from the other biological replicates in that they showed very little or no granulose accumulation, and degenerate colony morphologies. These were signs of degeneration, as identified by Humphreys *et al,* (2023), and so these isolates were discarded from further analysis.

5.4.1 Growth characteristics of *agrB* mutants

The *agrB* mutants brought forward for investigation were fully phenotypically analysed regarding growth rate, granulose accumulation, sporulation, and solvent formation. This study observed that the *agrB* mutants analysed demonstrated a similar growth rate to the WT, around 0.40 h^{-1}. However, all appeared to consistently plateau at a lower final OD_{600}. The growth rate is an important measurement as

it has been proposed that a reduced growth rate in *C. beijerinckii* can stabilise solvent formation (Li *et al.*, 2011). Furthermore, the *agrB* mutants were observed to consume lower amounts of glucose over 96 hours of growth in liquid media, when compared to the WT. This could also be genetically complemented to similar levels of the WT. It may be that the reduced glucose consumption contributed to the lower overall OD_{600} and vice versa, therefore these observations would need further analysis to determine if these were a direct or indirect result of the *agr* systems and whether the changes in OD were indicative of changes in cell numbers, or rather associated with cell size, shape or other OD influencing factors.

5.4.2 Reduced sporulation in *agrB* mutants

The *agrB* mutant isolates appeared to show similar levels of granulose accumulation compared to the WT 8052 strain, shown by a dark brown coloration after staining with Lugol's iodine solution. This finding contrasts with those observed by Steiner *et al.* (2012), who observed no or very little granulose accumulation in *C. acetobutylicum* ATCC 824 *agr* mutants. In the Steiner study, it was not clear whether granulose formation was directly or indirectly regulated by *agr* QS through the in-turn reduction in sporulation also, since it is known the granulose

accumulation is associated with the onset of sporulation in *Clostridium* species (Diallo, Kengen and López-Contreras, 2021). A study by Diallo *et al.* (2020) found that in a study on *C. beijerinckii* NCIMB 8052 *spoIIE* mutant, genes involved in granulose formation were upregulated, whilst genes involved in granulose degradation were unchanged when observing RNA-Seq data. In both *C. acetobutylicum* and *C. beijerinckii*, inactivation of the master regulator *spo0A* resulted in loss of both granulose and spore formation (Ravagnani *et al.*, 2000). Since granulose accumulation was largely unchanged in *agrB* mutant in this study, it could be that the *spo0A* gene is not directly affected by the Agr systems inactivated here.

Interestingly, whilst granulose accumulation in the *agrB* mutants in this study did not appear to be largely affected, heat resistant endospore formation was significantly reduced in the *agrB2* and *agrB4* mutants after 24 and 120 hours. Meanwhile the *agrB5* mutant had reduced heat-resistant endospore formation at 24 hours, but this then increased to similar CFU/ml as the 8052 WT, around 10^8. The AgrB2 system was found to be conserved in all *C. beijerinckii* strain genomes analysed in **chapter 3**. This system, when mutated, appeared to cause a large reduction in heat-resistant spore formation, suggesting this system holds an important regulatory role in endospore formation. A similar observation was made for the Agr4 system, present in 91% of the strains analyses in **chapter 3**. When these systems were knocked out in combination to produce the *agrB2B4* mutant, endospore formation was further reduced, with no spore-derived colonies being observed on the 24-hour plates. The Agr system has been shown to play an

important role in sporulation for various *Clostridium* species including *C. sporogenes*, group I *C. botulinum*, *C. perfringens* and *C. acetobutylicum* (Cooksley *et al.*, 2010; Li *et al.*, 2011; Steiner *et al.*, 2012).

However, even though the Agr5 system was highly conserved, in 97% of strains analysed, this study did not show much of an effect on heat resistant endospore formation resulting from *agrB5* mutation. Whilst this system did appear to delay sporulation at 24 hours compared to the WT, this was overcome by 120 hours. Therefore, whilst some *agr* systems appear to hold an important role in heat-resistant endospore formation, others do not. Supporting this hypothesis, when the *agrB2* and *agrB5* genes were knocked out in combination, the double mutant showed a 4-log reduction in heat-resistant CFU/ml, compared to the WT, that was only complemented by the *agrB2* gene, and not the *agrB5* gene. Similarly, in the *agrB4B5* double mutant, the reduced heat-resistant CFU/ml was complemented more so by the *agrB4* gene. These findings suggest that the role of the AgrB5 system in sporulation is limited compared to the AgrB2 and AgrB4 systems and could be indirect by influencing other aspects of the organism's physiology and metabolism.

In this study the WT and *agrB* mutants were further observed under phase contrast microscopy, TEM and high-resolution microscopy to examine the presence of spores in culture and cell morphology. As expected, the *agrB5* mutant was producing a high level of forespores, similar to the WT, by 24 hours. Meanwhile the *agrB2*, *agrB4* and

agrB2B4 mutants were producing much fewer. The *agrB4* mutant was also observed to have some cells where two asymmetric septa were present in the opposite ends of the same cell, and both the *agrB4* and *spo0A* mutants had many elongated mother cells. These were both seen in *spoIIE* mutants of *C. beijerinckii* NCIMB 8052 (Diallo, Kint, *et al.*, 2020) and *B. subtilis* (Barák and Youngman, 1996). In these studies, they hypothesised therefore that the role of SpoIIE in *C. beijerinckii* closely resembles that described in *B. subtilis*, where SpoIIE enables the activation of SigF and the positioning of asymmetric septa. Interestingly, the *agrB4*, *agrB2B4 and agrB4B5* mutants were also observed to be producing longer rod-shaped cells compared to the WT, suggesting that the *agrB4* mutation was a key factor resulting in this long-rod cell phenotype. There are also examples where stained DNA regions in long cells are more defined in the *spo0A* and *agrB4* mutants, suggesting that they represent distinct chromosomes distributed in regular intervals along the length of the cell. This suggests that these cells replicated their DNA, potentially for packing into daughter cells for forespore formation, but then did not happen and the cell continued growing and replicating.

The presence of forespores in the *agrB* mutants which showed a reduced CFU/ml of heat-resistant spores suggests abortive sporulation in which the cells initiate but fail to complete spore development. This has been observed in *B. subtilis* in minimal media by Mijakovic *et al.* (2016), who suggested this be a way for cells to delay or escape from unsuccessful sporulation. Whether AgrB4 regulates a similar escape

route in *C. beijerinckii* remains to be seen. The fact that *agrB2* and *agrB4* mutants still produce granulose and forespores suggests that regulation by QS is not directly involved in *spo0A* expression during the onset and progression of the sporulation process. This is in contrast to Bacillus subtilis, where RRNPP-type QS has been shown to control the degree of Spo0A phosphorylation (Perego and Hoch, 1996; Jiang, Grau and Perego, 2000).

5.4.3 Reduced solvent formation in *agrB* mutants

This study demonstrated a reduction in butanol and acetone concentration in the *C. beijerinckii* NCIMB 8052 *agrB2*, *agrB4*, *agrB5* single mutants and the *agrB2B4* and *agrB2B5* double mutants, however not in the *agrB4B5* double mutant. This reduced solvent concentration could be fully or partially complemented genetically suggesting the *agr* systems play a regulatory role in solvent formation. However, whether this is a direct result of the Agr systems or indirect result of lower glucose consumption/ final OD_{600} observed during growth has yet to be confirmed. There is evidence supporting the idea that the Agr systems may play a regulatory role in the switch from acidogenesis to solventogenesis. For instance, the higher butyrate concentrations observed in all *agrB* single mutants after 24 hours compared to the WT, would suggest that less butyric acid was being

converted into butanol in the mutants, leading to the reduced butanol concentrations observed. Likewise, increased acetate production in the *agrB2* and *agrB4* mutants would suggest less acetate-acetone conversion. This increased acid concentration in the *agrB* mutants could also be genetically complemented. It may therefore be that the Agr systems analysed affect the switch from acidogenesis to solventogenesis. Usually, *spo0A* inactivation has effects on solvent production in both *C. acetobutylicum* and *C. beijerinckii* (Ravagnani *et al.*, 2000; Harris, Welker and Papoutsakis, 2002). However, based on the observed reduction in sporulation with no effect on granulose, leading to the hypothesis that *spo0A* regulation in not affected in these *agrB* mutants, it is likely that this potential *agr* regulation of the acid-solvent shift would be independent of *spo0A*. This increased acid concentration in the media may explain why the *agrB* mutants observed a lower overall final OD_{600}, since an acidified medium would reduce the viability of bacterial cells, causing increased cell death. The study on *C. acetobutylicum* by Steiner *et al.* (2012) did not observe any significant change in solvent formation in the *agr* mutants generated. However, a study by Kotte *et al.* (2020), into the RRNPP QS system showed that a *qsrB* mutant of *C. acetobutylicum* was observed to have increased solvent formation, and overexpression of this gene significantly reduced both endospore and solvent production. Meanwhile, Feng *et al.* (2020), found that RRNPP systems positively regulated the *sol* operon in *C. saccharoperbutylacetonicum,* In the acetogen *C. autoethanogenum* Agr QS appears to affect the ratio of acetate and ethanol produced (Piatek *et al.*, 2022). Taken together, these studies suggest that QS regulation

of acid and solvent metabolism, as well as sporulation, appears to be a common trait in the clostridial species studied so far.

5.4.4 Conclusions

In summary, these findings establish a role for Agr QS in solventogenesis and endospore production. In contrast to the findings from *C. beijerinckii* in this study, Steiner *et al.* (2012) found that disruption of *agr* genes in *C. acetobutylicum* had no effect on acid and solvent formation from glucose. However, Kotte *et al.* (2020) did observe that the RRNPP-type QS does affect solvent formation and sporulation. Since *C. beijerinckii* NCIMB 8052 does not contain RRNPP-type QS systems, it may help explain why this study observed a more profound role for Agr QS systems.

The Agr systems analysed in this study are all shown to have an effect on solvent formation, especially for butanol and acetone formation. It may be that these systems are involved in the regulation of the switch from acidogenesis to solventogenesis for butyrate-butanol conversion and, in the case of Agr2 and Agr4, acetate-acetone conversion. However, the *agrB5* mutation only appeared to reduce acetone production and had no significant effect on acetate concentrations in the single *agrB5* mutant, therefore this Agr system may only hold a regulatory role later in acetone formation rather than during the acetate-acetone switch. When *agrB2* and *agrB5* were knocked-out in combination, there was a larger increase in acetate formation compared to the WT and the single *agrB2* mutant. Two of the systems, the Agr2 and Agr4 system, were also observed to have a large decrease in heat-

resistant endospores when disrupted. Whilst the *agrB2*, *agrB4* *agrB2B4*, *agrB2B5* and *agrB4B5* mutants were still observed to be producing forespore structures under the microscope, these were present in much fewer numbers compared to the WT and *agrB5* mutant.

Chapter 6:

Transcriptional changes in *agrB* mutants

6.1 Introduction

In solventogenic *Clostridium* species, accumulation of acetate and butyrate, in the presence of excess nutrients, are suggested to be the main trigger of sporulation, and solventogenesis, this is due to the reduced culture pH. These two processes provide short-term and long-term survival mechanisms for the bacterial cells (Sedlar *et al.*, 2019; Diallo, Kengen and López-Contreras, 2021).

The process of sporulation undergoes different stages. The first stage involves cells taking on swollen and cigar-like morphologies with granulose accumulation. During this stage DNA replication and axial filament formation are undertaken. Stage II occurs when cells form an asymmetric septum. The membranes from this stage begin to engulf the smaller compartment and thus stage III of engulfment is underway. At stage IV a cortex made of modified peptidoglycans is formed to separate the two membranes between the mother cell and forespore. In stage V the spore coat proteins are deposited on the outer membrane of the forespore and finally stages VI and VII involve maturation of the spore and its release into the surrounding environment, see **Figure 1.1** (Diallo et al., 2021).

The regulatory cascade of sporulation in *Clostridium acetobutylicum* is summarised by **Figure 6.1**, with the master regulator Spo0A taking up a key position in the initiation of the process. In *Clostridium* spp., in

contrast to the phosphorelay system employed by members of the genus *Bacillus*, it appears that phosphorylation of Spo0A is carried out by orphan histidine kinases. In *C. acetobutylicum* these were determined to be HK Cac903, HK Cac3319 and HK Cac0323 (Steiner *et al.*, 2011), whilst Xin *et al.* (2020), found Cbei4484 and Cbei2073 in *C. beijerinckii* have high sequence similarities to those found in *C. acetobutylicum*. Phosphorylated Spo0A ($Spo0A^P$) and σ^H regulate the expression of the *spoIIE* gene and the *spoIIA* operon, which encodes. SpoIIAA, an anti-anti-sigma factor, SpoIIAB, an anti-sigma factor, and σ^F, a sigma factor. The SpoIIE phosphatase, initiated by $Spo0A^P$, dephosphorylates SpoIIA which leads to the release of σ^F from SpoIIAB. This then leads to downstream expression of σ^E, specific to the mother cell. Meanwhile, the forespore specific σ^G becomes active after the mother cell has engulfed the forespore. In *B. subtilis* 80-plus genes, related to spore maturation, are regulated by σ^G. These include *spoIVB*, which activates the later σ^K, and the *spoVA* operon (Diallo, Kengen and López-Contreras, 2021). In *C. acetobutylicum* σ^K expression has been suggested to be necessary to initiate sporulation, through enabling Spo0A expression, as well as being crucial during late sporulation (Jones and Papoutsakis, 2014). The point at which QS systems affect this cascade is of interest if scientists are to manipulate solventogenic *Clostridium* species in the future. For instance, by controlling sporulation without interfering with solvent formation or other important functions, but also for a more fundamental and thorough understanding of this important biological phenomenon.

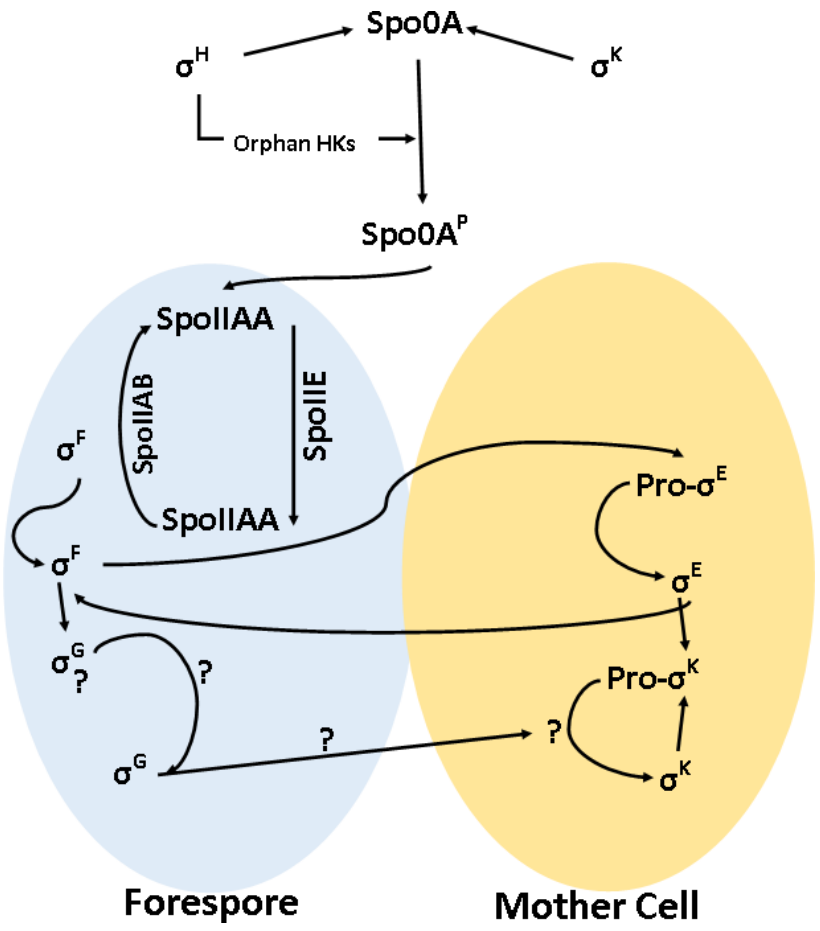


Figure 6.1 Summary of the sporulation signaling cascade in *Clostridium* (Adapted from Al-Hinai et al., 2015).

The whole genome of *C. beijerinckii* NCIMB 8052 has been sequenced and the identified genes catalogued (Wang *et al.*, 2011; Sun *et al.*, 2018). However, to gain further insight into gene regulation and functionality, transcriptomic methods are invaluable. The two main methods for transcriptomic analysis are DNA microarrays and RNA-

seq. The use of RNA-seq as a method for transcriptional analysis has proven to be advantageous over microarray due to having a higher dynamic range with no upper limits for expression levels (Wang, Gerstein and Snyder, 2009; Kukurba and Montgomery, 2015; D'Agostino, Li and Wang, 2022). Microarrays have poor background and sensitivity at very low and high gene expression levels. They only allow researchers to track the expression of pre-selected genes whereas RNA-seq can be used on the whole genome (Wang et al., 2009). In *C. acetobutylicum*, global gene expression has been widely studied using RNA-seq (Herman et al., 2017; Luo et al., 2020; Ralston & Papoutsakis, 2018; Servinsky et al., 2015; Venkataramanan et al., 2015). Other solventogenic clostridia such as *C. saccharoperbutylacetonicum* and *C. beijerinckii* have also been studied (Diallo et al., 2020; Li et al., 2020; Patakova et al., 2019, 2021, 2022; Sedlar et al., 2018, 2019). A study by Wang et al. (2013) undertook genome-wide RNA-seq in WT *C. beijerinckii* NCIMB 8052 across its growth which gave an insight into gene regulation in solventogenesis and sporulation. The use of transcriptomic data in this study will allow us to determine which agr QS gene clusters are used by the organism under the employed conditions and which genes are affected by the individual *agrB* mutations.

Transcriptomic data are often validated by techniques such as qPCR (Diallo et al., 2020). When the starting material is RNA, the process is called reverse transcriptase-qPCR, here a complementary DNA (cDNA) is first synthesised. This cDNA is then carried forward into real time-

PCR amplification. Amplified DNA is usually quantified by a fluorescent probe. When validating RNA-seq data, this process requires reference genes which show unchanged expression levels under different conditions. These can be used to normalise level of gene expression alongside the selected genes (Patakova et al., 2022).

The aims of this chapter were to:

- Extract total RNA from the *C. beijerinckii* NCIMB 8052 WT and *agrB* single KO mutants at both 12 hours and 24 hours.
- Send total RNA for RNA-seq and observe which specific genes are differentially regulated in relation to sporulation, solventogenesis and sugar-uptake genes or any other genes showing interesting or unexpected changes.
- Validate RNA-seq data using qPCR.

6.2 Results

6.2.1 RNA Extraction and RNA Sequencing

Total RNA samples were taken from three replicates of *agrB* mutant and three wild type replicate cultures that had been grown as described in **Section 2.2.2**. This was done at 12 and 24 hours of growth to observe differential gene expression (see **Figure 6.2**). These time points were chosen because a significant different in solvent formation was shown in the Δ*agrB* mutant strains compared to the WT. It was reasoned that at 12 hours, when differences were small but statistically significant, with regards to solvent formation, indirect changes resulting from the introduced mutations would be less pronounced and hence genes directly affected by QS more easily identified. By 24 hours the cells had reached the stationary phase, producing high amounts of mature endospores, with the WT producing 10^8 heat- resistant CFU/ml. Once extracted via phenol-chloroform based method, the RNA was quality checked using a agarose gel electrophoresis and an Agilent bioanalyser. After RNA integrity and quantity were confirmed, it was sent for sequencing on the Illumina NextSeq500 platform provided by the DeepSeq service (University of Nottingham).

The *C. beijerinckii* NCIMB 8052 WT was compared to each individual *agrB* mutant for both 12- and 24-hours RNA extractions. The fold

change values were provided in respect to the WT (WT vs Mutant), meaning a value of larger than 1-fold (because $\log^0 = 1$) change in gene expression meant reduced expression in the respective *agrB* mutant comparison.

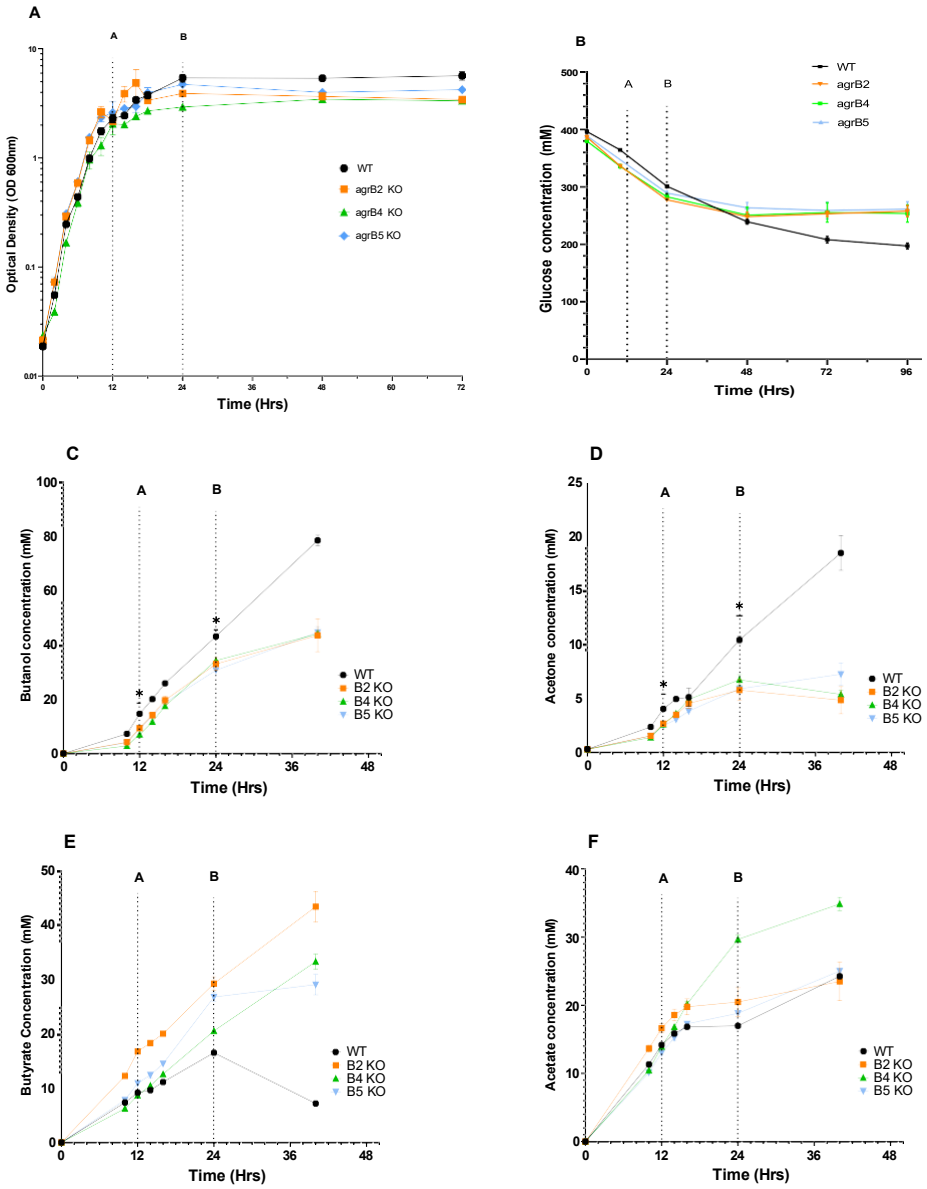


Figure 6.2 Growth, glucose consumption and solvent analysis of the *C. beijerinckii* NCIMB 8052 *agrB* mutants and WT for RNA extraction time point.

The growth curves for WT and *agrB* mutants are shown by graph A. The glucose consumption was measured using HPLC, shown in graph B. For solvents, 1 ml samples were taken from each 60 ml culture at 0, 10, 12, 14, 16 and 40 hours. Samples were centrifuged at the highest speed for 1 minute and the supernatant collected then stored at -20 °C. Gas chromatography was used to measure butanol (C), acetone (D), butyrate (E) and acetate (F) concentrations for the *C. beijerinckii* NCIMB 8052 *agrB2* (CBEI_RS03605/Cbei_0658), *agrB4* (CBEI_RS16315/Cbei_3171) and *agrB5* (CBEI_RS20300/Cbei_3965) mutants alongside the WT. The values plotted are the means with standard error of the mean (SEM) from technical triplicates. The T test was used to identify statistically significant (*) differences in butanol and acetone production at 12 hrs and 24 hrs sampling points, with P value ≤0.05. The degree of statistical significance is shown as <0.033 (*).

6.2.2 RNA-seq Result Analysis

A total of 5385 genes were found to be transcribed in the RNA-seq data for each sample set. At the 12 hour time point the RNA-Seq analysis revealed that; 3004 genes (56% of all transcribed genes) were significantly differentially regulated in the *agrB4* mutant compared to the wild type, with 1783 of those genes being downregulated and 1221 genes upregulated ($p \leq 0.05$, fold change ≥ 2, (WT v *agrB4* KO) ≥ 2). At 12 hours, 1642 genes (30% of all transcribed genes) were significantly differentially regulated in the *agrB2 KO* mutant compared to the wild type, with 969 of those genes being downregulated and 673 being upregulated ($p \leq 0.05$, fold change ≥ 2, (WT v *agrB2* KO) ≥ 2). At 12 hours, 188 genes (3% of all transcribed genes) were significantly differentially regulated in the *agrB5 KO* mutant compared to the wild type, of these 177 were downregulated and 11 genes upregulated ($p \leq 0.05$, fold change ≥ 2, (WT v *agrB5* KO) ≥ 2). The top 20 genes most upregulated and downregulated for these mutants at the 12 hour time point can be seen in **Table 6.1 - Table 6.3.**

At the 24 hour time point the RNA-Seq analysis revealed that; 3257 genes (60% of all transcribed genes) were significantly differentially regulated in the *agrB4* mutant compared to the wild type with 1923 of these being downregulated and 1334 being upregulated ($p \leq 0.05$, fold change ≥ 2, (WT v *agrB4* KO) ≥ 2). At 24 hours, 1537 genes (29% of all

transcribed genes) were significantly differentially regulated in the *agrB2* mutant compared to the wild type, of these 781 genes were downregulated and 756 upregulated ($p \leq 0.05$, fold change ≥ 2, (WT v *agrB2* KO) ≥ 2). At 24 hours, 448 genes (8% of all transcribed genes) were significantly differentially regulated in the *agrB5* mutant compared to the wild type with 247 of these genes being downregulated and 201 upregulated ($p \leq 0.05$, fold change ≥ 2, (WT v *agrB5* KO) ≥ 2). The top 20 genes most upregulated and downregulated for these mutants at the 24 hour time point can be seen in **Table 6.1 - Table 6.3.**

Table 6.1 Top 20 differentially upregulated and downregulated genes in the *C. beijerinckii* NCIMB 8052 *agrB2* KO at 12 hours and 24 hours.

Upregulated 12 hours				
Gene ID	**product**	**baseMean**	**padj**	**Fold Change**
CBEI_RS27230	hypothetical protein	49	6.5E-09	6.7
CBEI_RS10050	ABC transporter permease subunit	214	1.1E-03	6.7
CBEI_RS10405	P-II family nitrogen regulator	803	1.1E-02	6.9
CBEI_RS23880	L-fucose/L-arabinose isomerase family protein	4760	3.2E-36	7.0
CBEI_RS10400	P-II family nitrogen regulator	784	8.7E-03	7.2
CBEI_RS06625	hypothetical protein	2424	8.3E-30	7.3
nifH	nitrogenase iron protein	6096	5.2E-03	7.5
nifK	nitrogenase molybdenum-iron protein subunit beta	8687	6.5E-03	7.5
CBEI_RS10045	ABC transporter permease	228	1.6E-04	7.7
CBEI_RS10370	ATP-binding cassette domain-containing protein	630	2.1E-06	9.0

CBEI_RS08155	hypothetical protein	3958	4.5E-61	9.0
CBEI_RS07440	NADH peroxidase	215931	2.7E-62	9.1
CBEI_RS04890	HK97 family phage prohead protease	4	9.7E-03	10.3
trpA	tryptophan synthase subunit alpha	3928	1.3E-17	19.2
trpB	tryptophan synthase subunit beta	4032	9.8E-18	21.5
CBEI_RS09120	phosphoribosylanthranilate isomerase	2674	2.9E-19	23.4
trpC	indole-3-glycerol phosphate synthase TrpC	3333	5.7E-20	28.8
trpD	anthranilate phosphoribosyltransferase	4962	6.5E-21	31.2
trpE	anthranilate synthase component I	5557	1.1E-21	31.2
CBEI_RS09105	aminodeoxychorismate/anthranilate synthase component II	2123	4.1E-22	31.4
Downregulated 12 hours				
Gene ID	**product**	**baseMean**	**padj**	**Fold Change**
CBEI_RS10750	spore coat associated protein CotJA	43449	2.9E-99	534.6
CBEI_RS10740	manganese catalase family protein	54411	7.6E-103	484.6
CBEI_RS10000	hypothetical protein	139309	2.4E-83	472.7
CBEI_RS10745	spore coat protein CotJB	24176	5.0E-100	416.8
CBEI_RS05010	ABC transporter ATP-binding protein	9494	3.1E-94	218.5
CBEI_RS05005	ABC transporter substrate-binding protein	7142	6.1E-92	205.1
CBEI_RS13470	hypothetical protein	11067	1.4E-140	185.7
CBEI_RS03720	thioesterase domain-containing protein	1669	3.7E-11	128.9
CBEI_RS05015	ABC transporter permease	12712	2.9E-63	126.6
CBEI_RS16855	hypothetical protein	41045	1.1E-49	117.1
CBEI_RS06995	hypothetical protein	81419	6.3E-90	93.4
CBEI_RS07000	hypothetical protein	65241	1.0E-78	86.1
CBEI_RS09770	YmaF family protein	21404	1.2E-38	86.1
CBEI_RS24015	nucleotide sugar dehydrogenase	16791	5.0E-94	83.6
CBEI_RS13335	hypothetical protein	227140	1.2E-108	74.8

CBEI_RS27860	hypothetical protein	964	3.7E-31	73.4
CBEI_RS19080	hypothetical protein	3451	1.6E-41	72.6
CBEI_RS23975	hypothetical protein	49061	1.3E-96	69.1
CBEI_RS24010	glycosyltransferase	10374	1.1E-101	68.7
CBEI_RS03705	ATP-binding cassette domain-containing protein	5571	2.4E-09	66.5
Upregulated 24 hours				
Gene ID	**product**	**baseMean**	**padj**	**Fold Change**
sufD	Fe-S cluster assembly protein SufD	18106	2.5E-18	18.1
CBEI_RS21115	dihydrodipicolinate synthase family protein	292	8.5E-19	18.4
CBEI_RS11415	ABC transporter ATP-binding protein	300	1.8E-14	18.8
panB	3-methyl-2-oxobutanoate hydroxymethyltransferase	9315	1.4E-66	18.8
CBEI_RS03855	MFS transporter	3677	7.4E-33	19.7
CBEI_RS09625	SUF system NifU family Fe-S cluster assembly protein	13812	1.6E-19	20.1
phoU	phosphate signaling complex protein PhoU	16741	2.4E-23	20.5
CBEI_RS11420	ABC transporter permease	290	1.4E-15	20.5
CBEI_RS25715	response regulator transcription factor	2696	1.4E-22	20.6
CBEI_RS10685	DsbA family oxidoreductase	5581	1.1E-28	21.3
CBEI_RS09620	cysteine desulfurase	27993	6.0E-20	21.6
CBEI_RS15095	efflux RND transporter periplasmic adaptor subunit	1976	1.1E-22	22.7
CBEI_RS15085	DHA2 family efflux MFS transporter permease subunit	2311	1.8E-33	23.8
deoC	deoxyribose-phosphate aldolase	10465	1.4E-23	27.3
CBEI_RS15090	efflux RND transporter periplasmic adaptor subunit	1026	1.9E-30	33.1
CBEI_RS11560	enoyl-CoA hydratase/isomerase family protein	1098	6.8E-09	35.4
CBEI_RS11565	L-serine ammonia-lyase%2C iron-sulfur-dependent%2C subunit alpha	1215	1.5E-09	35.4
CBEI_RS15100	hypothetical protein	404	6.5E-27	37.1

CBEI_RS11550	M20 family metallopeptidase	4621	1.6E-17	118.3
CBEI_RS11555	enoyl-CoA hydratase/isomerase family protein	3929	4.2E-17	144.5
Downregulated 24 hours				
Gene ID	**product**	**baseMean**	**padj**	**Fold Change**
CBEI_RS03720	thioesterase domain-containing protein	1669	1.3E-16	639.9
CBEI_RS03725	hypothetical protein	271	1.0E-09	113.3
CBEI_RS03705	ATP-binding cassette domain-containing protein	5571	4.1E-10	83.6
CBEI_RS19050	YmaF family protein	84219	2.6E-81	77.0
CBEI_RS11950	phage holin family protein	12832	1.6E-102	71.8
CBEI_RS16020	hypothetical protein	162024	1.4E-49	67.6
CBEI_RS03685	AMP-binding protein	820	1.1E-09	54.8
CBEI_RS16785	phage holin	82299	4.4E-52	42.9
CBEI_RS12950	hypothetical protein	16873	2.5E-25	34.5
CBEI_RS03680	aldehyde dehydrogenase family protein	1379	1.6E-10	29.3
CBEI_RS15575	protease inhibitor I42 family protein	144515	4.5E-37	27.4
CBEI_RS07185	hypothetical protein	28	2.4E-12	26.6
CBEI_RS10905	hypothetical protein	15	3.1E-11	24.9
CBEI_RS20520	malate dehydrogenase	3	8.9E-03	22.8
CBEI_RS16215	hypothetical protein	104227	1.8E-22	22.1
CBEI_RS03745	AMP-binding protein	3543	4.2E-10	21.5
CBEI_RS03730	radical SAM protein	3558	1.7E-08	20.2
CBEI_RS11315	hypothetical protein	927	2.6E-08	19.3
CBEI_RS03740	alcohol dehydrogenase catalytic domain-containing protein	1349	8.0E-08	19.0
CBEI_RS13355	DNRLRE domain-containing protein	574977	1.9E-40	18.7

Table 6.2 Top 20 differentially upregulated and downregulated genes in the *C. beijerinckii* NCIMB 8052 *agrB4* KO at 12 hours and 24 hours.

Upregulated 12 hours				
Gene ID	**product**	**baseMean**	**padj**	**Fold change**
CBEI_RS20075	ATP-binding protein	2685	5.0E-101	15.0
CBEI_RS14725	TetR/AcrR family transcriptional regulator	292	1.3E-83	15.2
CBEI_RS06625	hypothetical protein	2424	6.6E-60	17.0
CBEI_RS01570	hypothetical protein	2280	4.9E-50	17.0
CBEI_RS20400	ISNCY-like element ISCb1 family transposase	1455	0.0E+00	17.1
CBEI_RS13040	Cof-type HAD-IIB family hydrolase	5960	9.3E-124	17.2
CBEI_RS14930	ISNCY-like element ISCb1 family transposase	1522	0.0E+00	18.6
CBEI_RS01565	ABC transporter ATP-binding protein	2627	1.6E-42	19.0
CBEI_RS01560	hypothetical protein	1857	4.1E-131	20.6
CBEI_RS01555	hypothetical protein	3383	6.5E-66	20.6
CBEI_RS12955	ISNCY-like element ISCb1 family transposase	12970	0.0E+00	23.7
CBEI_RS07735	ISNCY-like element ISCb1 family transposase	161	2.7E-176	24.8
CBEI_RS25545	ISNCY-like element ISCb1 family transposase	8918	0.0E+00	27.1
CBEI_RS17655	ISNCY-like element ISCb1 family transposase	8686	0.0E+00	27.2
CBEI_RS14845	ISNCY-like element ISCb1 family transposase	13059	0.0E+00	28.2

CBEI_RS07440	NADH peroxidase	215931	4.6E-170	39.0
CBEI_RS14850	SDR family oxidoreductase	7714	6.3E-233	53.8
CBEI_RS14710	CAP domain-containing protein	3179	0.0E+00	91.9
CBEI_RS14720	multidrug efflux MFS transporter	13169	0.0E+00	181.1
CBEI_RS14715	isochorismatase family protein	1924	0.0E+00	192.4
Downregulated 12 hours				
Gene ID	**product**	**baseMean**	**padj**	**Fold change**
CBEI_RS16950	DUF346 domain-containing protein	8948	3.5E-48	199676.2
CBEI_RS17070	DUF975 family protein	8862	1.2E-46	151512.7
CBEI_RS16865	YmaF family protein	41802	3.2E-39	102766.4
CBEI_RS17090	alkaline phosphatase family protein	10359	3.3E-43	98094.6
CBEI_RS17110	hypothetical protein	26873	4.8E-38	86927.0
CBEI_RS17175	desulfoferrodoxin	27462	1.1E-41	79837.5
CBEI_RS16855	hypothetical protein	41045	3.3E-37	78235.7
CBEI_RS17755	4Fe-4S binding protein	6716	1.1E-41	75445.6
CBEI_RS17075	hemolysin family protein	25321	2.0E-40	65582.7
CBEI_RS17100	ABC transporter substrate-binding protein	4359	2.5E-37	43250.9
CBEI_RS17935	FAD-dependent oxidoreductase	16766	2.0E-33	35628.5
CBEI_RS18170	glycine/betaine ABC transporter substrate-binding protein	1998	2.2E-33	30086.7
CBEI_RS18180	O-acetylhomoserine aminocarboxypropyltransferase/cysteine synthase	1934	2.8E-33	28487.9
CBEI_RS17010	GGDEF domain-containing protein	3068	9.8E-34	28485.6
CBEI_RS17095	ABC transporter ATP-binding protein	2505	1.4E-33	25764.5
trxB	thioredoxin-disulfide reductase	3063	5.4E-34	25614.9
CBEI_RS17245	cell wall-binding protein	2543	1.5E-33	22616.2

CBEI_RS17335	cell wall-binding repeat-containing protein	1789	1.3E-33	22577.7
CBEI_RS18165	ABC transporter ATP-binding protein	1524	1.2E-31	21025.7
CBEI_RS17105	ABC transporter permease subunit	2086	4.4E-32	18956.8
Upregulated 24 hours				
Gene ID	**product**	**baseMean**	**padj**	**Fold Change**
CBEI_RS26775	FeoB-associated Cys-rich membrane protein	1754	8.1E-159	63.9
CBEI_RS11560	enoyl-CoA hydratase/isomerase family protein	1098	1.2E-12	70.2
CBEI_RS11565	L-serine ammonia-lyase%2C iron-sulfur-dependent%2C subunit alpha	1215	6.3E-14	74.2
CBEI_RS19275	iron-siderophore ABC transporter substrate-binding protein	1092	1.3E-219	79.0
CBEI_RS14710	CAP domain-containing protein	3179	5.2E-284	82.2
CBEI_RS19270	iron ABC transporter permease	480	1.1E-120	85.5
CBEI_RS12465	flavodoxin	766	1.1E-103	94.7
CBEI_RS15100	hypothetical protein	404	1.9E-43	96.9
CBEI_RS15095	efflux RND transporter periplasmic adaptor subunit	1976	3.5E-49	100.5
CBEI_RS11420	ABC transporter permease	290	6.0E-36	101.4
CBEI_RS11415	ABC transporter ATP-binding protein	300	4.6E-36	106.4
CBEI_RS19265	iron chelate uptake ABC transporter family permease subunit	550	1.7E-201	124.3
CBEI_RS15085	DHA2 family efflux MFS transporter permease subunit	2311	9.2E-78	125.5
CBEI_RS15600	heavy-metal-associated domain-containing protein	354	1.9E-80	125.9
feoB	ferrous iron transport protein B	30566	3.8E-137	146.1

CBEI_RS15090	efflux RND transporter periplasmic adaptor subunit	1026	2.3E-64	161.8
CBEI_RS14720	multidrug efflux MFS transporter	13169	0.0E+00	169.3
CBEI_RS14715	isochorismatase family protein	1924	0.0E+00	187.5
CBEI_RS11550	M20 family metallopeptidase	4621	2.3E-25	305.4
CBEI_RS11555	enoyl-CoA hydratase/isomerase family protein	3929	2.9E-25	414.2
Downregulated 24 hours				
Gene ID	**product**	**baseMean**	**padj**	**Fold Change**
CBEI_RS16865	YmaF family protein	41802	5.1E-51	553509.7
CBEI_RS17075	hemolysin family protein	25321	1.2E-52	331784.0
CBEI_RS16880	HD domain-containing protein	21038	1.1E-51	298334.3
CBEI_RS17175	desulfoferrodoxin	27462	1.7E-49	228010.5
CBEI_RS17510	DUF6143 family protein	39463	7.5E-60	187631.3
CBEI_RS16855	hypothetical protein	41045	6.2E-81	82809.7
CBEI_RS17090	alkaline phosphatase family protein	10359	1.4E-41	78712.3
CBEI_RS17935	FAD-dependent oxidoreductase	16766	1.4E-47	65709.5
CBEI_RS17110	hypothetical protein	26873	7.3E-62	58808.6
CBEI_RS17130	hypothetical protein	7368	2.0E-35	49504.7
CBEI_RS27500	hypothetical protein	6990	4.3E-33	42925.7
CBEI_RS17755	4Fe-4S binding protein	6716	3.3E-37	39769.0
CBEI_RS17100	ABC transporter substrate-binding protein	4359	2.2E-36	37655.1
CBEI_RS17125	hypothetical protein	3986	2.5E-32	30254.7
CBEI_RS17140	sigma-70 family RNA polymerase sigma factor	4017	3.3E-32	27300.9
CBEI_RS17070	DUF975 family protein	8862	1.8E-34	26842.7
CBEI_RS17010	GGDEF domain-containing protein	3068	2.5E-33	26645.6

trxB	thioredoxin-disulfide reductase	3063	3.1E-34	26608.5
CBEI_RS16885	hypothetical protein	1666	1.2E-31	20547.0
CBEI_RS17095	ABC transporter ATP-binding protein	2505	3.7E-32	20479.5

Table 6.3 Top 20 differentially upregulated and downregulated genes in the *C. beijerinckii* NCIMB 8052 *agrB5* KO at 12 hours and 24 hours.

Upregulated 12 hours				
Gene ID	**product**	**Base Mean**	**padj**	**Fold Change**
CBEI_RS15425	NADH-quinone oxidoreductase subunit I	16	3.0E-02	2.0
CBEI_RS12320	xylose ABC transporter ATP-binding protein	45	2.2E-02	2.1
panC	pantoate--beta-alanine ligase	8633	4.8E-04	2.1
CBEI_RS03675	IS256 family transposase	90	2.9E-03	2.1
xylF	D-xylose ABC transporter substrate-binding protein	36	9.4E-04	2.2
CBEI_RS17930	MBL fold metallo-hydrolase	211	2.6E-02	2.4
CBEI_RS14715	isochorismatase family protein	1924	4.3E-13	2.7
CBEI_RS14720	multidrug efflux MFS transporter	13169	4.4E-17	2.8
CBEI_RS03710	class I SAM-dependent methyltransferase	1834	2.6E-02	3.5
CBEI_RS07370	MBL fold metallo-hydrolase	397	3.2E-45	3.6
CBEI_RS03695	MFS transporter	16048	2.6E-02	4.0
Downregulation 12 hours				
Gene ID	**product**	**baseMean**	**padj**	**Fold Change**
CBEI_RS01550	hypothetical protein	46434	1.3E-79	112.7

CBEI_RS18905	hypothetical protein	24185	3.9E-60	64.8
CBEI_RS18910	hypothetical protein	9924	2.6E-57	62.7
CBEI_RS19645	C1 family peptidase	14386	4.2E-104	43.1
CBEI_RS20295	CHAP domain-containing protein	7467	1.9E-62	38.2
CBEI_RS22515	hypothetical protein	5405	2.4E-53	36.9
CBEI_RS22510	hypothetical protein	6052	7.4E-63	34.3
CBEI_RS18900	ABC transporter permease	1894	4.2E-39	25.4
CBEI_RS01555	hypothetical protein	3383	1.8E-53	18.7
CBEI_RS22505	DnaJ domain-containing protein	525	5.1E-42	18.2
CBEI_RS01565	ABC transporter ATP-binding protein	2627	2.3E-35	17.5
CBEI_RS13805	discoidin domain-containing protein	2527	1.4E-113	16.7
CBEI_RS22500	FHA domain-containing protein	447	1.1E-46	16.4
CBEI_RS22520	hypothetical protein	3885	4.4E-54	16.4
CBEI_RS22495	FHA domain-containing protein	325	6.0E-27	14.1
CBEI_RS12120	aspartate ammonia-lyase	177783	1.1E-08	13.1
CBEI_RS01570	hypothetical protein	2280	3.9E-35	12.4
CBEI_RS18895	ABC transporter permease	2450	6.4E-51	11.9
CBEI_RS01560	hypothetical protein	1857	7.9E-60	10.0
CBEI_RS10405	P-II family nitrogen regulator	803	1.9E-02	9.1
Upregulated 24 hours				
Gene ID	**product**	**baseMean**	**padj**	**Fold Change**
CBEI_RS20645	hypothetical protein	27	1.2E-03	3.1
CBEI_RS03015	response regulator	14	1.9E-02	3.1
CBEI_RS17925	PhzF family isomerase	522	6.7E-15	3.2
CBEI_RS27905	hypothetical protein	117	2.0E-04	3.2
pheA	prephenate dehydratase	5293	8.4E-06	3.3
CBEI_RS21470	ferrous iron transport protein A	596	4.6E-04	3.3

CBEI_RS06460	riboflavin synthase	1582	1.4E-03	3.4
CBEI_RS21465	ferrous iron transport protein A	1291	2.4E-04	3.4
CBEI_RS27975	BMC domain-containing protein	39	6.7E-06	3.6
CBEI_RS20780	glycyl-radical enzyme activating protein	82	1.1E-06	3.7
CBEI_RS17920	MerR family transcriptional regulator	240	5.6E-13	3.7
CBEI_RS20785	glycyl radical protein	108	2.6E-09	3.9
CBEI_RS15100	hypothetical protein	404	2.8E-04	4.0
ribE	6%2C7-dimethyl-8-ribityllumazine synthase	1864	3.4E-05	4.3
CBEI_RS06465	bifunctional 3%2C4-dihydroxy-2-butanone-4-phosphate synthase/GTP cyclohydrolase II	4981	1.8E-05	4.4
ribD	bifunctional diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino) uracil reductase RibD	4529	2.6E-05	4.4
CBEI_RS02645	ECF transporter S component	7353	2.7E-08	4.5
CBEI_RS17120	MFS transporter	588	3.9E-08	6.7
CBEI_RS15715	ATP-binding cassette domain-containing protein	2038	2.9E-09	8.7
CBEI_RS17930	MBL fold metallo-hydrolase	211	2.0E-13	10.5
Downregulated 24 hours				
Gene ID	**product**	**baseMean**	**padj**	**Fold Change**
CBEI_RS01550	hypothetical protein	46434	8.2E-69	85.5
CBEI_RS22515	hypothetical protein	5405	2.8E-65	59.0
CBEI_RS18910	hypothetical protein	9924	1.7E-40	42.8
CBEI_RS18905	hypothetical protein	24185	8.8E-40	32.5
CBEI_RS19645	C1 family peptidase	14386	4.2E-76	26.2
CBEI_RS01565	ABC transporter ATP-binding protein	2627	2.5E-36	23.5
CBEI_RS01555	hypothetical protein	3383	3.8E-52	23.1

CBEI_RS22510	hypothetical protein	6052	6.8E-51	22.9
CBEI_RS22520	hypothetical protein	3885	1.1E-55	16.9
CBEI_RS22485	serine/threonine protein phosphatase	772	7.6E-88	15.1
CBEI_RS20295	CHAP domain-containing protein	7467	5.0E-33	14.9
CBEI_RS18900	ABC transporter permease	1894	1.6E-19	13.2
CBEI_RS22500	FHA domain-containing protein	447	3.7E-42	12.1
CBEI_RS13805	discoidin domain-containing protein	2527	1.3E-81	11.5
CBEI_RS22490	protein phosphatase 2C domain-containing protein	678	4.0E-47	11.1
CBEI_RS22505	DnaJ domain-containing protein	525	2.3E-31	9.5
CBEI_RS18895	ABC transporter permease	2450	1.2E-29	9.4
CBEI_RS22495	FHA domain-containing protein	325	2.0E-20	9.3
CBEI_RS13640	hypothetical protein	107	1.1E-09	8.8
CBEI_RS20300	accessory gene regulator B family protein	2033	7.8E-34	8.2

6.2.3 Granulose and sporulation related gene expression

Granulose-related genes were analysed in the RNA-seq data, summarised in **Table 6.4**. Here it was observed that at 12 hours, genes involved in granulose biosynthesis (*glgA*, *glgB*, *glgD*) were upregulated in the *agrB2* and *agrB4* mutants by around 2 to 3-fold change, whilst genes involved in granulose degradation (*glgP*) were downregulated by 3.9-fold change in the *agrB2* mutant at 12 hours and 16.9-fold change in the *agrB4* mutant at 24 hours. In the *agrB5* mutant these genes remained unchanged at 12 and 24 hours compared to WT regulation.

The RNA-seq data also revealed that in the *agrB2* KO and *agrB4* KO, similar spore-related genes were down-regulated at 12 hours. These genes included, but were not limited to, those encoding sigma factor K (CBEI_RS05910), downregulated by 8-fold and 63-fold in in the *agrB2* KO and *agrB4* KO respectively. A collection of spore-coat related genes were downregulated including those encoding the CotS family spore coat proteins (CBEI_RS02205 and CBEI_RS02195), alpha/beta-type small acid-soluble spore proteins (CBEI_RS12770, CBEI_RS02705, CBEI_RS16815, CBEI_RS16765, CBEI_RS07595, CBEI_RS12145, CBEI_RS15865, CBEI_RS16025, CBEI_RS16695). Also, genes encoding stage V sporulation proteins, *spoVAC*, *spoVAD* and *spoVAE* (CBEI_RS04425, CBEI_RS04430, CBEI_RS04435), GPR endopeptidase (CBEI_RS04460), an M50 family metallopeptidase

thought to encode SpoIVFB (CBEI_RS02840) and SpoIVB peptidase (CBEI_RS08915) of which were downregulated by 14 and 232-fold change in the *agrB2* and *agrB4* mutants respectively. Of particular interest were the spore coat associated proteins genes encoding CotJA (CBEI_RS10750), downregulated by 534-fold, 1552-fold and 3-fold, CotJB (CBEI_RS10745), downregulated by 416-fold, 1271-fold and 3-fold, and CotJC (CBEI_RS10740), downregulated by 484-fold, 948-fold and 3-fold in the *agrB2*, *agrB4* and *agrB5* KO mutants respectively at 12 hours when compared to the WT. Although there was a reduction in spore germination experimentally for the *agrB2* and *agrB4* KO, interestingly there was no significant difference observed in gene expression for the master regulator of sporulation, *spo0A*, in either mutant at both 12 and 24 hours. However, there was an upregulation of the orphan histidine kinase genes, at 12 hours for CBEI_RS10760 (Cbei_2073), and 24 hours for CBEI_RS22900 (Cbei_4484) in the *agrB4* mutant, thought to be involved in the phosphorylation of Spo0A (Xin *et al.*, 2020). Interestingly, this study gave a clear indication that later stage sporulation-related genes tended to be downregulated rather than earlier stage genes, with the exception of the CBEI_RS17310 gene in the *agrB4* mutant, a AbrB/MazE/SpoVT family DNA-binding domain-containing protein which was downregulated by 5242-fold-change compared to the WT at 12 hours. This is visualised in the *agrB2 and agrB4* mutants at 12 hours by **Figure 6.3 and Figure 6.4.** In both figures, key sporulation genes of *C. beijerinckii* NCIMB 8052 were ordered in terms of when the gene is expressed from early to late

spore formation (Diallo, Kengen and López-Contreras, 2021). The change in gene expression compared to the WT was then added for each gene. There is a very clear point at which gene expression is largely downregulated in the *agrB2* and *agrB4* mutants, appearing at *sigK* and *spoIVB* genes. See appendix **Figure 8.8** and **8.9** for visualisation of 24 hour time points. In fact, as can be seen from **Figure 6.3,** many of the early sporulation gene were mildly upregulated in the *agrB2* mutant, whilst in the *agrB4* mutant they are mostly at similar expression levels to the WT.

Table 6.4 The Fold change in key spore-formation gene expression, for each WT v *agrB* mutant, at 12 and 24 hours of growth according to RNA-seq data.

The degree of fold change in gene expression has been visualised as a 3-colour heat map where values are represented as the WT v *agrB* mutant gene expression therefore, a value of <1 (green) represents an upregulated gene in the *agrB* mutant, a fold change of 1 (amber) represents no change in gene expression in the *agrB* mutant, and a fold change of >1 (red) represents a downregulated gene in the *agrB* mutant. Key genes are highlighted and early to late gene expression is demonstrated in the first three columns by light to darker shading of green. Blue numbers represent insignificant values (P>0.05).

New locus tag	Old locus tag	Product	Fold change			Fold change			Fold change	
Granulose related genes			WT 12h vs B2 12h	WT 24h vs B2 24h		WT 12h vs B4 12h.	WT 24h vs B4 24h.		WT 12h vs B5 12h.	WT 24h vs B5 24h.
CBEI_RS25085|glgA	Cbei_4908	glycogen synthase GlgA	0.4	0.4		0.4	0.2		0.9	1.0
CBEI_RS25090|glgB	Cbei_4909	1%2C4-alpha-glucan branching protein GlgB	0.4	0.6		0.4	0.4		0.8	1.2
CBEI_RS25065|glgD	Cbei_4904	glucose-1-phosphate adenylyltransferase subunit GlgD	0.3	0.5		0.6	0.4		0.9	1.1
CBEI_RS01575|glgP	Cbei_0285	glycogen/starch/alpha-glucan family phosphorylase	3.9	5.2		1.3	16.9		1.2	1.2
Spore related genes										
CBEI_RS11500	Cbei_2219	AbrB/MazE/SpoVT family DNA-binding domain-containing protein	2.1	0.9		10.8	0.6		1.2	0.8
CBEI_RS11745	Cbei_2270	AbrB/MazE/SpoVT family DNA-binding domain-containing protein	2.1	1.8		2.2	1.1		1.8	1.0
CBEI_RS17310	Cbei_3375	AbrB/MazE/SpoVT family DNA-binding domain-containing protein	0.5	0.3		5241.9	3842.7		1.1	0.4
CBEI_RS00835|sigH	Cbei_0135	RNA polymerase sporulation sigma factor SigH	0.5	1.5		0.3	0.7		0.8	0.8
CBEI_RS10760	Cbei_2073	ATP-binding protein (orphan Histidine Kinase)	0.9	1.3		0.3	1.5		0.8	0.6
CBEI_RS22900	Cbei_4484	ATP-binding protein (orphan Histidine Kinase)	2.1	1.0		0.5	0.2		1.2	0.7
CBEI_RS08920|spo0A	Cbei_1712	sporulation transcription factor Spo0A	0.8	0.9		0.5	0.6		0.9	1.3
CBEI_RS04410|spoIIAA	Cbei_0812	anti-sigma F factor antagonist	0.4	0.9		1.2	0.5		1.0	1.2
CBEI_RS00590|spoIIE	Cbei_0097	stage II sporulation protein E	0.2	4.5		1.3	21.9		0.8	1.7
CBEI_RS04415|spoIIAB	Cbei_0813	anti-sigma F factor	0.5	0.8		1.2	0.4		0.9	1.3
CBEI_RS04420|sigF	Cbei_0814	RNA polymerase sporulation sigma factor SigF	0.5	0.9		1.3	0.4		0.9	1.2
CBEI_RS02260|spoIIR	Cbei_0395	stage II sporulation protein R	1.3	0.9		0.8	1.0		1.4	0.7
CBEI_RS02400|spoIID	Cbei_0422	stage II sporulation protein D	0.5	1.9		2.2	15.2		0.7	3.6
CBEI_RS10200|spoIIM	Cbei_1962	stage II sporulation protein M	1.2	0.7		9.2	5.3		0.7	2.1
CBEI_RS05930	Cbei_1119	sigma-E processing peptidase SpoIIGA	0.3	1.1		1.9	2.9		0.7	1.2
CBEI_RS05935|sigE	Cbei_1120	RNA polymerase sporulation sigma factor SigE	0.3	1.3		1.4	4.3		0.7	1.2
CBEI_RS08820|spoIIIAA	Cbei_1692	stage III sporulation protein AA	0.3	0.5		1.3	3.4		0.7	1.6
CBEI_RS08825|spoIIIAB	Cbei_1693	stage III sporulation protein SpoIIIAB	0.2	0.4		1.1	2.5		0.8	2.1
CBEI_RS08830|spoIIIAC	Cbei_1694	stage III sporulation protein AC	0.2	0.5		1.3	3.1		0.7	2.7
CBEI_RS08835|spoIIIAD	Cbei_1695	stage III sporulation protein AD	0.2	0.5		1.1	3.3		0.6	2.8
CBEI_RS08840|spoIIIAE	Cbei_1696	stage III sporulation protein AE	0.3	0.6		1.2	4.6		0.6	2.9
CBEI_RS08845|spoIIIAF	Cbei_1697	stage III sporulation protein AF	0.3	0.7		1.2	5.0		0.6	2.9
CBEI_RS08850|spoIIIAG	Cbei_1698	stage III sporulation protein AG	0.3	0.8		1.1	4.6		0.6	3.0
CBEI_RS08855	Cbei_1699	SpoIIIAH-like family protein	0.4	0.5		1.1	5.0		0.7	3.3
CBEI_RS02410|spoIIID	Cbei_0424	sporulation transcriptional regulator SpoIIID	0.4	1.5		2.2	10.2		0.7	2.3
CBEI_RS05940|sigG	Cbei_1121	RNA polymerase sporulation sigma factor SigG	1.0	1.6		5.8	6.4		0.9	1.5
CBEI_RS06015|spoIVA	Cbei_1136	stage IV sporulation protein A	0.9	0.9		4.7	5.3		0.8	2.4
CBEI_RS08915|SpoIVB	Cbei_1711	SpoIVB peptidase	14.4	3.3		132.4	23.0		1.5	1.6
CBEI_RS02840|spoIVFB	Cbei_0501	M50 family metallopeptidase	0.8	1.6		3.4	9.9		0.6	2.1
CBEI_RS05910|sigK	Cbei_1115	RNA polymerase sporulation sigma factor SigK	8.8	0.9		63.4	9.7		1.5	0.8
CBEI_RS00455|spoVG	Cbei_0080	septation regulator SpoVG	0.5	3.7		0.4	3.7		1.1	1.4
CBEI_RS04425|spoVAC	Cbei_0815	stage V sporulation protein AC	6.3	1.7		39.3	2.1		0.9	1.6
CBEI_RS04430|spoVAD	Cbei_0816	stage V sporulation protein AD	7.7	2.2		44.7	2.8		1.0	1.9
CBEI_RS04435|spoVAE	Cbei_0817	stage V sporulation protein AE	8.0	2.1		45.2	2.6		1.0	1.7
CBEI_RS08280|spoVE	Cbei_1583	stage V sporulation protein E	1.1	1.1		3.0	4.9		0.9	0.9
CBEI_RS00495|spoVT	Cbei_0088	stage V sporulation protein T	2.7	2.0		30.2	12.7		0.9	1.1
CBEI_RS02215	Cbei_0387	CotS family spore coat protein	0.7	1.2		3.7	5.8		0.8	2.6
CBEI_RS02205	Cbei_0385	CotS family spore coat protein	7.3	0.4		50.8	2.6		1.7	0.7
CBEI_RS02195	Cbei_0383	CotS family spore coat protein	22.1	0.6		186.4	4.3		1.6	0.7
CBEI_RS04600	Cbei_0850	CotS family spore coat protein	1.0	1.4		5.2	12.6		0.8	2.4
CBEI_RS10750|CotJA	Cbei_2071	spore coat associated protein CotJA	534.6	1.7		1552.8	35.1		3.1	0.7
CBEI_RS10745|CotJB	Cbei_2070	spore coat protein CotJB	416.8	1.4		1271.2	26.4		3.1	0.7
CBEI_RS10740	Cbei_2069	manganese catalase family protein (cotJC)	484.6	1.1		948.3	17.6		3.1	0.7
CBEI_RS12770	Cbei_2471	alpha/beta-type small acid-soluble spore protein	15.5	3.7		134.3	40.3		1.1	0.9
CBEI_RS02705	Cbei_0474	alpha/beta-type small acid-soluble spore protein	23.6	1.7		231.2	25.6		2.1	0.9
CBEI_RS16815	Cbei_3275	alpha/beta-type small acid-soluble spore protein	37.4	3.5		345.6	52.7		1.9	0.8
CBEI_RS16765	Cbei_3264	alpha/beta-type small acid-soluble spore protein	47.6	4.6		325.1	89.5		2.0	0.9
CBEI_RS07595	Cbei_1447	alpha/beta-type small acid-soluble spore protein	39.8	3.5		279.8	34.0		2.0	0.8
CBEI_RS12145	Cbei_2345	alpha/beta-type small acid-soluble spore protein	32.1	3.1		284.1	83.1		2.0	0.9
CBEI_RS15865	Cbei_3080	alpha/beta-type small acid-soluble spore protein	59.4	4.3		217.7	47.1		2.5	0.6
CBEI_RS16025	Cbei_3111	alpha/beta-type small acid-soluble spore protein	63.8	4.2		457.1	83.3		1.9	0.9
CBEI_RS16695	Cbei_3250	alpha/beta-type small acid-soluble spore protein	21.7	3.2		178.8	46.8		1.3	1.1

Lowest value 1 25<

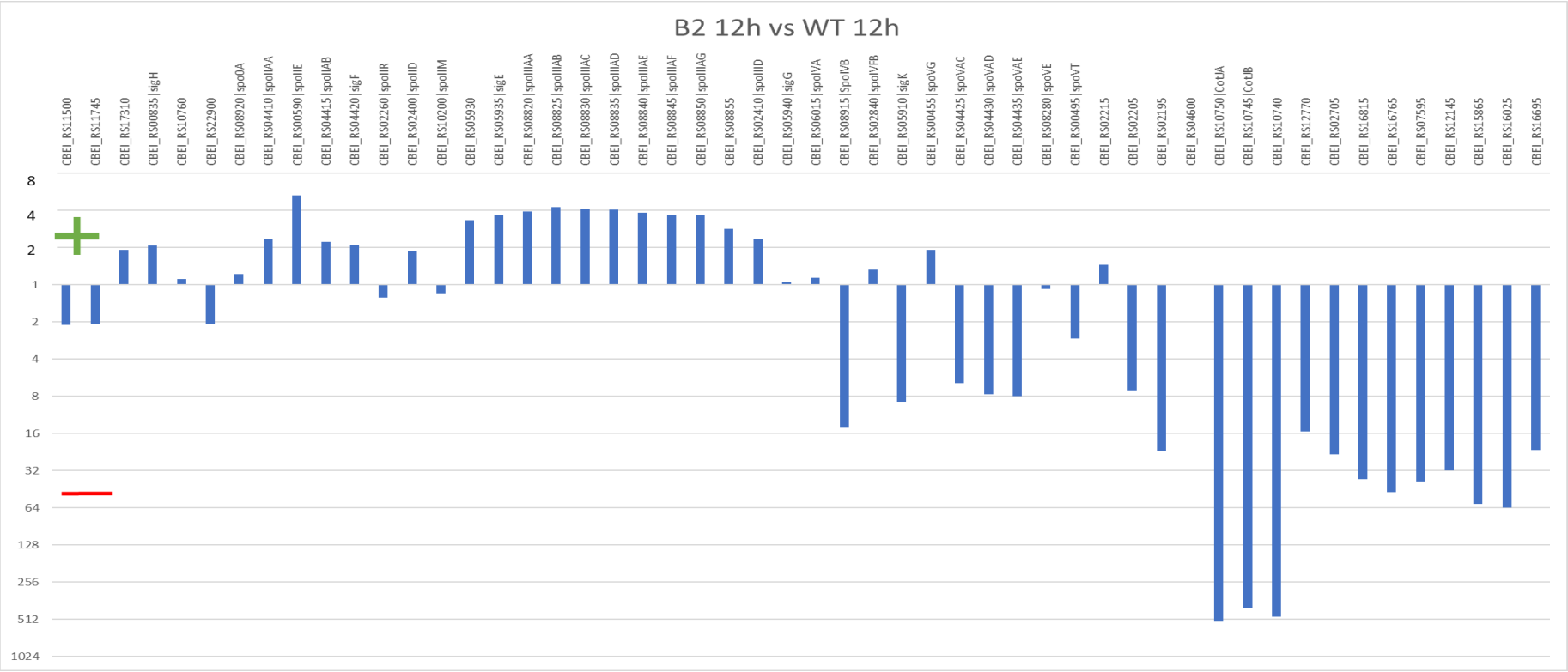


Figure 6.3 Change in key spore-formation gene expression in the *agrB2* mutant compared to the WT at 12 hours of growth.

The genes identified to be involved in *C. beijerinckii* sporulation were placed in order of early to late expression during the sporulation cycle, using Diallo et al. (2021) as a reference. The Fold change in gene expression, *agrB2* mutant versus WT, from the RNA-seq data was plotted for each gene. The green + indicated upregulation in the *agrB2* mutant compared to the WT whilst the red – sign represents downregulation in the mutant.

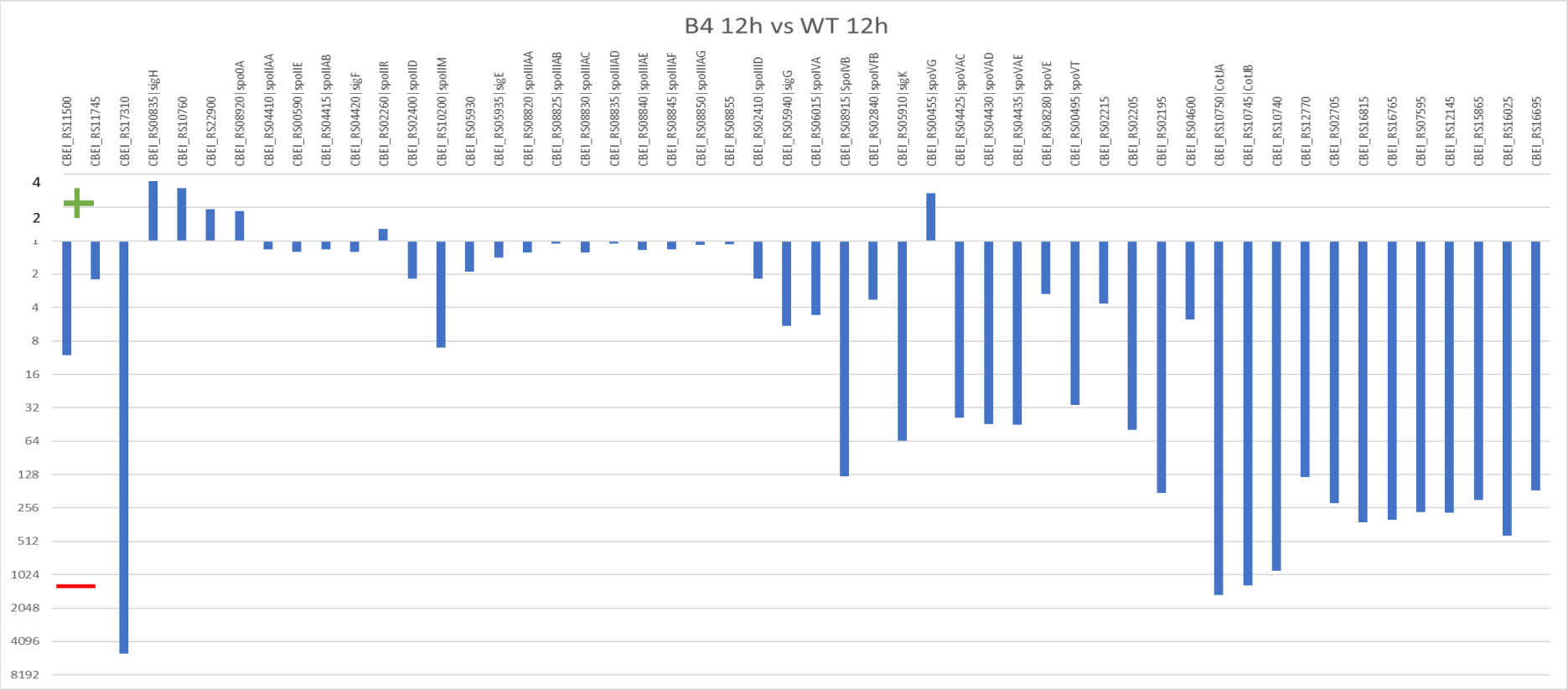


Figure 6.4 Change in key spore-formation gene expression in the *agrB4* mutant compared to the WT at 12 hours of growth.

The genes identified to be involved in *C. beijerinckii* sporulation were placed in order of expression during the sporulation cycle using Diallo et al. (2021) as a reference. The Fold change in gene expression, *agrB4* mutant versus WT, from the RNA-seq data was plotted for each gene. Key sporulation genes were ordered from early to late gene expression. The green + indicated upregulation in the *agrB4* mutant compared to the WT whilst the red – sign represents downregulation in the mutant.

6.2.4 Gene expression related to fermentation metabolism

The *agrB* KO mutants were shown to have a reduced level of solvent accumulation in **chapter 5**. For the *agrB2* and *agrB4* KO mutants the RNA-seq revealed common acid and solvent metabolism-related genes which were downregulated compared to the WT expression (see **Table 6.5**) however, in the *agrB5* KO mutant changes are not as evident and yet, solvents are also reduced, suggesting a different mechanism. **Figure 6.5** and **Figure 6.6** provide a visualisation of the differentially regulated genes in the *agrB* mutants compared to the WT which encode the enzymes involved in the acid and solvent metabolic pathways at 12 and 24 hours.

At 12 hours the key genes encoding solvent formation enzymes (**Figure 6.5**), such as, *adhE* (CBEI_RS01685) annotated as bifunctional acetaldehyde-CoA/alcohol dehydrogenase were downregulated in the *agrB2* mutant by 4-fold change and the *agrB4* mutant by 17-fold change compared to the WT. Other key genes involved in acid or solvent formation were either unchanged or upregulated compared to the WT. For example, the *pta* gene (CBEI_RS06155), encoding phosphate acetyltransferase was expressed at similar levels to the WT. Whilst the *ptb* gene (CBEI_RS01180) encoding phosphate butyryltransferase and the *buk* gene (CBEI_RS01185) encoding butyrate kinase being upregulated in the agrB4 mutant by 3-fold and 5-

fold change compared to the WT respectively, which contributes to butyrate formation.

By 24 hours the gene *adhE* (CBEI_RS01685) was still downregulated in both the *agrB2* and *agrB4* mutants, but only by 2-fold and 3-fold change compared to the WT respectively. The *thl* gene (CBEI_RS02340) encoding acetyl-CoA C-acetyltransferase was now being downregulated however, by 3-fold change in both the *agrB2* and *agrB4* mutant. Genes related to acid formation were still remaining at similar expression levels to that of the WT or upregulated in the case of *ptb* and *buk* genes in the *agrB4* mutant.

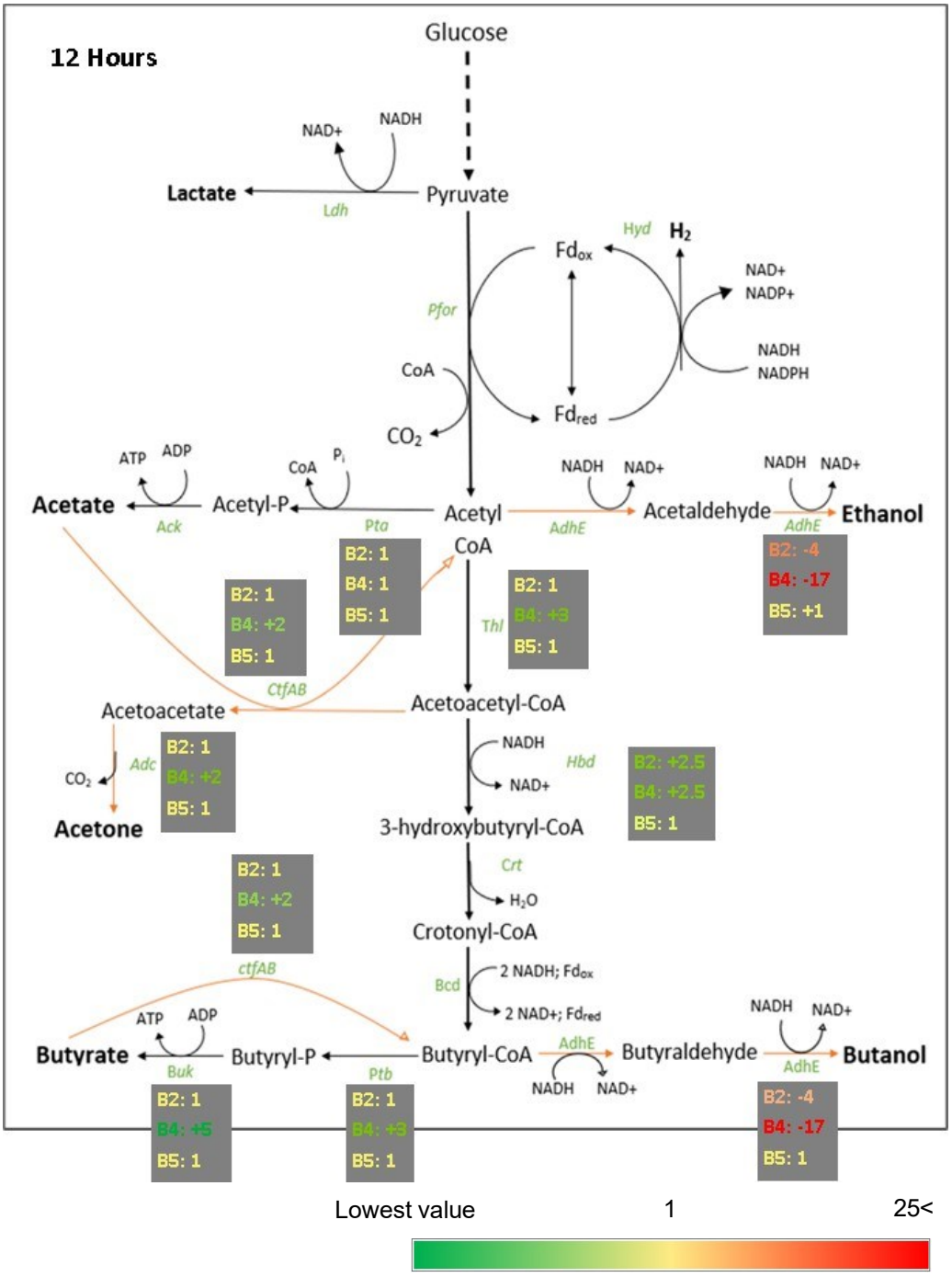


Figure 6.5 Fold-change in gene expression at 12 hours for the key genes involved in acid and/or solvent formation in the *C. beijerinckii* NCIMB 8052 *agrB* mutants.

The fold-change in gene expression compared to the WT has been shown next to key genes in the grey boxes for the *agrB2* mutant (B2), *agrB4* mutant (B4) and the *agrB5* mutant (B5). The degree of fold change in gene expression has been represented as a 3-colour heat map where values are expressed as the WT v *agrB* mutant gene expression. A fold change of <1 (green) represents an upregulated gene in the *agrB* mutant, a fold change of 1 (amber) represents no significant change in gene expression in the *agrB* mutant, and a fold change of >1 (red) represents a downregulated gene in the *agrB* mutant.

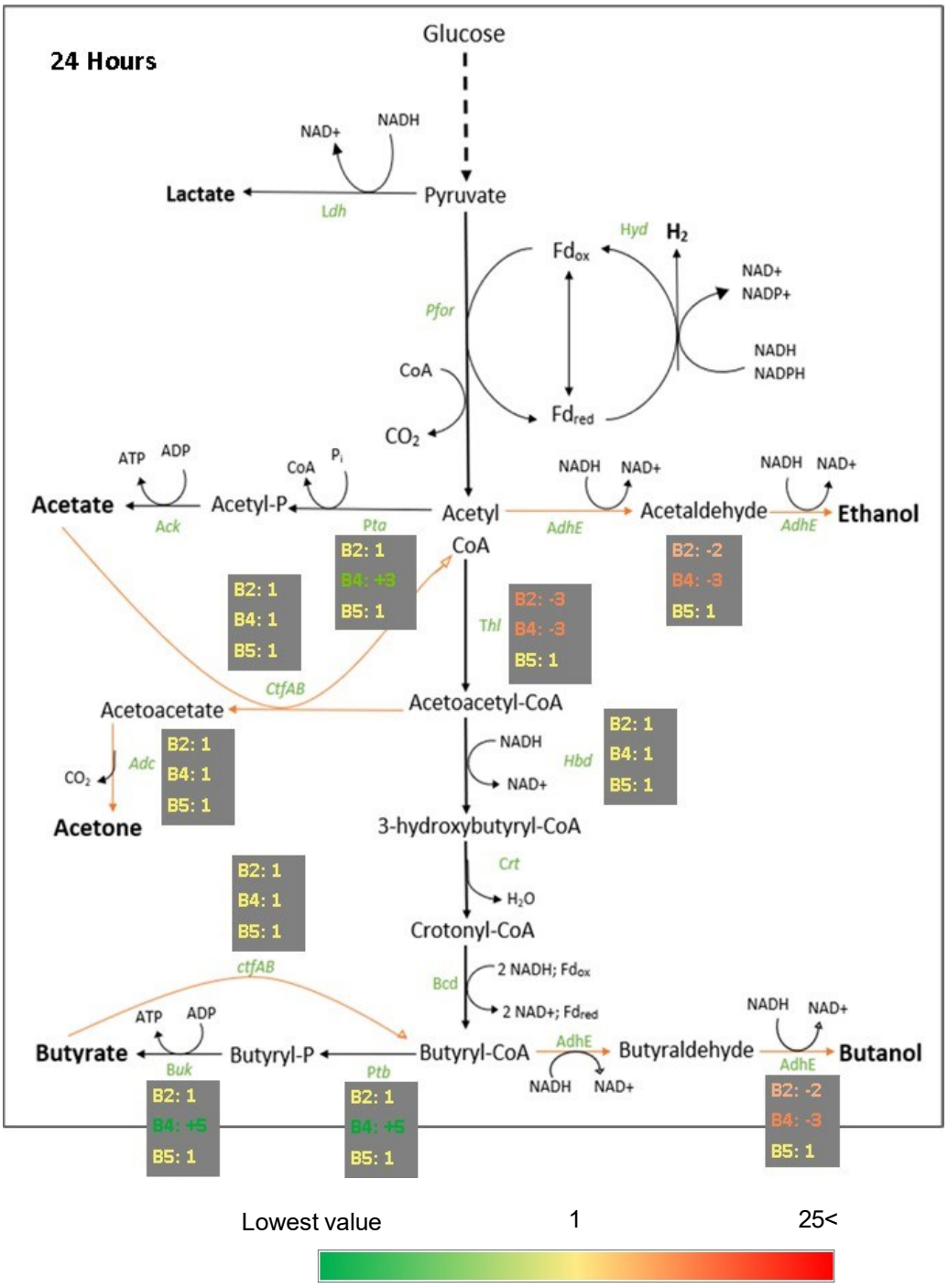


Figure 6.6 Fold-change in gene expression at 24 hours for the key genes involved in acid and/or solvent formation in the *C. beijerinckii* NCIMB 8052 *agrB* mutants.

The fold-change in gene expression compared to the WT has been shown next to key genes in the grey boxes for the *agrB2* mutant (B2), *agrB4* mutant (B4) and the *agrB5* mutant (B5). The degree of fold change in gene expression has been represented as a 3-colour heat map where values are expressed as the WT v *agrB* mutant gene expression. A fold change of <1 (green) represents an upregulated gene in the *agrB* mutant, a fold change of 1 (amber) represents no significant change in gene expression in the *agrB* mutant, and a fold change of >1 (red) represents a downregulated gene in the *agrB* mutant.

Various other genes potentially linked to fermentation metabolism were also affected. In the *agrB4* mutant there are two genes with a large downregulation compared to the WT by 12 hours, the putative acetyl-CoA C-acetyltransferase gene (CBEI_RS18610) downregulated by 1112-fold and the putative 2-oxoacid:ferredoxin oxidoreductase subunit beta gene (CBEI_RS18670) downregulated by 742-fold by 12 hours. This was not seen for the *agrB2* or *agrB5* mutants (**Table 6.5**).

In both the *agrB2* and *agrB4* mutants, genes affected included, but were not limited to, putative pyruvate:ferrodoxin oxidoreductase genes (*nifJ*) downregulated by around 6-fold in the *agrB2* mutant at 12 hours and 42-fold in the *agrB4* mutant by 12 hours. Putative acyl CoA dehydrogenase (CBEI_RS23185) was downregulated by 2-fold and 5-fold in the *agrB2* and *agrB4* mutants respectively at both time points, two putative butyrate kinases, *buk* (CBEI_RS20510 and CBEI_RS23520) and a putative bifunctional acetaldehyde-CoA/alcohol dehydrogenase (CBEI_RS01685) were also downregulated by 4-fold and 27-fold in the *agrB2* and *agrB4* mutants respectively by 12 hours. Other downregulated putative genes potentially related to fermentation metabolism include a putative aldehyde dehydrogenase family protein gene (CBEI_RS03680), putative iron-containing alcohol dehydrogenase genes (CBEI_RS08965 and CBEI_RS11295), acetolactate synthase large subunit gene (CBEI_RS14795), phosphophenolpyruvate synthase gene (CBEI_RS18040, CBEI_RS10710), pyruvate kinase gene (CBEI_RS4851) and dihydroxyacetone kinase phosphoryl donor subunit DhaM gene (CBEI_RS11140) (see appendix **section 8.3**).

Interestingly, some solvent-related genes were found to be upregulated in the *agrB* mutants compared to the WT. For example, the proposed gene *pflB* was upregulated by 4-fold in the *agrB2* mutant and *agrB4* mutant by 12 hours.

Fermentation metabolism-related genes downregulated in both *agrB4* and *agrB5* KO include the putative 2-oxoacid:ferrodoxin oxidoreductase subunit beta gene (CBEI_RS20685) downregulated 10 to 11-fold in the *agrB4* mutant at 12 and 24 hours, and 2-fold in the *agrB5* mutant at 24 hours. The putative 3-oxoacid CoA-transferase subunit B and subunit A genes (CBEI_RS13695) were downregulated between 11 and 15-fold at 12 and 24 hours in the *agrB4* mutant, and 3-fold at 24 hours in the *agrB5* mutant (**Table 6.5**).

Finally, potential fermentation metabolism-related genes found to be downregulated at 24 hours in all three *agrB* mutants included acyl CoA dehydrogenase gene (CBEI_RS14870) downregulated by 8, 6 and 3-fold in the *agrB2*, *agrB4* and *agrB5* mutants respectively at 24 hours. The CoA transferase subunit A gene (CBEI_RS13700) was also downregulated by 2 to 3-fold in the *agrB2* and *agrB5* mutants at 24 hours, whilst in the *agrB4* mutant being downregulated between 14 to 17-fold at both time points (**Table 6.5**). Other glycolytic and fermentation metabolism-associated genes which were downregulated include those encoding and putative L-Lactate permease (CBEI_RS14880) and lactate utilization (CBEI_RS07490) proteins (appendix **section 8.3**).

Table 6.5 The fold change in key solvent formation gene expression, for each WT v *agrB* mutant, at 12 hours of growth according to RNA-seq data.

The degree of fold change in gene expression has been visualised as a 3-colour heat map where values are represented as the WT v *agrB* mutant gene expression therefore, a value of <1 (green) represents an upregulated gene in the *agrB* mutant, a fold change of 1 (amber) represents no change in gene expression in the *agrB* mutant, and a fold change of >1 (red) represents a downregulated gene in the *agrB* mutant.

Acetyl coA formation genes			WT 12h vs B2 12h	WT 24h vs B2 24h	WT 12h vs B4 12h.	WT 24h vs B4 24h.	WT 12h vs B5 12h.	WT 24h vs B5 24h.
pflB	Cbei_1009	formate C-acetyltransferase	0.5	0.9	0.4	0.2	0.8	0.7
pflB	Cbei_1011	formate C-acetyltransferase	0.3	0.5	0.3	0.6	0.8	0.7
nifJ	Cbei_1458	pyruvate:ferredoxin (flavodoxin) oxidoreductase	5.8	2.5	42.0	26.7	0.8	1.8
nifJ	Cbei_1853	pyruvate:ferredoxin (flavodoxin) oxidoreductase	6.5	0.8	41.3	0.4	0.9	1.3
nifJ	Cbei_4318	pyruvate:ferredoxin (flavodoxin) oxidoreductase	0.5	0.8	0.1	0.5	0.7	1.2
CBEI_RS18670	Cbei_3642	2-oxoacid:ferredoxin oxidoreductase subunit beta	0.8	0.3	741.7	2979.5	1.2	0.4
CBEI_RS20685	Cbei_4041	2-oxoacid:ferredoxin oxidoreductase subunit beta	1.7	1.7	9.6	10.6	0.8	2.2
CBEI_RS20690	Cbei_4042	2-oxoacid:acceptor oxidoreductase subunit alpha	1.6	2.0	9.5	11.7	0.9	2.0
CBEI_RS19560	Cbei_3820	3-hydroxyacyl-CoA dehydrogenase family protein	0.8	0.5	0.7	0.3	0.7	0.8
Butyryl coA formation genes								
CBEI_RS02340	Cbei_0411	acetyl-CoA C-acetyltransferase (thl)	0.8	2.7	0.3	3.0	0.8	1.3
CBEI_RS18610	Cbei_3630	acetyl-CoA C-acetyltransferase	0.6	0.3	1111.9	7711.8	0.9	0.6
CBEI_RS01770	Cbei_0322	acyl-CoA dehydrogenase	0.5	0.6	0.4	0.7	0.8	1.2
CBEI_RS01785	Cbei_0325	3-hydroxybutyryl-CoA dehydrogenase (hbd)	0.4	0.6	0.4	0.8	0.7	1.2
CBEI_RS10565	Cbei_2035	acyl-CoA dehydrogenase	0.5	0.6	0.4	0.7	0.7	1.2
CBEI_RS14870	Cbei_2883	acyl-CoA dehydrogenase	1.4	8.3	0.3	5.9	0.8	2.6
CBEI_RS23185	Cbei_4542	acyl-CoA dehydrogenase family protein	2.1	2.4	4.6	4.6	0.9	1.1
CBEI_RS13700	Cbei_2654	CoA transferase subunit A	1.9	2.0	14.6	17.1	1.0	2.6
CBEI_RS19630	Cbei_3833	3-oxoacid CoA-transferase subunit A (ctfA)	1.1	1.2	0.4	0.8	1.0	1.8
CBEI_RS19635	Cbei_3834	3-oxoacid CoA-transferase subunit B (ctfB)	1.2	1.2	0.5	0.8	1.1	1.8
CBEI_RS10825	Cbei_2086	trans-2-enoyl-CoA reductase family protein	1.4	1.4	1.9	0.8	0.8	1.2
Acetate formation genes								
CBEI_RS06155|pta	Cbei_1164	phosphate acetyltransferase	1.0	0.6	0.6	0.3	1.1	0.7
CBEI_RS13695	Cbei_2653	3-oxoacid CoA-transferase subunit B	1.7	1.7	11.3	15.4	0.8	2.4
CBEI_RS03955	Cbei_0727	aldehyde dehydrogenase family protein	1.7	0.4	1.4	0.2	1.1	1.1
CBEI_RS03965	Cbei_0729	aldehyde dehydrogenase	1.2	0.4	0.2	0.3	1.9	0.5
CBEI_RS10135	Cbei_1953	aldehyde dehydrogenase	0.8	1.2	0.5	0.9	0.8	1.1
Butyrate formation genes								
CBEI_RS01180|ptb	Cbei_0203	phosphate butyryltransferase	0.7	0.9	0.3	0.2	0.6	1.0
CBEI_RS01185|buk	Cbei_0204	butyrate kinase (buk)	0.6	0.7	0.2	0.2	0.6	0.9
CBEI_RS23520|buk	Cbei_4609	butyrate kinase	8.5	2.6	26.7	6.9	1.1	1.6
CBEI_RS20510|buk	Cbei_4006	butyrate kinase	4.3	2.3	8.7	6.0	1.1	0.8
Solventogenesis genes								
CBEI_RS19640	Cbei_3835	acetoacetate decarboxylase (adc)	1.1	1.1	0.5	0.7	1.1	1.8
CBEI_RS01685|adhE	Cbei_0305	bifunctional acetaldehyde-CoA/alcohol dehydrogenase	4.1	2.1	16.7	2.7	0.8	1.1
CBEI_RS19640	Cbei_3835	acetoacetate decarboxylase (adc)	1.1	1.1	0.5	0.7	1.1	1.8
CBEI_RS12520	Cbei_2421	iron-containing alcohol dehydrogenase	0.7	1.0	0.6	0.8	0.9	0.9

6.2.5 Gene expression related to glucose consumption

The *agrB* mutants and WT were all grown in media containing glucose. It was therefore of interest to observe if sugar-uptake related genes were differentially expressed. Glucose-uptake related genes (see **Table 6.6**) were also downregulated in the *agrB*2 and *agrB4* mutants, particularly at 12 hours. Some key genes glucose-uptake genes include N-acetylglucosamine-specific PTS transporter subunit IIBC (*nagE/*BEI_RS23135) and PTS glucose transporter subunit IIA gene (CBEI_RS23140) which show a 2 to 5-fold downregulation in the *agrB2* and *agrB4* mutants at 24 hours. Other possible glucose-uptake genes, include, but not limited to, putative phosphotransferase system (PTS) sugar transporter subunits such as the IIA subunit genes (CBEI_RS14155, CBEI_RS24005, CBEI_RS24495 and (CBEI_RS25665), IIB subunit genes (CBEI_RS23895 and CBEI_RS14150) and IIC subunit genes (CBEI_RS23890 and CBEI_RS14145) (Appendix **section 8.3**).

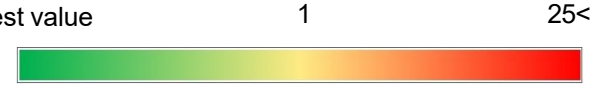
Various glycolysis-related genes were downregulated in the *agrB2*, *agrB4* and *agrB5* mutants by 24 hours respectively with key genes being those for fructose bisphosphate aldolase (CBEI_RS15665) downregulated by 3, 10 and 2-fold, triose-phosphate isomerase (CBEI_RS03315) downregulated by 4, 6 and 2-fold, type I glyceraldehyde-3-phosphate dehydrogenase (CBEI_RS03305)

downregulated by 6, 11 and 2-fold, and phosphoglycerate kinase (CBEI_RS03310) downregulated by 6, 11 and 3-fold. Furthermore, in the *agrB2* and *agrB4* mutants, the putative *gpmA* gene, encoding 2,2C3-diphosphoglycerate-dependent phosphoglycerate mutase, was hugely downregulated by 58.6 and 184-fold change respectively by 12 hours. A notable downregulated gene annotated as a ROK family protein is CBEI_RS18045, reduced by 258-fold change by 12 hours and 19-fold by 24 hours in the *agrB2* mutant compared to the WT (**Table 6.6**).

Table 6.6 The Fold change in key glucose-uptake and glycolysis gene expression, for each WT v *agrB* mutant, at 12 and 24 hours of growth according to RNA-seq data.

The degree of fold change in gene expression has been visualised as a 3-colour heat map where values are represented as the WT v *agrB* mutant gene expression therefore, a value of <1 (green) represents an upregulated gene in the *agrB* mutant, a fold change of 1 (amber) represents no change in gene expression in the *agrB* mutant, and a fold change of >1 (red) represents a downregulated gene in the *agrB* mutant.

Glucose uptake genes			WT 12h vs B2 12h	WT 24h vs B2 24h		WT 12h vs B4 12h.	WT 24h vs B4 24h.		WT 12h vs B5 12h.	WT 24h vs B5 24h.
CBEI_RS19300	Cbei_3768	PTS glucose transporter subunit IIA	0.4	0.6		0.5	0.3		0.8	0.9
CBEI_RS24495	Cbei_4804	PTS glucose transporter subunit IIA	4.2	3.7		9.7	4.7		1.8	0.9
CBEI_RS24005	Cbei_4706	PTS glucose transporter subunit IIA	17.3	0.4		12.3	2.6		1.6	0.5
CBEI_RS25485	Cbei_4982	PTS glucose transporter subunit IIA	1.1	1.0		0.7	0.7		1.0	0.8
CBEI_RS25490	Cbei_4983	PTS transporter subunit EIIC	2.2	1.8		2.1	1.5		1.0	1.1
CBEI_RS25665	Cbei_5017	PTS glucose transporter subunit IIA	2.3	1.3		5.2	3.0		1.4	1.2
CBEI_RS25675	Cbei_5019	PTS transporter subunit EIIC	1.6	1.8		2.6	2.0		0.7	1.6
CBEI_RS04080	Cbei_0751	PTS transporter subunit IIABC	0.9	1.5		0.5	2.2		0.9	1.1
CBEI_RS23135|nagE	Cbei_4532	N-acetylglucosamine-specific PTS transporter subunit IIBC	0.9	3.2		0.4	2.0		0.6	1.6
CBEI_RS23140	Cbei_4533	PTS glucose transporter subunit IIA	0.9	4.1		0.5	5.1		0.9	1.3
Glycolysis genes										
CBEI_RS18045	Cbei_3517	ROK family protein	2.1	0.4		258.1	18.8		1.2	0.7
CBEI_RS19515	Cbei_3810	ROK family protein	1.2	0.4		2.2	0.1		0.9	1.0
CBEI_RS23310	Cbei_4567	ROK family protein	1.3	0.4		1.2	0.3		0.9	1.4
CBEI_RS24965	Cbei_4895	ROK family protein (glcK)	0.9	1.2		0.8	1.2		0.8	1.3
CBEI_RS03240	Cbei_0584	6-phosphofructokinase (pfk)	1.3	0.7		0.5	0.5		0.9	0.8
CBEI_RS05325	Cbei_0998	6-phosphofructokinase	0.3	0.9		0.9	2.1		0.9	1.0
CBEI_RS24735	Cbei_4852	6-phosphofructokinase	1.7	3.3		0.6	5.7		0.9	1.4
CBEI_RS01865	Cbei_0341	glucose-6-phosphate isomerase (pgi)	0.5	2.5		0.4	2.5		0.8	1.3
CBEI_RS09875|fba	Cbei_1903	class II fructose-1%2C6-bisphosphate aldolase	0.6	0.6		0.3	0.3		0.9	1.0
CBEI_RS03305|gap	Cbei_0597	type I glyceraldehyde-3-phosphate dehydrogenase	1.4	5.5		0.4	10.5		0.8	2.3
CBEI_RS05255	Cbei_0983	phospho-sugar mutase	0.4	0.6		0.3	0.3		0.9	0.7
CBEI_RS22560	Cbei_4415	HAD family phosphatase	0.4	0.8		0.4	0.7		0.9	0.7
CBEI_RS03330	Cbei_0602	phosphopyruvate hydratase	1.5	0.7		3.6	0.1		0.8	1.3
CBEI_RS09165|gpmA	Cbei_1759	2%2C3-diphosphoglycerate-dependent phosphoglycerate mutase	58.6	1.8		184.2	46.1		2.4	0.7
CBEI_RS03320	Cbei_0600	2%2C3-bisphosphoglycerate-independent phosphoglycerate mutase	0.9	1.4		0.4	2.1		1.0	0.7
CBEI_RS24730|pyk	Cbei_0485	pyruvate kinase	1.0	1.0		0.5	1.0		0.9	1.2
CBEI_RS07420	Cbei_1412	pyruvate kinase	1.2	1.1		6.6	4.3		0.9	2.1
CBEI_RS24730	Cbei_4851	pyruvate kinase	1.8	3.1		0.6	6.0		0.9	1.5
CBEI_RS04595|ppdK	Cbei_0849	pyruvate%2C phosphate dikinase	0.5	2.2		1.2	3.3		0.8	1.4
CBEI_RS23065|argJ	Cbei_4518	bifunctional glutamate N-acetyltransferase/amino-acid acetyltransferase ArgJ	3.2	1.7		2.6	2.0		1.0	1.2
CBEI_RS22810	Cbei_4465	galactose mutarotase	1.9	0.5		1.3	0.6		0.9	1.1
CBEI_RS12750	Cbei_2467	fructose-1%2C6-bisphosphatase	0.7	1.0		0.8	0.7		1.0	0.9
CBEI_RS23180	Cbei_4541	fructose-1%2C6-bisphosphatase	0.9	0.9		1.1	0.8		0.9	1.2
CBEI_RS15665	Cbei_3039	fructose bisphosphate aldolase	1.9	2.8		9.2	10.3		0.8	2.1
CBEI_RS03315	Cbei_0599	triose-phosphate isomerase	1.8	4.0		0.5	6.1		1.0	2.2
CBEI_RS03305	Cbei_0597	type I glyceraldehyde-3-phosphate dehydrogenase	1.4	5.5		0.4	10.5		0.8	2.3
CBEI_RS03310	Cbei_0598	phosphoglycerate kinase	1.7	5.6		0.4	10.9		0.9	2.6
CBEI_RS11815	Cbei_2282	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	1.1	4.7		2.5	9.6		0.7	1.1
CBEI_RS13280	Cbei_2572	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	1.9	1.4		3.7	1.0		1.0	0.8



6.2.6 Other highly differentially regulated genes in *agrB* mutants

There were also other genes in which regulation was affected by the *agrB* mutations. The **Table 6.1, Table 6.2** and **Table 6.3** show the top 20 most upregulated and downregulated genes listed in each *agrB* mutant at 12 and 24 hours. Some notable examples being the gene cluster encoding the tryptophan synthase pathway genes (CBEI_RS09105 to CBEI_RS09130), these genes being upregulated in the *agrB2* mutant when compared to the WT by between 19 to 31-fold and upregulated between 5-fold to 3-fold in the *agrB4* mutant at 12 and 24 hours. Tryptophan is an amino acid that acts as an important nitrogen source for many bacteria (Gao *et al.*, 2020).

In the *agrB2* mutant key gene encoding nitrogen-fixing enzymes, nitrogenase iron protein (*nifH)* and nitrogenase molybdenum-iron protein subunit beta (*nifK),* were highly upregulated at 12 hours by 8-fold change compared to the WT at 12 hours (**Table 6.1)**. In contrast, the *nifH* and *nifK* genes were highly downregulated in the *agrB4* mutant by 62-fold compared to the WT at 12 hours. In the *agrB2* mutant, genes encoding the synthesis of the iron–molybdenum cofactor (FeMoco) of nitrogenase, *nifE* and *nifB* (Chen, Toth and Kasap, 2001), were upregulated by 4 to 6-fold change at 12 hours whilst the *nifV* remained at similar levels to the WT. Meanwhile, in the *agrB4* mutant at 12 hours the *nifE* and *nifB* genes were both downregulated by 38-fold compared to the WT. AT 24 hours, all the nitrogen-fixing related genes in the *agrB2* mutant had similar levels of gene expression to that of the WT, whilst in the *agrB4* mutant the genes *nifH*, *nifK*, *nifB* and *nifE* were all still downregulated compared to the WT by 6 to 9-fold change. All these listed nitrogen-fixing genes remained at similar levels to the WT gene expression at both 12 and 24 hours in the *agrB5* mutant. In all three *agrB* mutants analysed in this study a group of nitrogenase and nitrogenase accessory genes (CBEI_RS10370 to CBEI_RS10430) were affected in terms of regulation. For the *agrB2* mutant, genes in this region tended to be upregulated between 2-fold to 9-fold change at 12 hours. However, in the *agrB4* and *agrB5* mutants, these genes tended to be downregulated at 12 hours between 2 and 5-fold and 47-fold for the CBEI_RS10430 in the *agrB4* mutant (see appendix **section 8.3**).

6.2.7 Quorum-sensing gene expression

It was of interest to see how inactivation of these *agrB* genes affected the expression of the respective encoding *agr* gene cluster, and whether there was evidence for cross regulation. It was observed that, after 12 and 24 hours, other putative Agr systems appeared to be differentially regulated in the *agrB* mutants above (**Table 6.7**). For instance, at 12 hours, in the *agrB4* mutant the *agrBDC* genes from systems 2, 3 and 4, and the *agrBDCA* genes from systems 1, 5 and 6 were upregulated between 2.5-fold to 12.5-fold. Similarly, in the *agrB2* mutant, *agrA1* gene, *agrBDC* genes from system 2, *agrBD* genes from systems 3, *agrC* genes from system 4, *agrBDCA* genes from systems 5 and 6, were slightly upregulated between 2-fold and 4-fold. Knocking out the *agrB5* gene, only appeared to have a significant impression on *agrBDCA5*, downregulated by 4-fold to 9-fold, and *agrD*4 downregulated by 5.5-fold change compared to the WT at 12 hours.

By 24 hours the RNA-seq results showed a change in Agr system regulation also. For instance, in the *agrB2* mutant the *agrBDC* genes from system 4 and the *agrBD* genes from system 5 were now downregulated between 3-fold to 4-fold change compared to the WT. However, the *agrA5* gene (CBEI_RS20310) and *agrBDCA* genes from system 6 were still being upregulated by 3-fold to 5-fold change compared to the WT. The Agr3 system was no longer being differentially regulated compared to the WT by 24 hours in this *agrB*

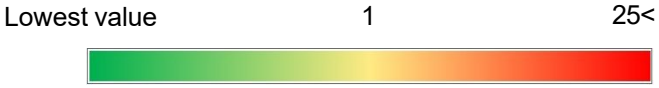
mutant. In the *agrB4* mutant, *agrBD2*, *agrBDC4 and agrB5* genes were being downregulated between 2-fold to 6-fold change compared to the WT. Meanwhile, *agrCA1*, *agrCA5* and *agrBDCA6* genes were all still being upregulated, between 2-fold to 11-fold change from the WT at 24 hours. In the *agrB5* mutant there was still only a significant effect on the Agr5 and Agr4 system, at 24 hours the *agrBDC4* genes were downregulated by 2-fold to 5-fold change compared to the WT.

Since LytTR-family regulators, such as AgrA, are those often involved in the expression of quorum sensing systems (Nikolskaya and Galperin, 2002) it was of interest to investigate how such genes, not directly part of the already annotated Agr systems, were affected in the agrB mutants. In the agrB4 mutant, the LytTR family transcriptional regulator genes CBEI_RS18730 and CBEI_RS17975 were downregulated between 2290 to 2530-Fold change and 839 to 903-Fold change at both 12 and 24 hours respectively. These were not significantly differentially expressed compared to the WT in the *agrB2* or *agrB5* mutants. Another LytTR family protein (CBEI_RS03110) was upregulated in the *agrB2* and *agrB4* mutant at 24 hours by 2 to 3-fold change compared to the WT (Appendix **section 8.3.**

Table 6.7 The Fold change in *agr* gene expression, for each WT v *agrB* mutant, at 12 and 24 hours of growth according to RNA-seq data.

The degree of fold change in gene expression has been visualised as a 3-colour heat map where values are represented as the WT v *agrB* mutant gene expression therefore, a value of <1 (green) represents an upregulated gene in the *agrB* mutant, a fold change of 1 (amber) represents no change in gene expression in the *agrB* mutant, and a fold change of >1 (red) represents a downregulated gene in the *agrB* mutant.

Agr systems			WT 12h vs B2 12h	WT 24h vs B2 24h		WT 12h vs B4 12h.	WT 24h vs B4 24h.		WT 12h vs B5 12h.	WT 24h vs B5 24h.
CBEI_RS26420	Cbei_0269	accessory gene regulator B family protein (agrB1)	0.8	1.2		0.3	0.8		1.4	1.2
CBEI_RS26850		Cyclic lactone autoinducer peptide (agrD1)	0.9	1.9		0.3	1.3		1.7	1.2
CBEI_RS01490	Cbei_0268	GHKL domain-containing protein (putative agrC1)	0.8	1.0		0.3	0.4		1.1	1.1
CBEI_RS01485	Cbei_0267	LytTR family DNA-binding domain-containing protein (putative agrA1)	0.5	0.7		0.1	0.2		0.7	0.9
CBEI_RS03605	Cbei_0658	accessory gene regulator B family protein (agrB2)	0.4	1.1		0.1	3.4		0.8	0.5
AgrD2		Putative AgrD2	0.3	0.9		0.1	2.3		1.9	0.5
CBEI_RS03610	Cbei_0659	ATP-binding protein (putative agrC2)	0.5	0.6		0.1	1.4		0.8	0.3
CBEI_RS03615	Cbei_0660	accessory gene regulator B family protein (agrB3)	0.4	1.4		0.1	1.1		1.0	0.9
AgrD3		Putative AgrD3	0.3	1.7		0.1	1.4		0.9	0.9
CBEI_RS16315	Cbei_3171	accessory gene regulator B family protein (agrB4)	0.6	3.6		0.1	4.9		1.6	3.7
CBEI_RS26650		cyclic lactone autoinducer peptide (agrD4)	0.7	3.3		0.3	5.7		5.5	2.4
CBEI_RS16305	Cbei_3169	PAS domain-containing sensor histidine kinase (putative agrC4)	0.4	3.2		0.2	3.4		1.4	5.1
CBEI_RS20295	Cbei_3964	CHAP domain-containing protein	1.1	3.6		0.4	5.1		38.2	14.9
CBEI_RS20300	Cbei_3965	accessory gene regulator B family protein (agrB5)	0.4	4.1		0.4	2.1		8.7	8.2
CBEI_RS26925		cyclic lactone autoinducer peptide (agrD5)	0.5	3.0		0.4	1.3		7.0	8.1
CBEI_RS20305	Cbei_3966	GHKL domain-containing protein (putative agrC5)	0.4	0.9		0.3	0.3		4.3	1.7
CBEI_RS20310	Cbei_3967	LytTR family DNA-binding domain-containing protein (putative agrA5)	0.4	0.3		0.1	0.1		1.1	0.5
CBEI_RS23365	Cbei_4578	accessory gene regulator ArgB-like protein (agrB6)	0.3	0.3		0.1	0.2		0.7	0.6
CBEI_RS26810		cyclic lactone autoinducer peptide (agrD6)	0.4	0.2		0.2	0.1		0.9	0.6
CBEI_RS23370	Cbei_4579	GHKL domain-containing protein (putative agrC6)	0.6	0.3		0.2	0.2		1.1	0.5
CBEI_RS23375	Cbei_4580	LytTR family DNA-binding domain-containing protein (putative agrA6)	0.5	0.3		0.2	0.2		0.9	0.5



6.2.8 Reverse Transcriptase qPCR

Reverse transcriptase qPCR was used to independently confirm the relative changes in transcript levels according to RNA-seq, both in terms of direction and order of magnitude. For this, three replicates were used for each *agrB* mutant and the WT, when testing each gene and the average taken to calculate the quantity. The *pepT* housekeeping gene (CBEI_RS12555) was used as an endogenous control as this gene showed constant expression levels to the WT in all samples, this has also been used in other transcriptomic studies on *C. beijerinckii* NCIMB 8052 to normalize gene expression against (Wang et al., 2012). The Pffafl method (Pfaffl, 2001), which takes into account different primer amplification efficiencies, was used for relative quantification of gene expression from the qPCR data. As demonstrated by **Figure 6.7**, when compared to the RNA-seq data, the qPCR data was generally highly similar when calculating the relative gene expression in the *agrB* mutants compared to the WT, with exceptions being almost entirely limited to transcripts showing <2-fold changes. This suggests that the RNA-seq data is highly reliable.

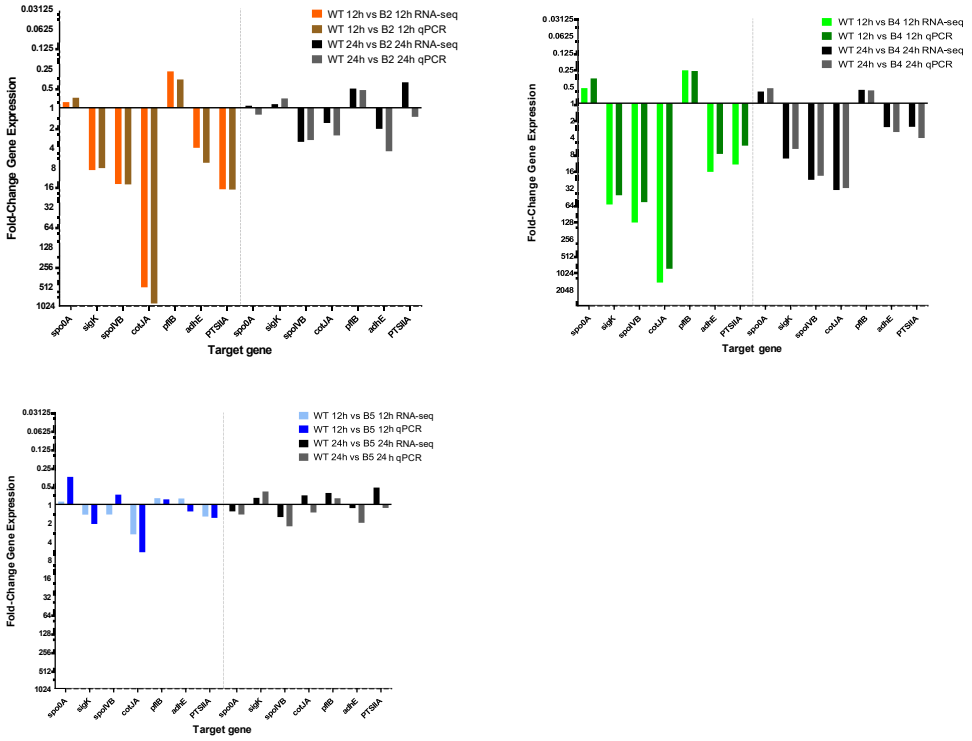


Figure 6.7 qPCR data alongside RNA-seq data showing fold change in gene expression compared to the WT at 12 and 24 hours in the *C. beijerinckii* NCIMB 8052 *agrB* mutants.

The genes used in the qPCR analysis were *spo0A* (CBEI_RS08920), *sigK* (CBEI_RS05910), *spoIVB* (CBEI_RS08915), *cotJA* (CBEI_RS10750), *ptflB* (CBEI_RS05390), *adhE* (CBEI_RS01685) and PTSIIA (CBEI_RS24005). The different *agrB* mutants were shown, A, *agrB2*, B, *agrB4*, and C, *agrB5*.

6.3 Discussion

The aims of this study were to extract total RNA from the *C. beijerinckii* NCIMB 8052 WT and *agrB* single KO mutants at both 12 hours and 24 hours. The total RNA was sent for RNA-seq which was observed for specific genes which were differentially regulated in relation to sporulation, solventogenesis and sugar-uptake genes or any other genes showing interesting or unexpected changes. The relative changes in genes expression was also checked for selected genes by qPCR. These two independent methods showed matching gene expression levels for selected genes and so the genome wide RNA-seq was seen as sufficiently reliable.

6.3.1 Sporulation and granulose related genes

The number of heat-resistant endospores produced in the *agrB2* and *agrB4* mutants was shown to be drastically reduced when compared to the WT (**Chapter 5**). In this study it was observed that there were common sporulation-related genes differentially regulated in these two *agrB* mutants. Of particular interest was when the key sporulation-related genes of *C. beijerinckii* NCIMB 8052 were placed in order of early to late gene expression, as shown in **Figure 6.3** and **Figure 6.4**, there appeared a very clear trend. Later stage sporulation genes were

often strongly downregulated in these mutants, whilst earlier stage sporulation genes tended to be of similar expression levels to the WT or in the case for the *agrB2* mutant, mildly upregulated. This was with the exception of CBEI_RS11500 and CBEI_RS17310 in the *agrB4* mutant, the latter of which is annotated AbrB/MazE/SpoVT family DNA-binding domain-containing protein, listed to encode an AbrB regulator (Wang *et al.*, 2012; Seo *et al.*, 2017). However, this study hardly observed any major consequences for most early sporulation-related genes in the *agrB4* mutant. Given that there was a downregulation of 5342-fold change at 12 hours for the gene CBEI_RS17310 it may be inaccurate to assume this is an *abrB* regulator gene. The study by Wang *et al.* (2012), showing RNA-seq data across the course of batch fermentation, indicates that in the *C. beijerinckii* NCIMB 8052 WT the *sigK* gene started to show increased expression at time point B (6 h) and was strongly expressed from time point C onwards (12 h), whereas CBEI_RS17310 regulation peaked at B (6 h) and then declined, a pattern compatible with a role in triggering *spoIVB* activation. Most early sporulation genes are already transcribed at time point A (4 h) so CBEI_RS17310 does not fit in the early sporulation gene expression trend. It may therefore be that this gene is encoding a SpoVT version that acts in later stage sporulation.

In both mutants however there was a clear point at which there was a 'flip' in this gene expression, and this occurred around the point of *sigK* (CBEI_RS05910) and *spoIVB* (CBEI_RS08915) expression in the sporulation cascade. The latest sporulation sigma-factor (*sigK*) is

imperative for the transcription of genes specific to the mother cell. These genes include spore cortex, spore coat and germination genes (Li, Di Donato and Piggot, 2004; Sun *et al.*, 2021). In *B. subtilis* SigK maturation is positively regulated by SpoIVB (Sun *et al.*, 2021). In B. subtillis the maturation of SigK in the mother cell occurs after the engulfment stage of sporulation in response to the SpoIVB signalling protein produced by the forespore. The SpoIVB peptidase signals by passing through a membrane and activating SpoIVFB, which then activates SigK (Resnekov and Losick, 1998; Hoa, Brannigan and Cutting, 2002). The *spoIVFB* gene was found to be encoded in most clostridial species. The product of CBEI_RS02840/Cbei_0501 (annotated as peptidase M50) in *C. beijerinckii* was found to exhibit 62% amino acid identity (169 out of 273) to SpoIVFB from *C. acetobutylicum* by BLASTP (Wang *et al.*, 2012). This putative *spoIVFB* gene only appeared to be downregulated in the *agrB4* mutant at 12 and 24 hours and only slightly downregulated in the *agrB5* mutant at 24 hours. In this study the *spoIVB* gene was downregulated in the *agrB2* and *agrB4* single KO strains by 15 and 132-fold change respectively at 12 hours. This would therefore contribute to the reduced expression of *sigK* gene observed in these mutants at 12 hours.

This downregulation of gene expression also occurs in those genes involved in the sporulation stages IV onwards, likely dependent on the downregulation of the *sigK* gene. These stages are responsible for, stage IV cortex formation to separate the two membranes between the

mother cell and forespore. Stage V spore coat proteins are deposited on the outer membrane of the forespore and stages VI and VII spore maturation and release. This study showed a downfold change in spore coat associated genes by 12 hours such as *cotJA* and *cotJB* in all three single *agrB* mutants. When observed under TEM and high-resolution microscopy, the *agrB2*, *agrB4 and agrB5* mutants were observed to still be producing forespores (**chapter 5**). However, when testing heat-resistant endospore formation in **chapter 5**, the *agrB2* and *agrB4* mutants had a significant reduction in heat-resistant CFU/ml. Along with the observations here where RNA-seq and qPCR data shows later stage sporulation genes, required for spore maturation and spore coat proteins, being largely downregulated in the *agrB2* and *agrB4* mutants. This supports the idea that AgrB2 and AgrB4 systems are likely involved in the regulation of later stage sporulation via the downregulation of *spoIVB* and *sigK* gene expression.

During the phenotypic characterisation of *agrB* mutants in **chapter 5** granulose accumulation in the *agrB* mutants appeared to be similar to that of the WT. Since granulose is known to be linked to the onset of sporulation, granulose-related genes were analysed in the RNA-seq data. Here it was observed that by 12 hours, genes involved in granulose biosynthesis (*glgA*, *glgB*, *glgD*) were upregulated in the *agrB2* and *agrB4* mutants, whilst genes involved in granulose degradation (*glgP*) were downregulated compared to WT gene expression at 12 hours for the *agrB2* mutant and 24 hours for the *agrB4* mutant. A study by Diallo *et al.* (2020) observed that in a *C. beijerinckii*

NCIMB 8052 *spoIIE* mutant, genes involved in granulose formation were upregulated, whilst genes involved in granulose degradation were unchanged when observing RNA-Seq data. These results contrast with the effect of the *spo0A* mutation which results in a granulose negative phenotype (Humphreys *et al.*, 2023). The observed changes here in the *agrB* single mutants suggest that granulose may be enhanced in the mutants, although this was not confirmed at the phenotypic level. Together these findings suggest that the generated *agrB* mutations are unlikely to strongly affect *spo0A* expression and phosphorylation, which are in line with the obtained RNAseq data.

6.3.2 Acidogenesis and solventogenesis related genes

In **chapter 5** it was suggested that the switch from acidogenesis to solventogenesis was likely differentially regulated in the *agrB* mutants. In all of the *agrB* single mutants there was higher butyrate and lower butanol concentrations compared to the WT at 24 hours. This trend was also seen for acetate and acetone concentrations respectively in the *agrB2* and *agrB4* mutants compared to the WT. However, unlike for sporulation which showed a clear point at which sporulation genes tended to be downregulated, the RNA-seq and qPCR results did not suggest a clear point of differential gene expression in the solventogenesis process, however, not all key genes in the process were found and put in order of expression. Nevertheless, it was

observed from the RNA-seq data that genes encoding enzymes involved in acid formation, such as *pta* (CBEI_RS06155) involved in acetate formation, or *ptb* (CBEI_RS01180) and *buk (*CBEI_RS01185*)* genes involved in butyrate formation were either expressed at similar levels to the WT or upregulated at 12 hours. Whilst two other *buk* genes (CBEI_RS20510 and RS23520) were downregulated in the *agrB2* and *agrB4* mutant by 12 hours, their transcript levels were very low and this suggests they may not hold a large influence compared to the unchanged/upregulated *buk* gene (CBEI_RS01185) with much higher transcript levels. Furthermore, genes such as *adhE* (CBEI_RS01685), encoding enzymes involved in ethanol and butanol formation, were downregulated in the *agrB2* and *agrB4* mutants compared to the WT. For the *agrB2* and *agrB4* mutants, this would help explain why there was an accumulation in acids seen in **chapter 5**, since less acid would likely be converted into solvents here. Since this change in gene expression was not observed for the *agrB5* mutant, it suggests the difference in solvent accumulation, observed in **chapter 5,** is likely due to an alternative reason.

There were common genes thought to be involved in acid and solvent formation downregulated in all three *agrB* mutants, such as those encoding coA transferase subunit A (CBEI_RS13700), downregulated between 2 to 3-fold change in the *agrB2* and *agrB5* mutants and downregulated 17-fold change in the *agrB4* mutant compared to the WT at 24 hours. Generally, coA transferase genes in solventogenic

Clostridia are involved in acetate and butyrate uptake to reduce the toxicity of the growth medium, this is done by catalysing the transfer of CoA from acetoacetyl-CoA to acetate, butyrate and propionate. In effect, this produces acetoacetate which is decarboxylated to acetone, and butyryl-CoA which is further processed into butanol via solventogenesis, this would further explain the reduced solvent formation in *agrB* mutant (Yan *et al.*, 1988; Wiesenborn, Rudolph and Papoutsakis, 1989). Whilst the *spo0A* gene has been shown to be a regulatory element of solvent production in *C. beijerinckii* (Harris, Welker and Papoutsakis, 2002; Alsaker, Spitzer and Papoutsakis, 2004; Shi and Blaschek, 2008; Seo *et al.*, 2017; Humphreys *et al.*, 2023), it can be confirmed that the reduced butanol and acetone production observed in the *C. beijerinckii* NCIMB 8052 *agrB2*, *agrB4*, *agrB5* single mutants from this study, is not as a result of the *spo0A* gene, since the *spo0A* gene was shown not to be differentially regulated in the RNA-seq or qPCR data. As previously mentioned, other studies into the role of QS in clostridial solvent formation have shown varied results. For instance, Steiner *et al.* (2012) did not observe any significant change in solvent formation in the *agr* mutants of *C. acetobutylicum*. However, this may be due to the apparent role of RRNPP systems in solvent and spore formation in this species (Kotte *et al.*, 2020). These RRNPP systems have not been identified in *C. beijerinckii* NCIMB 8052 genome, and thus there is likely a more important role for Agr systems.

6.3.2.1 Nitrogen-fixation metabolism-related genes

Both *C. acetobutylicum* and *C. beijerinckii*, were previously found to be nitrogen-fixing organisms. Nitrogen fixation and the production of solvents are two reductive metabolic processes that can possibly compete against each other for the same source of reductant in the cell, if the two processes occur simultaneously. The *nif* genes are known to be the nitrogen-fixation genes in nitrogen-fixing *Clostridium* species (Chen, Toth and Kasap, 2001). These were differentially regulated in both the *agrB2* and *agrB4* mutant compared to the WT but in contrasting directions. At 12 hours, where cells are just reaching the very early stages of the stationary phase, in the agrB2 mutant the genes encoding *nifH*, *nifK*, *nifE* and *nifB* were all upregulated between 4 to 80-fold change compared to the WT. In contrast to this, the *agrB4* mutant showed these genes being downregulated at both 12 hours, by 38 to 62-fold, and 24 hours by 6 to 9-fold compared to the WT. The upregulation of these genes in *agrB2* could help explain the reduced solvent formation if nitrogen fixation does indeed compete with solvent formation genes for reductant sources in the cell. Studies by Kasap *et al.* (2002) found that when *C. beijerinckii* was grown in medium containing 4 mM ammonium acetate, the onset of nitrogen fixation coincided with the onset of solvent production and prevented accumulation of solvents to high levels. This

suggested competition between alcohol-producing enzymes and nitrogenase for the reductant. Whilst this may provide some explanation for reduced solvent formation in the *agrB2* mutant, the opposite effect on nitrogen-fixation genes was true for the *agrB4* mutant, where *nif* genes were largely downregulated. This suggests that alternative processes affected solvent production in this mutant and the *agrB5* mutant where *nif* genes were unaffected.

6.3.3 Sugar-uptake and Glycolysis related genes

In **Chapter 5,** during phenotypic charcterisations, it was observed that the *agrB* mutants consumed lower amounts of glucose over 96 hours of growth in liquid media, when compared to the WT. Solventogenic *Clostridia* can metabolise pentose and hexose sugars, with PTS being responsible for hexose sugar uptake, such as glucose, whilst non-PTS uptake pentose sugars (Mitchell, 2016). Glucose uptake is complicated in *Clostridium* species such as *C. beijerinckii* due to the greater number of phosphotransferase systems (PTS) present, indicated to encode 43 complete systems (Mitchell, 2015). The PTS is a multicomponent enzyme complex which includes IIA, IIB and IIC (Mitchell, 2015). Some noted PTS components from this study include the IIA subunit genes (CBEI_RS14155, CBEI_RS24005, CBEI_RS24495 and CBEI_RS25665), IIB subunit genes (CBEI_RS23895 and

CBEI_RS14150) and IIC subunit genes (CBEI_RS23890 and CBEI_RS14145). These were downregulated in the *agrB2* and *agrB4* mutants by 12 hours. However, these genes have not been confirmed experimentally to be involved in glucose uptake. Genes which have been experimental confirmed to do this include CBEI_RS04080 (Essalem and Mitchell, 2016), showing similar expression to the WT, CBEI_RS23135 and CBEI_RS23140 (Al Makishah and Mitchell, 2013) which show a 2 to 5-fold downregulation in the *agrB2* and *agrB4* mutants at 24 hours. As mentioned previously however, it may be that the reduced glucose consumption contributes to the lower overall OD 600 nm and vice versa, therefore these observations would need further analysis to determine if this was a direct or indirect result of the Agr systems. As typical saccharolytic species, both *C. acetobutylicum* and *C. beijerinckii* generate acetyl-CoA via the glycolysis pathway. In this study, various glycolysis-related genes were shown to be downregulated in the *agrB* single mutants by 24 hours in the RNA-Seq data. One of these genes, annotated as *gpmA*, was largely downregulated in *agrB2* and *agrB4* mutants by 12 hours. This gene encodes a putative phosphoglycerate mutase enzyme which catalyses the interconversion of 2-phosphoglycerateand 3-phosphoglycerate during glycolysis (Roth *et al.*, 2022). Another key glycolytic gene downregulated in the *agrB4* mutant is the ROK family protein CBEI_RS18045, reduced by 258-fold change by 12 hours and 19-fold by 24 hours compared to the WT. This downregulation may lead

to reduced conversion of glucose into pyruvate and hence affect the ABE fermentation metabolism.

6.3.4 Agr system hierarchy

In the *agrB2*, *agrB4* and *agrB5* single mutants it was observed that other annotated *agr* loci were differentially regulated. At 12 hours the *agrB2* mutant showed a slight upregulation in the *agrA1*, *agrBDC2*, *agrBD3*, *agrC4*, *agrBDCA5* and *agrBDCA6* systems. Similarly, the *agrB4* mutation showed to have caused an upregulation in *agrBDCA1*, *agrBDC2*, *agrBD3*, *agrBDC4*, *agrBDCA5* and *agrBDCA6* genes in the RNA-seq data. At this time point, the *agrB* mutants and the WT were in the very early stages of stationary phase, where cell density would be high and therefore quorum sensing systems would be expected to be active to encourage cell-cell communications. The AgrB2 and AgrB4 systems are highly conserved in *C. beijerinckii* strain genomes, as described in **Chapter 3**. Here Agr2 was present in 100 % and Agr 4 present in 91 % of strains analysed. Therefore, in the single mutants lacking either of these highly conserved systems it was unsurprising that there was upregulation in most or all of the other Agr systems identified. This upregulation of alternative Agr systems may help provide some form of compensation for gene expression due to the lack of Agr2 or Agr4 systems via a regulatory feedback loop. This also

suggests that when these systems are functional in the WT, they negatively regulate the other Agr systems during exponential phase of growth. This phenomenon of compensation of one QS system by another has been indicated in other bacteria. For instance, in *Pseudomonas aeruginosa* which is known to have multiple QS systems, Lee and Zhang (2015) indicated that the *rhl* QS system could compensate for the loss of the *las* system and maintain the expression of QS-dependent virulence factors. This was suggested due to the findings that RhlR was shown to upregulate the expression of the *lasI* gene in the *lasR* mutant, this is usually specifically regulated by LasR, this resulted in 3-oxo-C_{12}-HSL signal molecule production consequently being increased (Dekimpe and Déziel, 2009). The complexity of the *P. aeruginosa* multi-component QS network may allow the bacteria to adapt to various environmental conditions (Lee and Zhang, 2015).

By 24 hours, the bacterial cultures had reached the stationary phase of growth. At this stage, the WT and *agrB* mutants are all producing heat-resistant spores and solvents, with the *agrB2* and *agrB4* mutants in particular producing significantly less spores and all three *agrB* mutants producing less solvents (see **Chapter 5**). In the *agrB2* mutant the RNA-seq data now shows a downregulation of the AgrBDC4 and AgrBD5 system components suggesting the Agr2 system helps, directly or indirectly, to positively regulate these system components during later stages of growth in the WT. Meanwhile, the *agrA5* gene (CBEI_RS20310) and the *agrBDCA6* genes are all upregulated in this

mutant at early stationary phase growth, hence Agr2 encourages negative regulation of these in the WT. Similarly, in the *agrB4* mutant at 24 hours, the *agrBD2*, *agrBDC4* and *agrB5* genes were all downregulated whilst the *agrCA1*, *agrCA5* genes and the *agrBDCA6* genes are all upregulated which suggests that the Agr4 system under WT conditions directly or indirectly upregulates systems 2 and 4, and the *agrB5*, whilst downregulating system 6 and the *agrCA* genes of systems 1 and 5.

The Agr5 system has an altogether different affect when mutated. At 12 hours, this system mutation appeared to cause a downregulation in the *agrD4* gene (CBEI_RS26650), as well as the Agr5 system. By 24 hours this downregulation was extended to the *agrBC4* genes (CBEI_RS16315 and CBEI_RS16305) also, suggesting the Agr5 system may positively regulate the Agr4 system and self-regulate in the WT strain.

These results could suggest a system hierarchy in which specific systems are preferentially active at specific points of growth or cell densities. More conserved systems appear to be preferentially active over lesser conserved systems. For instance, it appears that the Agr2, 4 and 5 systems tend to positively regulate each other during the stationary phase (24 hours) whilst also negatively regulating lesser conserved systems such as Agr1 and 6, present in 14.3 % and 22.9 % of strains respectively (see **Chapter 3**). This QS system hierarchy is seen in other bacterial species also, the pathogen *P. aeruginosa* has 4

major interconnected AHL-based QS systems, *las, iqs, pqs* and *rhl*, with the *las* system sitting at the top of the hierarchy (Latifi *et al.*, 1996).

This study also identified LytTR-family regulators which were differentially regulated in the *agrB2* and *agrB4* mutants. All AgrA regulators contain the LytTR DNA-binding domain (Sidote *et al.*, 2008). Two identified LytTR-family regulator genes, CBEI_RS18730 and CBEI_RS17975 were found to be downregulated in the *agrB4* mutant, with the CBEI_RS03110 gene was upregulated slightly in the *agrB2* and *agrB4* mutant. These did not appear to belong to a specific gene cluster and therefore may be orphan regulators.

Chapter 7: General Discussion

This study started out by asking what role Agr QS has in the processes of solvent formation and sporulation in the industrially relevant bacterium *C. beijerinckii* NCIMB 8052. This analysis supports the hypothesis that some Agr systems affect either both or one of the processes. The key findings from this project show that at least two highly conserved Agr systems from *C. beijerinckii* species, denoted as Agr2 and Agr4, were found to significantly reduce the number of heat-resistant endospore formation in *C. beijerinckii* NCIMB 8052. This was observed by both sporulation assays to determine reduced heat-resistant CFU/ml and microscopic techniques observing reduced extracellular spores present in the *agrB2* and *agrB4* mutant samples when compared to the WT. In contrast, another highly conserved system, denoted as Agr5, did not appear to have a large effect on sporulation. Furthermore, there was no observed change in granulose accumulation in any of the generated mutants. However, all three *agrB* mutants had reduced glucose consumption and reduced formation of acetone and butanol when observed through GC and HPLC. Importantly, these differences in heat-resistant CFU/ml and reduced solvent formation could be genetically complemented, confirming that they were indeed linked to the generated mutations. These results are further supported by RNA-seq of each *agrB* mutant. This data showed many later stage sporulation genes to be downregulated in the *agrB2* and *agrB4* mutants. It was particularly interesting to observe a shared trend in these mutants which identified the *sigK* gene as a key point in the sporulation cascade in which gene expression was flipped from

similar expression to downregulation in comparison to the WT. This gene is imperative to the regulation of later stage sporulation genes such as spore cortex, spore coat and germination genes (Li, Di Donato and Piggot, 2004; Sun *et al.*, 2021). Genes involved in acid and solvent formation were also observed to be differentially regulated in the *agrB2* and *agrB4* mutants. The RNA-seq data in this study showed that genes encoding enzymes involved in acid formation, such as *pta* (CBEI_RS06155) involved in acetate formation, or *ptb* (CBEI_RS01180) and *buk (*CBEI_RS01185*)* genes involved in butyrate formation were either expressed at similar levels to the WT or upregulated at 12 hours whilst genes such as *adhE* (CBEI_RS01685), encoding enzymes involved in ethanol and butanol formation, were downregulated in the *agrB2* and *agrB4* mutants compared to the WT, explaining higher acid accumulation and reduced solvent phenotypes in these mutants. The RNA-seq data did not show obvious changes in key solvent-related gene expression in the *agrB5* mutant, but this may be because not all solvent related genes have been identified in *C. beijerinckii* and so an unknown gene may hold an important role. It was also observed that upon disrupting one *agr* system, alternative *agr* loci in the same isolate were differentially regulated, suggesting systems can positively or negatively affect each other. For example, when the *agrB2* and *agrB4* genes were disrupted other *agr* loci in the same mutant were often upregulated, which suggested the *agr* loci may to some degree compensate for loss of another in some cases.

There are a number of studies mentioned supporting the direction of these findings. Patakova *et al.* (2019) hypothesised that sporulation-related genes were regulated by Agr QS in *C. beijerinckii* NRRL B-598, this was based on in-depth transcriptional studies and the timely expression of all three Agr systems identified with the onset of sporulation. They also found that butanol shock led to differential gene expression in at least three putative *agr* gene loci. The study by Steiner *et al.* (2012), which created *agrB*, *agrC*, and *agrA* mutants of *C. acetobutylicum* ATCC 824, found that the number of heat-resistant endospores formed by mutants was significantly reduced when compared to the WT. However, in contrast to *C. beijerinckii* in the current study which showed similar levels of granulose accumulation to the WT, the C*. acetobutylicum* mutants did show reduced granulose accumulation. Since granulose accumulation has been linked to the onset of sporulation (Diallo, Kengen and López-Contreras, 2021), it would be expected that a reduction in sporulation would result in reduced granulose, however a study by Diallo *et al.* (2020) demonstrated that in a *C. beijerinckii* NCIMB 8052 *spoIIE* mutant, granulose formation genes were upregulated whilst granulose degradation genes such as *glgP* were unchanged. The current study shows the *glgP* gene, encoding glycogen phosphorylse, in *agrB* mutants was downregulated whilst granulose formation genes were unchanged. Whilst granulose and sporulation appear to be linked in *C. beijerinckii*, they are also independent of each other since mutants have been isolated that are deficient in one but not the other (Diallo, Kint, *et al.*, 2020). However, if

the central *spo0A* gene is mutated granulose accumulation is prevented (Ravagnani *et al.*, 2000; Humphreys *et al.*, 2023), showing that Spo0A is not affected directly by the *agrB* mutations in this study, a point which was proven in this studies RNA-seq results, where spo0A expression in the *agrB* mutants remained like that in the WT. The Steiner et al. (2012) study also did not see a change in solvent formation in the *C. acetobutylicum agr* mutants whilst the current study on *C. beijerinckii agrB* mutants did. This difference may be due to the presence of RRNPP-type QS in *C. acetobutylicum,* which is not present in *C. beijerinckii*, and therefore the Agr QS may hold more important regulatory roles in a larger variety of genes compared to in *C. acetobutylicum*. This idea is supported by the study by Kotte *et al.* (2020) who demonstrated that the RRNPP-type QS in *C. acetobutylicum* contributes to changes in solvent-related gene expression. The same study by Kotte et al. (2020), identified six putative Agr systems in *C. beijerinckii* NCIMB 8052, the current study on *C. beijerinckii* NCIMB 8052 also identified six putative systems with a possible seventh partial system that was denoted as ‘orphan’, during **chapter 3** bioinformatic studies. A phenomenon that has been suggested by the current study is the occurrence of abortive sporulation in *C. beijerinckii* NCIMB 8052. This suggestion was brought about by the observation that the *agrB2* and *agrB4* mutants, which showed reduced heat-resistant CFU/ml, still produced forespores. This suggests that QS-based regulation of sporulation occurs when forespores have already

formed rather than at the level of initiation. This would commit a lot of energy and resources without a means of exit, this brings about the question as to why QS regulation should happen at this late stage, unless there is indeed an exit route. It could be that abortive sporulation, in which the cells initiate but fail to complete spore development, had occurred. This has been observed in *B. subtilis* in minimal media by Mijakovic *et al.* (2016), who suggested this be a way for cells to delay or escape from unsuccessful sporulation.

This study did face an implication due to the genetic manipulation technique used to create the *agrB* mutations. It is known that genetic tools with high efficiency in *Clostridium* are not common and still being developed to their full potential. The technique used (Seys *et al.*, 2023) was a development of the CRISPR-Cas9 tool, relying on point mutations to result in gene disruption. Although this tool has been used in a number of *Clostridium* species (Nishida *et al.*, 2016; Banno *et al.*, 2018; Li *et al.*, 2019), a recent study did observe off-target mutations in similar genes and so full genome sequencing was advised after mutant isolation (Seys *et al.*, 2023). Therefore, the *agrB* mutant isolates were sent off for full genome sequencing and future analysis will determine if off-target point mutations occurred. The *agrB* mutants were successfully genetically complemented however, therefore polar effects of any off-target mutations are extremely limited/if any. Other future work into this area should include research into the AIP signaling molecules, with respect to their chemical composition. These could then be used to undertake chemical complementation studies using the *agr*

mutants produced from this study, as was conducted in the Steiner *et al.* (2012) study. The identification of specific AIP structures would also allow for cross-regulation studies, for example via shared AgrC histidine kinases responding to different AIPs. There were also the unstudied Agr systems identified in **chapter 3** denoted as Agr1, Agr3, Agr6 and Agr7 which could be studied further by gene disruption, phenotypic characterisation and RNA-seq as carried out in this study for systems Agr2, Agr4 and Agr5. This study also identified a common stage in the *agrB2* and *agrB4* mutants at which sporulation-related genes were downregulated, this being the stage of *spoIVB* and *sigK* regulation. Looking into why this stage is drastically affected and how. This could involve identifying corresponding *agrA* genes and direct targets of the AgrA response regulators.

This study into Agr QS in *C. beijerinckii* NCIMB 8052 has provided new evidence into the understanding of this industrially important bacteria. Whilst Agr systems have been identified in this strain, no studies had previously conducted in-depth experimental analysis to conclude this. Understanding the regulatory elements of processes affecting solvent formation, including sporulation, in *C. beijerinckii* is important as it can better our ability to increase the efficiency of solvent formation in this organism via processes such as strain improvement. This research also greatly benefits the fundamental understanding of how QS, which occurs at the high cell volumes in batch cultures, can be manipulated to make for a more successful fermentation process. A more successful and efficient fermentation system will create an increasingly competitive process that can act as a renewable source of fuels instead

of the use of petroleum-based fuels in industry. This would reduce the global dependence on non-renewable sources of energy making for a more sustainable future.

Bibliography

Abel-Santos, E. (2014) 'Endospores, Sporulation and Germination', in *Molecular Medical Microbiology: Second Edition*. Elsevier Ltd, pp. 163–178. doi: 10.1016/B978-0-12-397169-2.00009-3.

Abisado, R. G., Benomar, S., Klaus, J. R., Dandekar, A. A. and Chandler, J. R. (2018) 'Bacterial quorum sensing and microbial community interactions', *mBio*. doi: 10.1128/mBio.02331-17.

Aframian, N. and Eldar, A. (2020) 'A Bacterial Tower of Babel: Quorum-Sensing Signaling Diversity and Its Evolution', *Annual Review of Microbiology*, 74(1), pp. 587–606. doi: 10.1146/annurev-micro-012220-063740.

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', *Applied and Environmental Microbiology*. American Society for Microbiology, 78(22), pp. 8112–8121. doi: 10.1128/AEM.02214-12.

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G. and Nielsen, H. (2019) 'SignalP 5.0 improves signal peptide predictions using deep neural networks', *Nature Biotechnology*. Nature Publishing Group, 37(4), pp. 420–423. doi: 10.1038/s41587-019-0036-z.

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’, *Journal of Bacteriology*. American Society for Microbiology (ASM), 186(7), pp. 1959–1971. doi: 10.1128/JB.186.7.1959-1971.2004.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) ‘Gapped BLAST and PSI-BLAST: A new generation of protein database search programs’, *Nucleic Acids Research*, pp. 3389–3402. doi: 10.1093/nar/25.17.3389.

ANADÓN, L. D. (2014) Energy from Fossil Fuels: Challenges and Opportunities for Technology Innovation--Laura Daz Anadn | *Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2013 Symposium* | *The National Academies Press*. Available at: https://www.nap.edu/read/18558/chapter/15 (Accessed: 1 April 2020).

Andreesen, J. R., Bahl, H. and Gottschalk, G. (1989) ‘Introduction to the Physiology and Biochemistry of the Genus *Clostridium*’, in Minton, N. and Clarke, D. (eds) *Clostridia*. Boston: Springer US, pp. 27–62. doi: 10.1007/978-1-4757-9718-3_2.

Argyros, D. A., Tripathi, S. A., Barrett, T. F., Rogers, S. R., Feinberg, L. F., Olson, D. G., Foden, J. M., Miller, B. B., Lynd, L. R., Hogsett, D. A. and Caiazza, N. C. (2011) ‘High ethanol Titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes’, *Applied and Environmental Microbiology*. American Society for Microbiology, 77(23), pp. 8288–8294. doi: 10.1128/AEM.00646-11.

Auger, S., Krin, E., Aymerich, S. and Gohar, M. (2006) 'Autoinducer 2 affects biofilm formation by *Bacillus cereus*', *Applied and Environmental Microbiology*. American Society for Microbiology, 72(1), pp. 937–941. doi: 10.1128/AEM.72.1.937-941.2006.

Autret, N., Raynaud, C., Dubail, I., Berche, P. and Charbit, A. (2003) 'Identification of the agr locus of *Listeria monocytogenes*: Role in bacterial virulence', *Infection and Immunity*. American Society for Microbiology (ASM), 71(8), pp. 4463–4471. doi: 10.1128/IAI.71.8.4463-4471.2003.

Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H. and Kondo, A. (2018) 'Deaminase-mediated multiplex genome editing in *Escherichia coli*', *Nature Microbiology 2018 3:4*. Nature Publishing Group, 3(4), pp. 423–429. doi: 10.1038/s41564-017-0102-6.

Barák, I. and Youngman, P. (1996) 'SpoIIE mutants of *Bacillus subtilis* comprise two distinct phenotypic classes consistent with a dual functional role for the SpoIIE protein', *Journal of Bacteriology*. American Society for Microbiology, 178(16), pp. 4984–4989. doi: 10.1128/JB.178.16.4984-4989.1996.

Barketi-Klai, A., Hoys, S., Lambert-Bordes, S., Collignon, A. and Kansau, I. (2011) 'Role of fibronectin-binding protein A in *Clostridium difficile* intestinal colonization', *Journal of medical microbiology*. J Med Microbiol, 60(Pt 8), pp. 1155–1161. doi: 10.1099/JMM.0.029553-0.

Bassler, B. L. and Losick, R. (2006) 'Bacterially Speaking', *Cell*, pp.

237–246. doi: 10.1016/j.cell.2006.04.001.

Bassler, B. L., Wright, M. and Silverman, M. R. (1994) ‘Multiple signalling systems controlling expression of luminescence in *Vibrio harveyi* : sequence and function of genes encoding a second sensory pathway’, *Molecular Microbiology*. John Wiley & Sons, Ltd, 13(2), pp. 273–286. doi: 10.1111/j.1365-2958.1994.tb00422.x.

Bernsel, A., Viklund, H., Hennerdal, A. and Elofsson, A. (2009) ‘TOPCONS: Consensus prediction of membrane protein topology’, *Nucleic Acids Research*. Nucleic Acids Res, 37(SUPPL. 2). doi: 10.1093/nar/gkp363.

Bhatt, V. S. (2018) ‘Quorum sensing mechanisms in gram positive bacteria’, in Implication of Quorum Sensing System in Biofilm Formation and Virulence. Springer Singapore, pp. 297–311. doi: 10.1007/978-981-13-2429-1_20.

Birgen, C., Dürre, P., Preisig, H. A. and Wentzel, A. (2019) ‘Butanol production from lignocellulosic biomass: Revisiting fermentation performance indicators with exploratory data analysis’, *Biotechnology for Biofuels*. BioMed Central Ltd., 12(1), pp. 1–15. doi: 10.1186/s13068-019-1508-6.

Bolotin, A., Quinquis, B., Sorokin, A. and Dusko Ehrlich, S. (2005) ‘Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin’, *Microbiology*. Microbiology Society, 151(8), pp. 2551–2561. doi: 10.1099/mic.0.28048-0.

BP (2019) *Full report – BP Statistical Review of World Energy 2019*.

Burbulys, D., Trach, K. A. and Hoch, J. A. (1991) 'Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay', *Cell*. doi: 10.1016/0092-8674(91)90238-T.

Cabrera, R., Rocha, J., Flores, V., Vázquez-Moreno, L., Guarneros, G., Olmedo, G., Rodríguez-Romero, A. and de la Torre, M. (2014) 'Regulation of sporulation initiation by NprR and its signaling peptide NprRB: molecular recognition and conformational changes', *Applied Microbiology and Biotechnology*. Springer Verlag, 98(22), pp. 9399–9412. doi: 10.1007/s00253-014-6094-8.

Campbell, J., Lin, Q., Geske, G. D. and Blackwell, H. E. (2009) 'New and unexpected insights into the modulation of LuxR-type quorum sensing by cyclic dipeptides', *ACS Chemical Biology*. NIH Public Access, 4(12), pp. 1051–1059. doi: 10.1021/cb900165y.

Campos-Fernández, J., Arnal, J. M., Gómez, J. and Dorado, M. P. (2012) 'A comparison of performance of higher alcohols/diesel fuel blends in a diesel engine', *Applied Energy*. Elsevier Ltd, 95, pp. 267–275. doi: 10.1016/j.apenergy.2012.02.051.

Canovas, J., Baldry, M., Bojer, M. S., Andersen, P. S., Grzeskowiak, P. K., Stegger, M., Damborg, P., Olsen, C. A. and Ingmer, H. (2016) 'Cross-talk between *Staphylococcus aureus* and other staphylococcal species via the agr quorum sensing system', *Frontiers in Microbiology*. Frontiers Media S.A., 7, p. 1733. doi: 10.3389/fmicb.2016.01733.

Carter, G. P., Purdy, D., Williams, P. and Minton, N. P. (2005) 'Quorum sensing in *Clostridium difficile*: Analysis of a luxS-type signalling system', in *Journal of Medical Microbiology*. doi: 10.1099/jmm.0.45817-0.

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', *Applied and Environmental Microbiology*, 78(13), pp. 4683–4690. doi: 10.1128/AEM.00249-12.

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*. Springer Science and Business Media LLC, 3(5), pp. 415–431. doi: 10.1007/s13205-013-0167-8.

Chen, C., Sun, C. and Wu, Y. R. (2018) 'The Draft Genome Sequence of a Novel High-Efficient Butanol-Producing Bacterium *Clostridium Diolis* Strain WST', *Current Microbiology*. Springer New York LLC, 75, pp. 1011–1015. doi: 10.1007/s00284-018-1481-5.

Chen, J.-S., Toth, J. and Kasap, M. (2001) 'Nitrogen-fixation genes and nitrogenase activity in *Clostridium acetobutylicum* and *Clostridium beijerinckii*', *Journal of Industrial Microbiology & Biotechnology*, 27, pp. 281–286.

Chen, J., Ma, M., Uzal, F. A. and McClane, B. A. (2014) 'Host cell-

induced signaling causes *Clostridium perfringens* to upregulate production of toxins important for intestinal infections', *Gut Microbes*. Landes Bioscience, 5(1), pp. 96–107. doi: 10.4161/gmic.26419.

Chen, X., Schauder, S., Potier, N., Van Dorsselaer, A., Pelczer, I., Bassler, B. L. and Hughson, F. M. (2002) 'Structural identification of a bacterial quorum-sensing signal containing boron', *Nature*. Nature Publishing Group, 415(6871), pp. 545–549. doi: 10.1038/415545a.

Chen, Y., McClane, B. A., Fisher, D. J., Rood, J. I. and Gupta, P. (2005) 'Construction of an alpha toxin gene knockout mutant of *Clostridium perfringens* type A by use of a mobile group II intron', *Applied and Environmental Microbiology*. American Society for Microbiology (ASM), 71(11), pp. 7542–7547. doi: 10.1128/AEM.71.11.7542-7547.2005.

Chen, Y., Yang, Y., Ji, X., Zhao, R., Li, G., Gu, Y., Shi, A., Jiang, W. and Zhang, Q. (2020) 'The SCIFF-derived ranthipeptides participate in quorum sensing in solventogenic clostridia', *Biotechnology Journal*. Wiley, p. 2000136. doi: 10.1002/biot.202000136.

Cheung, J. K., Keyburn, A. L., Carter, G. P., Lanckriet, A. L., Van Immerseel, F., Moore, R. J. and Rood, J. I. (2010) 'The VirSR two-component signal transduction system regulates NetB toxin production in *Clostridium perfringens*', *Infection and Immunity*, 78(7), pp. 3064–3072. doi: 10.1128/IAI.00123-10.

Choi, J., Shin, D., Kim, M., Park, J., Lim, S. and Ryu, S. (2012) 'LsrR-mediated quorum sensing controls invasiveness of *Salmonella*

typhimurium by regulating SPI-1 and flagella genes', *PLoS ONE*. Public Library of Science, 7(5). doi: 10.1371/journal.pone.0037059.

Choi, S. J., Lee, J., Jang, Y. S., Park, J. H., Lee, S. Y. and Kim, I. H. (2012) 'Effects of nutritional enrichment on the production of acetone-butanol-ethanol (ABE) by *Clostridium acetobutylicum*', *Journal of Microbiology*, 50(6), pp. 1063–1066. doi: 10.1007/s12275-012-2373-1.

Choudhary, K. S., Mih, N., Monk, J., Kavvas, E., Yurkovich, J. T., Sakoulas, G. and Palsson, B. O. (2018) 'The *Staphylococcus aureus* two-component system AgrAC displays four distinct genomic arrangements that delineate genomic virulence factor signatures', *Frontiers in Microbiology*. Frontiers Media S.A., 9(1082). doi: 10.3389/fmicb.2018.01082.

Collins, M. D., Lawson, P. A., Willems, A., Cordoba, J. J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H. and Farrow, J. A. E. (1994) 'The phylogeny of the genus *Clostridium*: Proposal of five new genera and eleven new species combinations', *International Journal of Systematic Bacteriology*. Microbiology Society, 44(4), pp. 812–826. doi: 10.1099/00207713-44-4-812.

Cooksley, C. M., Davis, I. J., Winzer, K., Chan, W. C., Peck, M. W. and Minton, N. P. (2010) 'Regulation of neurotoxin production and sporulation by a putative *agrBD* signaling system in proteolytic *Clostridium botulinum*', *Applied and Environmental Microbiology*. doi: 10.1128/AEM.03038-09.

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', *Metabolic Engineering*, 14(6), pp. 630–641. doi: 10.1016/j.ymben.2012.09.001.

Cruz-Morales, P., Orellana, C. A., Moutafis, G., Moonen, G., Rincon, G., Nielsen, L. K. and Marcellin, E. (2019) 'Revisiting the Evolution and Taxonomy of *Clostridia*, a Phylogenomic Update', *Genome biology and evolution*. NLM (Medline), pp. 2035–2044. doi: 10.1093/gbe/evz096.

D'Agostino, N., Li, W. and Wang, D. (2022) 'High-throughput transcriptomics', *Scientific Reports*. doi: 10.1038/s41598-022-23985-1.

Darkoh, C., Dupont, H. L., Norris, S. J. and Kaplan, H. B. (2015) 'Toxin synthesis by *Clostridium difficile* is regulated through quorum signaling', *mBio*. American Society for Microbiology, 6(2). doi: 10.1128/mBio.02569-14.

Declerck, N., Bouillaut, L., Chaix, D., Rugani, N., Slamti, L., Hoh, F., Lereclus, D. and Arold, S. T. (2007) 'Structure of PlcR: Insights into virulence regulation and evolution of quorum sensing in Gram-positive bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, 104(47), pp. 18490–18495. doi: 10.1073/pnas.0704501104.

Dekimpe, V. and Déziel, E. (2009) 'Revisiting the quorum-sensing hierarchy in *Pseudomonas aeruginosa*: The transcriptional regulator RhlR regulates LasR-specific factors', *Microbiology*, 155(3). doi:

10.1099/mic.0.022764-0.

Diallo, M., Hocq, R., Collas, F., Chartier, G., Wasels, F., Wijaya, H. S., Werten, M. W. T., Wolbert, E. J. H., Kengen, S. W. M., van der Oost, J., Ferreira, N. L. and López-Contreras, A. M. (2020a) 'Adaptation and application of a two-plasmid inducible CRISPR-Cas9 system in *Clostridium beijerinckii*', *Methods*. Academic Press Inc., 172, pp. 51–60. doi: 10.1016/j.ymeth.2019.07.022.

Diallo, M., Hocq, R., Collas, F., Chartier, G., Wasels, F., Wijaya, H. S., Werten, M. W. T., Wolbert, E. J. H., Kengen, S. W. M., van der Oost, J., Ferreira, N. L. and López-Contreras, A. M. (2020b) 'Adaptation and application of a two-plasmid inducible CRISPR-Cas9 system in *Clostridium beijerinckii'*, *Methods*. Academic Press Inc., 172, pp. 51–60. doi: 10.1016/J.YMETH.2019.07.022.

Diallo, M., Kengen, S. W. M. and López-Contreras, A. M. (2021) 'Sporulation in solventogenic and acetogenic clostridia', *Applied Microbiology and Biotechnology*. Springer Science and Business Media Deutschland GmbH, 105(9), pp. 3533–3557. doi: 10.1007/S00253-021-11289-9.

Diallo, M., Kint, N., Monot, M., Collas, F., Martin-Verstraete, I., van der Oost, J., Kengen, S. W. M. and López-Contreras, A. M. (2020) 'Transcriptomic and Phenotypic Analysis of a *spoIIE* Mutant in *Clostridium beijerinckii'*, *Frontiers in Microbiology*. Frontiers Media S.A., 11, p. 2225. doi: 10.3389/FMICB.2020.556064/BIBTEX.

Dingle, T. C., Mulvey, G. L. and Armstrong, G. D. (2011) 'Mutagenic analysis of the *Clostridium difficile* flagellar proteins, flic and flid, and their contribution to virulence in hamsters', *Infection and Immunity*, 79(10), pp. 4061–4067. doi: 10.1128/IAI.05305-11.

Do, H. and Kumaraswami, M. (2016) 'Structural Mechanisms of Peptide Recognition and Allosteric Modulation of Gene Regulation by the RRNPP Family of Quorum-Sensing Regulators', *Journal of Molecular Biology*. doi: 10.1016/j.jmb.2016.05.026.

Do, H., Makthal, N., VanderWal, A. R., Saavedra, M. O., Olsen, R. J., Musser, J. M. and Kumaraswami, M. (2019) 'Environmental pH and peptide signaling control virulence of *Streptococcus pyogenes* via a quorum-sensing pathway', *Nature Communications*. doi: 10.1038/s41467-019-10556-8.

Dürre, P. (2014) 'Physiology and Sporulation in *Clostridium*', *Microbiology Spectrum*. American Society for Microbiology, 2(4). doi: 10.1128/microbiolspec.tbs-0010-2012.

Ehsaan, M., Kuehne, S. A. and Minton, N. P. (2016) '*Clostridium difficile* genome editing using *pyrE* alleles', in *Methods in Molecular Biology*. Humana Press Inc., pp. 35–52. doi: 10.1007/978-1-4939-6361-4_4.

Ehsaan, M., Kuit, W., Zhang, Y., Cartman, S. T., Heap, J. T., Winzer, K. and Minton, N. P. (2016) 'Mutant generation by allelic exchange and genome resequencing of the biobutanol organism *Clostridium acetobutylicum* ATCC 824', *Biotechnology for Biofuels*. doi:

10.1186/s13068-015-0410-0.

Engebrecht, J. A. and Silverman, M. (1984) ‘Identification of genes and gene products necessary for bacterial bioluminescence’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 81(13), pp. 4154–4158. doi: 10.1073/pnas.81.13.4154.

Essalem, M. E. E. and Mitchell, W. J. (2016) ‘Identification of a glucose–mannose phosphotransferase system in *Clostridium beijerinckii*’, *FEMS Microbiology Letters*. Oxford Academic, 363(8), p. 53. doi: 10.1093/FEMSLE/FNW053.

Even-Tov, E., Omer Bendori, S., Valastyan, J., Ke, X., Pollak, S., Bareia, T., Ben-Zion, I., Bassler, B. L. and Eldar, A. (2016) ‘Social Evolution Selects for Redundancy in Bacterial Quorum Sensing’, *PLoS Biology*. doi: 10.1371/journal.pbio.1002386.

Ezeji, T. C., Qureshi, N. and Blaschek, H. P. (2004) ‘Butanol fermentation research: Upstream and downstream manipulations’, *Chemical Record*. John Wiley & Sons, Ltd, 4(5), pp. 305–314. doi: 10.1002/tcr.20023.

Feng, J., Zong, W., Wang, P., Zhang, Z.-T., Gu, Y., Dougherty, M., Borovok, I. and Wang, Y. (2020) ‘RRNPP-type quorum-sensing systems regulate solvent formation, sporulation and cell motility in *Clostridium saccharoperbutylacetonicum*’, *Biotechnology for Biofuels*. Springer Science and Business Media LLC, 13(1). doi:

10.1186/s13068-020-01723-x.

Fischer, R. J., Helms, J. and Durre, P. (1993) ‘Cloning, sequencing, and molecular analysis of the sol operon of *Clostridium acetobutylicum*, a chromosomal locus involved in solventogenesis’, *Journal of Bacteriology*. American Society for Microbiology (ASM), 175(21), pp. 6959–6969. doi: 10.1128/jb.175.21.6959-6969.1993.

Fonseca, B. C., Riaño-Pachón, D. M., Guazzaroni, M. E. and Reginatto, V. (2019) ‘Genome sequence of the H2-producing *Clostridium beijerinckii* strain Br21 isolated from a sugarcane vinasse treatment plant’, *Genetics and Molecular Biology*. Brazilian Journal of Genetics, 42(1), pp. 139–144. doi: 10.1590/1678-4685-gmb-2017-0315.

Formanek, J., Mackie, R. and Blaschek, H. P. (1997) ‘Enhanced butanol production by *Clostridium beijerinckii* BA101 grown in semidefined P2 medium containing 6 percent maltodextrin or glucose’, *Applied and Environmental Microbiology*, 63(6), pp. 2306–2310. doi: 10.1128/aem.63.6.2306-2310.1997.

Foulquier, C., Huang, C.-N., Nguyen, N.-P.-T., Thiel, A., Wilding-Steel, T., Soula, J., Yoo, M., Ehrenreich, A., Meynial-Salles, I., Liebl, W. and Soucaille, P. (2019) ‘An efficient method for markerless mutant generation by allelic exchange in *Clostridium acetobutylicum* and *Clostridium saccharobutylicum* using suicide vectors’, *Biotechnology for Biofuels*, 12(31). doi: 10.1186/s13068-019-1364-4.

Fox, M. E., Lemmon, M. J., Mauchline, M. L., Davis, T. O., Giaccia, A.

J., Minton, N. P. and Brown, J. M. (1996) ‘Anaerobic bacteria as a delivery system for cancer gene therapy: In vitro activation of 5-fluorocytosine by genetically engineered clostridia’, *Gene Therapy*, 3(2).

Frazier, C. L., San Filippo, J., Lambowitz, A. M. and Mills, D. A. (2003) ‘Genetic manipulation of *Lactococcus lactis* by using targeted group II introns: Generation of stable insertions without selection’, *Applied and Environmental Microbiology*. American Society for Microbiology, 69(2), pp. 1121–1128. doi: 10.1128/AEM.69.2.1121-1128.2003.

Friedlingstein, P., Jones, M. W., O’Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., DBakker, O. C. E., Canadell1, J. G., Ciais1, P., Jackson, R. B., Anthoni1, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Goldewijk, K. K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S. I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., Van Der Werf, G. R., Wiltshire, A. J. and Zaehle, S. (2019) ‘Global carbon budget 2019’, *Earth System Science Data*. Copernicus GmbH,

11(4), pp. 1783–1838. doi: 10.5194/essd-11-1783-2019.

Gao, K., Mu, C.-L., Farzi, A. and Zhu, W.-Y. (2020) 'Tryptophan Metabolism: A Link Between the Gut Microbiota and Brain', 11(3). doi: 10.1093/advances/nmz127.

Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. (2012) 'Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 109(39). doi: 10.1073/PNAS.1208507109.

Gerischer, U. and Durre, P. (1990) 'Cloning, sequencing, and molecular analysis of the acetoacetate decarboxylase gene region from *Clostridium acetobutylicum*', *Journal of Bacteriology*, 172(12), pp. 6907–6918. doi: 10.1128/jb.172.12.6907-6918.1990.

Gottschal, J. C. and Morris, J. G. (1981) 'Non-production of acetone and butanol by *Clostridium acetobutylicum* - during glucose- and ammonium- limitation in continuous culture', *Biotechnology Letters*, 3(9), pp. 525–530.

Gu, H., Yang, Y., Wang, M., Chen, S., Wang, H., Li, S., Ma, Y. and Wang, J. (2018) 'Novel cysteine desulfidase CdsB involved in releasing cysteine repression of toxin synthesis in *Clostridium difficile*', *Frontiers in Cellular and Infection Microbiology*. Frontiers Media S.A., 7(531). doi: 10.3389/fcimb.2017.00531.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. and

Gascuel, O. (2010) ‘New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0’, *Systematic Biology*, 59(3), pp. 307–321. doi: 10.1093/sysbio/syq010.

Guo, Y., Zhang, Y., Li, J. L. and Wang, N. (2012) ‘Diffusible signal factor-mediated quorum sensing plays a central role in coordinating gene expression of *Xanthomonas citri* subsp. citri’, *Molecular Plant-Microbe Interactions*, 25(2), pp. 165–179. doi: 10.1094/MPMI-07-11-0184.

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’, *Journal of Bacteriology*, 184(13), pp. 3586–3597. doi: 10.1128/JB.184.13.3586-3597.2002.

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 Research*. doi: 10.1093/nar/gkr1321.

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’, *Journal of Microbiological Methods*. J Microbiol Methods, 80(1), pp. 49–55. doi: 10.1016/j.mimet.2009.10.018.

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'*, *Journal of Microbiological Methods*. Elsevier, 70(3), pp. 452–464. doi: 10.1016/j.mimet.2007.05.021.

Heap, J. T., Pennington, O. J., Cartman, S. T. and Minton, N. P. (2009) 'A modular system for *Clostridium* shuttle plasmids', *Journal of Microbiological Methods*. J Microbiol Methods, 78(1), pp. 79–85. doi: 10.1016/j.mimet.2009.05.004.

Hendriks, A. T. W. M. and Zeeman, G. (2009) 'Pretreatments to enhance the digestibility of lignocellulosic biomass', *Bioresource Technology*. Elsevier Ltd, 100(1), pp. 10–18. doi: 10.1016/j.biortech.2008.05.027.

Herman, N. A., Kim, S. J., Li, J. S., Cai, W., Koshino, H. and Zhang, W. (2017) 'The industrial anaerobe *Clostridium acetobutylicum* uses polyketides to regulate cellular differentiation', *Nature Communications*. doi: 10.1038/s41467-017-01809-5.

Herring, S. C., Christidis, N., Hoell, A., Kossin, J., Schreck III, C. J. and Stott, P.(2018) 'Explaining Extreme Events of 2016 From A Climate Perspective', *Bulletin of the American Meteorological Society*, 99(1). doi: 10.1175/BAMS-D-17-0118.1.

Hoa, N. T., Brannigan, J. A. and Cutting, S. M. (2002) 'The *Bacillus subtilis* Signaling Protein SpoIVB Defines a New Family of Serine Peptidases', *Journal of Bacteriology*. American Society for Microbiology

(ASM), 184(1), p. 191. doi: 10.1128/JB.184.1.191-199.2002.

Huang, C. N., Liebl, W. and Ehrenreich, A. (2018) 'Restriction-deficient mutants and marker-less genomic modification for metabolic engineering of the solvent producer *Clostridium saccharobutylicum*', *Biotechnology for Biofuels*. BioMed Central Ltd., 11(264). doi: 10.1186/s13068-018-1260-3.

Humphreys, J. R. (2019) *Clostridial strain degeneration: new approaches to an old problem*.

Humphreys, J. R., Debebe, B. J., Diggle, S. P. and Winzer, K. (2023) '*Clostridium beijerinckii* strain degeneration is driven by the loss of Spo0A activity', *Frontiers in Microbiology*. Frontiers Media S.A., 13. doi: 10.3389/FMICB.2022.1075609/FULL.

IEA (2019) *World Energy Outlook 2019 – Analysis - IEA*. Available at: https://www.iea.org/reports/world-energy-outlook-2019 (Accessed: 1 April 2020).

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987) 'Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product.', *Journal of Bacteriology*, 169(12), p. 5429. doi: 10.1128/jb.169.12.5429-5433.1987.

Jabbari, S., Steiner, E., Heap, J. T., Winzer, K., Minton, N. P. and King, J. R. (2013) 'The putative influence of the agr operon upon survival mechanisms used by *Clostridium acetobutylicum*', *Mathematical*

Biosciences. doi: 10.1016/j.mbs.2013.03.005.

Jackson, R. B., Le Quéré, C., Andrew, R. M., Canadell, J. G., Korsbakken, J. I., Liu, Z., Peters, G. P., Zheng, B. and Friedlingstein, P. (2019) *GLOBAL ENERGY GROWTH IS OUTPACING DECARBONIZATION*. doi: 10.1088/1748-9326/aaf303.

Jansen, R., Van Embden, J. D. A., Gaastra, W. and Schouls, L. M. (2002) 'Identification of genes that are associated with DNA repeats in prokaryotes', *Molecular Microbiology*. John Wiley & Sons, Ltd, 43(6), pp. 1565–1575. doi: 10.1046/j.1365-2958.2002.02839.x.

Ji, G., Beavis, R. and Novick, R. P. (1997) 'Bacterial interference caused by autoinducing peptide variants', *Science*, 276(5321), pp. 2027–2030. doi: 10.1126/science.276.5321.2027.

Ji, G., Pei, W., Zhang, L., Qiu, R., Lin, J., Benito, Y., Lina, G. and Novick, R. P. (2005) '*Staphylococcus intermedius* produces a functional agr autoinducing peptide containing a cyclic lactone', *Journal of Bacteriology*. American Society for Microbiology (ASM), 187(9), pp. 3139–3150. doi: 10.1128/JB.187.9.3139-3150.2005.

Jiang, M., Grau, R. and Perego, M. (2000) 'Differential processing of propeptide inhibitors of rap phosphatases in *Bacillus subtilis'*, *Journal of Bacteriology*, 182(2), pp. 303–310. doi: 10.1128/JB.182.2.303-310.2000/FORMAT/EPUB.

Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L. A. (2013) 'RNA-guided editing of bacterial genomes using CRISPR-Cas systems',

Nature Biotechnology 2013 31:3. Nature Publishing Group, 31(3), pp. 233–239. doi: 10.1038/nbt.2508.

Jiang, Y., Chen, T., Dong, W., Zhang, M., Zhang, W., Wu, H., Ma, J., Jiang, M. and Xin, F. (2018) 'The Draft Genome Sequence of *Clostridium beijerinckii* NJP7, a Unique Bacterium Capable of Producing Isopropanol–Butanol from Hemicellulose Through Consolidated Bioprocessing', *Current Microbiology*. Springer New York LLC, 75(3), pp. 305–308. doi: 10.1007/s00284-017-1380-1.

Jin, Q., Qureshi, N., Wang, H. and Huang, H. (2019) 'Acetone-butanol-ethanol (ABE) fermentation of soluble and hydrolyzed sugars in apple pomace by *Clostridium beijerinckii* P260', *Fuel*. Elsevier Ltd, 244, pp. 536–544. doi: 10.1016/j.fuel.2019.01.177.

Johnson, J. L. and Francis, B. S. (1975) 'Taxonomy of the *clostridia*: ribosomal ribonucleic acid homologies among the species', *Journal of General Microbiology*, 88(2), pp. 229–244. doi: 10.1099/00221287-88-2-229.

Jones, D. T., Schulz, F., Roux, S. and Brown, S. D. (2023) 'Solvent-Producing Clostridia Revisited', *Microorganisms*, 11(9). doi: 10.3390/microorganisms11092253.

Jones, D. T., Van Der Westhuizen, A., Long, S., Allcock, E. R., Reid, S. J. and Woods, D. R. (1982) 'Solvent Production and Morphological Changes in *Clostridium acetobutylicum*', *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 43(6), pp. 1434–1439.

Jones, D. T. and Woods, D. R. (1986) ‘Acetone-Butanol Fermentation Revisited’, *MICROBIOLOGICAL REVIEWS*, 50(4), pp. 484–524.

Jones, M. A. A. H. and Papoutsakis, E. T. (2014) ‘σK of *clostridium acetobutylicum* is the first known sporulation-: Specific sigma factor with two developmentally separated roles, one early and one late in sporulation’, *Journal of Bacteriology*, 196(2), pp. 287–299. doi: 10.1128/JB.01103-13/SUPPL_FILE/ZJB999092976SO1.PDF.

Jönsson, L. J., Alriksson, B. and Nilvebrant, N. O. (2013) ‘Bioconversion of lignocellulose: Inhibitors and detoxification’, *Biotechnology for Biofuels*. BioMed Central, 6(1), pp. 1–10. doi: 10.1186/1754-6834-6-16.

Jørgensen, H. and Pinelo, M. (2017) ‘Enzyme recycling in lignocellulosic biorefineries’, *Biofuels, Bioproducts and Biorefining*. John Wiley and Sons Ltd, 11(1), pp. 150–167. doi: 10.1002/bbb.1724.

Junier, T. and Zdobnov, E. M. (2010) ‘The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell’, *Bioinformatics*, 26(13), pp. 1669–1670. doi: 10.1093/bioinformatics/btq243.

Kalamara, M., Spacapan, M., Mandic-Mulec, I. and Stanley-Wall, N. R. (2018) ‘Social behaviours by *Bacillus subtilis*: quorum sensing, kin discrimination and beyond’, *Molecular Microbiology*. Blackwell Publishing Ltd, 110(6), pp. 863–878. doi: 10.1111/mmi.14127.

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’, *Nature Biotechnology*. Nature Publishing Group, 19(12), pp. 1162–1167. doi: 10.1038/nbt1201-1162.

Kasap, M., Newton, W. E., White, R. H., Gregory, E. M. and Claus, G. W. (2002) ‘NITROGEN METABOLISM AND SOLVENT PRODUCTION IN CLOSTRIDIUM BEIJERINCKII NRRL B593’.

Kashket, E. R. and Cao, Z.-Y. (1995) ‘Clostridial strain degeneration’, *FEMS Microbiology Reviews*, 17, pp. 307–315. doi: 10.1111/j.1574-6976.1995.tb00214.x.

Keis, S., Bennett, C. F., Ward, V. K. and Jones, D. T. (1995) *Taxonomy and Phylogeny of Industrial Solvent-Producing* Clostridia, *INTERNATIONAL JOURNAL OF SYSTEMATIC BACTERIOLOGY*. doi: https://doi.org/10.1099/00207713-45-4-693.

Keis, S., Shaheen, R. and Jones, D. T. (2001) ‘Emended descriptions of *Clostridium acetobutylicum* and *Clostridium beijerinckii*, and descriptions of *Clostridium saccharoperbutylacetonicum* sp. nov. and *Clostridium saccharobutylicum* sp. nov.’, *International Journal of Systematic and Evolutionary Microbiology*. doi: 10.1099/00207713-51-6-2095.

Kobayashi, H., Tanizawa, Y., Sakamoto, M., Nakamura, Y., Ohkuma, M. and Tohno, M. (2020) ‘Reclassification of *Clostridium diolis* Biebl and Spröer 2003 as a later heterotypic synonym of *Clostridium beijerinckii* Donker 1926 (Approved Lists 1980) emend. Keis et al.

2001', *International Journal of Systematic and Evolutionary Microbiology*. Microbiology Society, 70(4), pp. 2463–2466. doi: 10.1099/ijsem.0.004059.

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', *Bioscience, Biotechnology and Biochemistry*, 71(1), pp. 58–68. doi: 10.1271/bbb.60370.

Kotte, A.-K., Severn, O., Bean, Z., Schwarz, K., Minton, N. P. and Winzer, K. (2020) 'RNPP-type quorum sensing regulates solvent formation and sporulation in *Clostridium acetobutylicum*', *bioRxiv*. doi: 10.1101/106666.

Kuan Ng, Y., 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(2). doi: 10.1371/journal.pone.0056051.

Kuehne, S. A., Heap, J. T., Cooksley, C. M., Cartman, S. T. and Minton, N. P. (2011) 'ClosTron-mediated engineering of *Clostridium*', *Methods in Molecular Biology*. doi: 10.1007/978-1-61779-197-0_23.

Kuipers, O. P., De Ruyter, P. G. G. A., Kleerebezem, M. and De Vos, W. M. (1998) 'Quorum sensing-controlled gene expression in lactic acid bacteria', *Journal of Biotechnology*. Elsevier Sci B.V., pp. 15–21. doi:

10.1016/S0168-1656(98)00100-X.

Kukurba, K. R. and Montgomery, S. B. (2015) ‘RNA Sequencing and Analysis’, *Cold Spring Harbor protocols*. NIH Public Access, 2015(11), p. 951. doi: 10.1101/PDB.TOP084970.

Kumar, R. and Kumar, P. (2017) ‘Future Microbial Applications for Bioenergy Production: A Perspective’, *Frontiers in Microbiology*. Frontiers Research Foundation, 8(MAR), p. 450. doi: 10.3389/fmicb.2017.00450.

de la Fuente-Núñez, C. and Lu, T. K. (2017) ‘CRISPR-Cas9 technology: applications in genome engineering, development of sequence-specific antimicrobials, and future prospects’, *Integrative Biology*. Oxford Academic, 9(2), pp. 109–122. doi: 10.1039/C6IB00140H.

Lang, J. and Faure, D. (2014) ‘Functions and regulation of quorum-sensing in *Agrobacterium tumefaciens*’, *Frontiers in Plant Science*. Frontiers Research Foundation, 5(14). doi: 10.3389/fpls.2014.00014.

Lapuerta, M., Rodríguez-Fernández, J., Fernández-Rodríguez, D. and Patiño-Camino, R. (2017) ‘Modeling viscosity of butanol and ethanol blends with diesel and biodiesel fuels’, *Fuel*. Elsevier Ltd, 199, pp. 332–338. doi: 10.1016/j.fuel.2017.02.101.

Latifi, A., Foglino, M., Tanaka, K., Williams, P. and Lazdunski, A. (1996) ‘A hierarchical quorum-sensing cascade in *Pseudomonas aeruginosa* links the transcriptional activators LasR and RhIR (VsmR) to expression

of the stationary-phase sigma factor RpoS', *Molecular Microbiology*, 21(6). doi: 10.1046/j.1365-2958.1996.00063.x.

Lawson, P. A. and Rainey, F. A. (2016) 'Proposal to restrict the genus *Clostridium prazmowski* to *Clostridium butyricum* and related species', *International Journal of Systematic and Evolutionary Microbiology*. Microbiology Society, 66(2), pp. 1009–1016. doi: 10.1099/ijsem.0.000824.

Le, K. Y. and Otto, M. (2015) 'Quorum-sensing regulation in *staphylococci*-an overview', *Frontiers in Microbiology*. Frontiers Media S.A., 6(OCT). doi: 10.3389/fmicb.2015.01174.

Lee, J., Seo, E., Kweon, D. H., Park, K. and Jin, Y. S. (2009) 'Fermentation of Rice Bran and Defatted Rice Bran for Butanol Production Using *Clostridium beijerinckii* NCIMB 8052', *Journal of Microbiology and Biotechnology*. The Korean Society for Microbiology and Biotechnology, 19(5), pp. 482–490. doi: 10.4014/JMB.0804.275.

Lee, J. and Zhang, L. (2015) 'The hierarchy quorum sensing network in *Pseudomonas aeruginosa'*, *Protein Cell*, 6(1). doi: 10.1007/s13238-014-0100-x.

Lee, R. A. and Lavoie, J.-M. (2013) 'From first- to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity', *Animal Frontiers*. Oxford University Press (OUP), 3(2), pp. 6–11. doi: 10.2527/af.2013-0010.

Lee, S. H., Kim, H. J., Shin, Y. A., Kim, K. H. and Lee, S. J. (2016)

‘Single Crossover-Mediated Markerless Genome Engineering in *Clostridium acetobutylicum’*, *Journal of Microbiology and Biotechnology*. The Korean Society for Microbiology and Biotechnology, 26(4), pp. 725–729. doi: 10.4014/JMB.1512.12012.

Lefort, V., Longueville, J. E. and Gascuel, O. (2017) ‘SMS: Smart Model Selection in PhyML’, *Molecular biology and evolution*, 34(9), pp. 2422–2424. doi: 10.1093/molbev/msx149.

Lemmon, M. J., Van Zijl, P., Fox, M. E., Mauchline, M. L., Giaccia, A. J., Minton, N. P. and Brown, J. M. (1997) ‘Anaerobic bacteria as a gene delivery system that is controlled by the tumor microenvironment’, *Gene Therapy*, 4(8). doi: 10.1038/sj.gt.3300468.

Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S. and Gascuel, O. (2019) ‘NGPhylogeny.fr: new generation phylogenetic services for non-specialists’, *Web Server issue Published online*, 47. doi: 10.1093/nar/gkz303.

Lemoine, F., Domelevo Entfellner, J. B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T. and Gascuel, O. (2018) ‘Renewing Felsenstein’s phylogenetic bootstrap in the era of big data’, *Nature*. doi: 10.1038/s41586-018-0043-0.

Lerat, E. and Moran, N. A. (2004) ‘The Evolutionary History of Quorum-Sensing Systems in Bacteria’, *Molecular Biology and Evolution*, 21(5), pp. 903–913. doi: 10.1093/molbev/msh097.

Letunic, I. and Bork, P. (2019) ‘Interactive Tree of Life (iTOL) v4:

Recent updates and new developments', *Nucleic Acids Research*. Oxford University Press, 47(W1). doi: 10.1093/nar/gkz239.

Li, H., Yang, Y., Hong, W., Huang, M., Wu, M. and Zhao, X. (2020) 'Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects', 5(1). doi: 10.1038/s41392-019-0089-y.

Li, J., Chen, J., Vidal, J. E. and McClane, B. A. (2011) 'The Agr-like quorum-sensing system regulates sporulation and production of enterotoxin and beta2 toxin by *Clostridium perfringens* type a non-food-borne human gastrointestinal disease strain F5603', *Infection and Immunity*. doi: 10.1128/IAI.00169-11.

Li, Q., Chen, J., Minton, N. P., Zhang, Y., Wen, Z., Liu, J., Yang, H., Zeng, Z., Ren, X., Yang, J., Gu, Y., Jiang, W., Jiang, Y. and Yang, S. (2016) 'CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*', *Biotechnology Journal*. doi: 10.1002/biot.201600053.

Li, Q., Seys, F. M., Minton, N. P., Yang, J., Jiang, Y., Jiang, W. and Yang, S. (2019) 'CRISPR–Cas9D10A nickase-assisted base editing in the solvent producer *Clostridium beijerinckii*', *Biotechnology and Bioengineering*. John Wiley & Sons, Ltd, 116(6), pp. 1475–1483. doi: 10.1002/BIT.26949.

Li, S., Huang, L., Ke, C., Pang, Z. and Liu, L. (2020) 'Pathway dissection, regulation, engineering and application: Lessons learned

from biobutanol production by solventogenic clostridia', *Biotechnology for Biofuels*. BioMed Central Ltd., 13(1). doi: 10.1186/s13068-020-01674-3.

Li, Z., Di Donato, F. and Piggot, P. J. (2004) 'Compartmentalization of Gene Expression during Sporulation of *Bacillus subtilis* Is Compromised in Mutants Blocked at Stage III of Sporulation', *Journal of Bacteriology*. American Society for Microbiology, 186(7), pp. 2221–2223. doi: 10.1128/JB.186.7.2221-2223.2003.

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', *Metabolic Engineering*. Academic Press Inc., 40, pp. 104–114. doi: 10.1016/j.ymben.2017.01.007.

Little, G. T., Willson, B. J., Heap, J. T., Winzer, K. and Minton, N. P. (2018) 'The Butanol Producing Microbe *Clostridium beijerinckii* NCIMB 14988 Manipulated Using Forward and Reverse Genetic Tools', *Biotechnology Journal*. Wiley-VCH Verlag, 13(11), p. 1700711. doi: 10.1002/biot.201700711.

Liu, Z., Ying, Y., Li, F., Ma, C. and Xu, P. (2010) 'Butanol production by *Clostridium beijerinckii* ATCC 55025 from wheat bran', *Journal of Industrial Microbiology and Biotechnology*. Springer, 37(5), pp. 495–501. doi: 10.1007/s10295-010-0695-8.

Long, S., Jones, D. T. and Woods, D. R. (1984) 'Initiation of solvent

production, clostridial stage and endospore formation in *Clostridium acetobutylicum*', *erobiology Biotechnology © Springer*, 20, pp. 256–261.

López-Linares, J. C., García-Cubero, M. T., Coca, M. and Lucas, S. (2021) 'Efficient biobutanol production by acetone-butanol-ethanol fermentation from spent coffee grounds with microwave assisted dilute sulfuric acid pretreatment', *Bioresource Technology*, 320, p. 124348. doi: https://doi.org/10.1016/j.biortech.2020.124348.

Lütke-Eversloh, T. (2014) 'Application of new metabolic engineering tools for *Clostridium acetobutylicum*', *Applied Microbiology and Biotechnology*. doi: 10.1007/s00253-014-5785-5.

Ma, M., Li, J. and McClane, B. A. (2015) 'Structure-function analysis of peptide signaling in the *Clostridium perfringens* Agr-like quorum sensing system', *Journal of Bacteriology*. American Society for Microbiology, 197(10), pp. 1807–1818. doi: 10.1128/JB.02614-14.

Maddox, I. S. (1989) 'The Acetone-Butanol-Ethanol Fermentation: Recent Progress in Technology', *Biotechnology and Genetic Engineering Reviews*, 7(1), pp. 189–220. doi: 10.1080/02648725.1989.10647859.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D. and Lopez, R. (2019) 'The EMBL-EBI search and sequence analysis tools APIs in 2019', *Nucleic Acids Research*. doi: 10.1093/nar/gkz268.

Magal, R. (2019) Understanding the similarity and diversity of the

accessory gene regulator quorum sensing systems in in the genus *Clostridium*.

Al Makishah, N. H. and Mitchell, W. J. (2013) 'Dual substrate specificity of an n-acetylglucosamine phosphotransferase system in *Clostridium beijerinckii*', *Applied and Environmental Microbiology*. American Society for Microbiology, 79(21), pp. 6712–6718. doi: 10.1128/AEM.01866-13.

Martin, M. J., Clare, S., Goulding, D., Faulds-Pain, A., Barquist, L., Browne, H. P., Pettit, L., Dougan, G., Lawley, T. D. and Wren, B. W. (2013) 'The agr locus regulates virulence and colonization genes in *Clostridium difficile* 027', *Journal of Bacteriology*. American Society for Microbiology (ASM), 195(16), pp. 3672–3681. doi: 10.1128/JB.00473-13.

Martinussen, J. and Hammer, K. (1994) 'Cloning and characterization of upp, a gene encoding uracil phosphoribosyltransferase from *Lactococcus lactis*', *Journal of Bacteriology*. American Society for Microbiology, 176(21), pp. 6457–6463. doi: 10.1128/jb.176.21.6457-6463.1994.

Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J. B. R., Chen, Y., Zhou, X., Gomis, M. I., Lonnoy, E., Maycock, T., Tignor, M. and Waterfield, T. (2019) *Global warming of 1.5°C An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the*

global response to the threat of climate change, sustainable development, and efforts to eradicate poverty Edited by Science Officer Science Assistant Graphics Officer Working Group I Technical Support Unit.

Mehta, P., Goyal, S., Long, T., Bassler, B. L. and Wingreen, N. S. (2009) 'Information processing and signal integration in bacterial quorum sensing', *Molecular Systems Biology*. John Wiley & Sons, Ltd, 5(1), p. 325. doi: 10.1038/msb.2009.79.

Merritt, J., Qi, F., Goodman, S. D., Anderson, M. H. and Shi, W. (2003) 'Mutation of luxS affects biofilm formation in *Streptococcus mutans*', *Infection and Immunity*, 71(4), pp. 1972–1979. doi: 10.1128/IAI.71.4.1972-1979.2003.

Mijakovic, I., Denham, E., Claessen, D., Joël, H. and Soufo, D. (2016) 'A Novel Cell Type Enables B. subtilis to Escape from Unsuccessful Sporulation in Minimal Medium', *Frontiers in Microbiology | www.frontiersin.org*, 7, p. 1810. doi: 10.3389/fmicb.2016.01810.

Miller, M. B. and Bassler, B. L. (2001) 'Quorum Sensing in Bacteria', *Annual Review of Microbiology*. Annual Reviews, 55(1), pp. 165–199. doi: 10.1146/annurev.micro.55.1.165.

Miller, S. T., Xavier, K. B., Campagna, S. R., Taga, M. E., Semmelhack, M. F., Bassler, B. L. and Hughson, F. M. (2004) '*Salmonella typhimurium* recognizes a chemically distinct form of the bacterial quorum-sensing signal AI-2', *Molecular Cell*. doi:

10.1016/j.molcel.2004.07.020.

Minton, N. P., Ehsaan, M., Humphreys, C. M., Little, G. T., Baker, J., Henstra, A. M., Liew, F., Kelly, M. L., Sheng, L., Schwarz, K. and Zhang, Y. (2016) 'A roadmap for gene system development in *Clostridium*', *Anaerobe*. Academic Press, 41, pp. 104–112. doi: 10.1016/j.anaerobe.2016.05.011.

Mitchell, W. J. (2015) 'The Phosphotransferase System in Solventogenic Clostridia', *Journal of Molecular Microbiology and Biotechnology*. S. Karger AG, 25(2–3), pp. 129–142. doi: 10.1159/000375125.

Mitchell, W. J. (2016) 'Sugar uptake by the solventogenic clostridia', *World Journal of Microbiology and Biotechnology*. doi: 10.1007/s11274-015-1981-4.

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. and Soria, E. (2005) 'Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements', *Journal of Molecular Evolution*. Springer, 60(2), pp. 174–182. doi: 10.1007/s00239-004-0046-3.

Naik, S. N., Goud, V. V., Rout, P. K. and Dalai, A. K. (2010) 'Production of first- and second-generation biofuels: A comprehensive review', *Renewable and Sustainable Energy Reviews*. Pergamon, 14(2), pp. 578–597. doi: 10.1016/j.rser.2009.10.003.

Nair, R. V., Bennett, G. N. and Papoutsakis, E. T. (1994) 'Molecular characterization of an aldehyde/alcohol dehydrogenase gene from

Clostridium acetobutylicum ATCC 824', *Journal of Bacteriology*. American Society for Microbiology, 176(3), pp. 871–885. doi: 10.1128/jb.176.3.871-885.1994.

National Center for Biotechnology Information (NCBI) (2020) *Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information*. Available at: https://www.ncbi.nlm.nih.gov/ (Accessed: 20 October 2020).

Navarro, M. A., Li, J., Beingesser, J., McClane, B. A. and Uzal, F. A. (2020) 'The Agr-Like Quorum-Sensing System Is Important for *Clostridium perfringens* Type A Strain ATCC 3624 To Cause Gas Gangrene in a Mouse Model', *mSphere*. American Society for Microbiology, 5(3). doi: 10.1128/msphere.00500-20.

Neiditch, M. B., Capodagli, G. C., Prehna, G. and Federle, M. J. (2017) 'Genetic and Structural Analyses of RRNPP Intercellular Peptide Signaling of Gram-Positive Bacteria', *Annual Review of Genetics*. doi: 10.1146/annurev-genet-120116-023507.

Nguyen, N. P. T., Raynaud, C., Meynial-Salles, I. and Soucaille, P. (2018) 'Reviving the Weizmann process for commercial n-butanol production', *Nature Communications*. Nature Publishing Group, 9(1), pp. 1–8. doi: 10.1038/s41467-018-05661-z.

Nikolskaya, A. N. and Galperin, M. Y. (2002) 'A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family', *Nucleic Acids Research*, 30(11), pp. 2453–

2459.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K., Shimatani, Z. and Kondo, A. (2016) 'Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems', *Science (New York, N.Y.)*. Science, 353(6305). doi: 10.1126/SCIENCE.AAF8729.

Novick, R. P. and Geisinger, E. (2008) 'Quorum Sensing in *Staphylococci*', *Annual Review of Genetics*. Annual Reviews, 42(1), pp. 541–564. doi: 10.1146/annurev.genet.42.110807.091640.

Ohlsen, K. L., Grimsley, J. K. and Hoch, J. A. (1994) 'Deactivation of the sporulation transcription factor Spo0A by the Spo0E protein phosphatase', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 91(5), pp. 1756–1760. doi: 10.1073/pnas.91.5.1756.

Olson, M. E., Todd, D. A., Schaeffer, C. R., Paharik, A. E., Van Dyke, M. J., Bu¨ttner, H., Dunman, P. M., Rohde, H., Cech, N. B., Fey, P. D. and Horswill, A. R. (2014) '*Staphylococcus epidermidis* agr quorum-sensing system: Signal identification, cross talk, and importance in colonization', *Journal of Bacteriology*. American Society for Microbiology, 196(19), pp. 3482–3493. doi: 10.1128/JB.01882-14.

Ourisson, G., Albrecht, P. and Rohmer, M. (1984) 'The Microbial Origin of Fossil Fuels', *Scientific American*, 251(2), pp. 44–51. doi: 10.2307/24969433.

Papoutsakis, E. T. (2008) ‘Engineering solventogenic clostridia’, *Current Opinion in Biotechnology*. Elsevier Current Trends, 19(5), pp. 420–429. doi: 10.1016/j.copbio.2008.08.003.

Parashar, V., Aggarwal, C., Federle, M. J. and Neiditch, M. B. (2015) ‘Rgg protein structure - Function and inhibition by cyclic peptide compounds’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 112(16), pp. 5177–5182. doi: 10.1073/pnas.1500357112.

Parashar, V., Jeffrey, P. D. and Neiditch, M. B. (2013) ‘Conformational Change-Induced Repeat Domain Expansion Regulates Rap Phosphatase Quorum-Sensing Signal Receptors’, *PLoS Biology*, 11(3). doi: 10.1371/journal.pbio.1001512.

Paredes, C. J., Rigoutsos, I. and Papoutsakis, T. (2004) ‘Transcriptional organization of the *Clostridium acetobutylicum* genome’, *Nucleic Acids Research*. Oxford University Press, 32(6), pp. 1973–1981. doi: 10.1093/nar/gkh509.

Patakova, P., Branska, B., Sedlar, K., Vasylkivska, M., Jureckova, K., Kolek, J., Koscova, P. and Provaznik, I. (2019) ‘Acidogenesis, solventogenesis, metabolic stress response and life cycle changes in *Clostridium beijerinckii* NRRL B-598 at the transcriptomic level’, *Scientific Reports*. doi: 10.1038/s41598-018-37679-0.

Patakova, P., Kolek, J., Jureckova, K., Branska, B., Sedlar, K., Vasylkivska, M. and Provaznik, I. (2021) ‘Deeper below the surface—

transcriptional changes in selected genes of Clostridium beijerinckii in response to butanol shock', *MicrobiologyOpen*. Blackwell Publishing Ltd, 10(1). doi: 10.1002/mbo3.1146.

Perego, M., Hanstein, C., Welsh, K. M., Djavakhishvili, T., Glaser, P. and Hoch, J. A. (1994) 'Multiple protein-aspartate phosphatases provide a mechanism for the integration of diverse signals in the control of development in *B. subtilis*', *Cell*. doi: 10.1016/0092-8674(94)90035-3.

Perego, M. and Hoch, J. A. (1996) 'Cell-cell communication regulates the effects of protein aspartate phosphatases on the phosphorelay controlling development in *Bacillus subtilis'*, *Proceedings of the National Academy of Sciences of the United States of America*, 93(4). doi: 10.1073/pnas.93.4.1549.

Pesci, E. C., Milbank, J. B. J., Pearson, J. P., Mcknight, S., Kende, A. S., Greenberg, E. P. and Iglewski, B. H. (1999) 'Quinolone signaling in the cell-to-cell communication system of *Pseudomonas aeruginosa*', *Proceedings of the National Academy of Sciences of the United States of America*, 96(20), pp. 11229–11234. doi: 10.1073/pnas.96.20.11229.

Peters, J. M., Silvis, M. R., Zhao, D., Hawkins, J. S., Gross, C. A. and Qi, L. S. (2015) 'Bacterial CRISPR: Accomplishments and Prospects', *Current opinion in microbiology*. NIH Public Access, 27, p. 121. doi: 10.1016/J.MIB.2015.08.007.

Pfaffl, M. W. (2001) 'A new mathematical model for relative

quantification in real-time RT–PCR', *Nucleic Acids Research*. Oxford Academic, 29(9), pp. e45–e45. doi: 10.1093/NAR/29.9.E45.

Piatek, P., Humphreys, C., Raut, M. P., Wright, P. C., Simpson, S., Köpke, M., Minton, N. P. and Winzer, K. (2022) 'Agr Quorum Sensing influences the Wood-Ljungdahl pathway in *Clostridium autoethanogenum*', *Scientific Reports*. Nature Research, 12(1). doi: 10.1038/S41598-021-03999-X.

Piazza, F., Tortosa, P. and Dubnau, D. (1999) 'Mutational analysis and membrane topology of ComP, a quorum-sensing histidine kinase of *Bacillus subtilis* controlling competence development', *Journal of Bacteriology*. doi: 10.1128/jb.181.15.4540-4548.1999.

Piggot, P. J. and Hilbert, D. W. (2004) 'Sporulation of *Bacillus subtilis*', *Current Opinion in Microbiology*, 7(6), pp. 579–586. doi: 10.1016/j.mib.2004.10.001.

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. and Dürre, P. (2017) 'Microbial solvent formation revisited by comparative genome analysis', *Biotechnology for Biofuels*. BioMed Central Ltd., 10(1), p. 58. doi: 10.1186/s13068-017-0742-z.

Pollak, S., Omer-Bendori, S., Even-Tov, E., Lipsman, V., Bareia, T., Ben-Zion, I. and Eldar, A. (2016) 'Facultative cheating supports the coexistence of diverse quorum-sensing alleles', *Proceedings of the National Academy of Sciences of the United States of America*.

National Academy of Sciences, 113(8), pp. 2152–2157. doi: 10.1073/pnas.1520615113.

Pollitt, E. J. G., West, S. A., Crusz, S. A., Burton-Chellew, M. N. and Diggle, S. P. (2014) 'Cooperation, quorum sensing, and evolution of virulence in *Staphylococcus aureus*', *Infection and Immunity*. American Society for Microbiology Journals, 82(3), pp. 1045–1051. doi: 10.1128/IAI.01216-13.

Pottathil, M. and Lazazzera, B. A. (2003) 'The extracellular PHR peptide-rap phosphatase signaling circuit of *bacillus subtilis*', *Frontiers in Bioscience*, 8. doi: 10.2741/913.

Pourcel, C., Salvignol, G. and Vergnaud, G. (2005) 'CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies', *Microbiology*. Microbiology Society, 151(3), pp. 653–663. doi: 10.1099/mic.0.27437-0.

Qiu, R., Pei, W., Zhang, L., Lin, J. and Ji, G. (2005) 'Identification of the putative staphylococcal AgrB catalytic residues involving the proteolytic cleavage of AgrD to generate autoinducing peptide', *Journal of Biological Chemistry*, 280(17), pp. 16695–16704. doi: 10.1074/jbc.M411372200.

Qureshi, N. and Blaschek, H. P. (2001) 'Recent advances in ABE fermentation: Hyper-butanol producing *Clostridium beijerinckii* BA101', *Journal of Industrial Microbiology and Biotechnology*. doi:

10.1038/sj.jim.7000114.

Rader, B. A., Campagna, S. R., Semmelhack, M. F., Bassler, B. L. and Guillemin, K. (2007) 'The quorum-sensing molecule autoinducer 2 regulates motility and flagellar morphogenesis in *Helicobacter pylori*', *Journal of Bacteriology*. American Society for Microbiology Journals, 189(17), pp. 6109–6117. doi: 10.1128/JB.00246-07.

Ramos, J. L., Martínez-Bueno, M., Molina-Henares, A. J., Terán, W., Watanabe, K., Zhang, X., Gallegos, M. T., Brennan, R. and Tobes, R. (2005) 'The TetR Family of Transcriptional Repressors', *Microbiology and Molecular Biology Reviews*. American Society for Microbiology, 69(2), pp. 326–356. doi: 10.1128/mmbr.69.2.326-356.2005.

Ravagnani, A., Jennert, K. C. B., Steiner, E., Grünberg, R., Jefferies, J. R., Wilkinson, S. R., Young, D. I., Tidswell, E. C., Brown, D. P., Youngman, P., Gareth Morris, J. and Young, M. (2000) 'Spo0A directly controls the switch from acid to solvent production in solvent-forming *clostridia*', *Molecular Microbiology*, 37(5), pp. 1172–1185. doi: 10.1046/j.1365-2958.2000.02071.x.

Ren, C., Gu, Y., Wu, Y., Zhang, W., Yang, C., Yang, S. and Jiang, W. (2012) 'Pleiotropic functions of catabolite control protein CcpA in Butanol-producing *Clostridium acetobutylicum*', *BMC Genomics*, 13(1). doi: 10.1186/1471-2164-13-349.

Resnekov, O. and Losick, R. (1998) 'Negative regulation of the proteolytic activation of a developmental transcription factor in Bacillus

subtilis', *Proceedings of the National Academy of Sciences of the United States of America*. The National Academy of Sciences, 95(6), pp. 3162–3167. doi: 10.1073/PNAS.95.6.3162/ASSET/AEC1A7FA-EEE2-47C1-BAEC-0365C461252A/ASSETS/GRAPHIC/PQ0680170005.JPEG.

Reysenbach, A. L., Ravenschroft, N., Long, S., Jones, D. T. and Woods, D. R. (1986) 'Characterization, biosynthesis, and regulation of granulose in *Clostridium acetobutylicum*', *Applied and Environmental Microbiology*. American Society for Microbiology (ASM), 52(1), pp. 185–190. doi: 10.1128/aem.52.1.185-190.1986.

Rocha-Estrada, J., Aceves-Diez, A. E., Guarneros, G. and De La Torre, M. (2010) 'The RNPP family of quorum-sensing proteins in Gram-positive bacteria', *Applied Microbiology and Biotechnology*, pp. 913–923. doi: 10.1007/s00253-010-2651-y.

Roth, M., Goodall, E. C. A., Pullela, K., Jaquet, V., François, P., Henderson, I. R. and Krause, K. H. (2022) 'Transposon-Directed Insertion-Site Sequencing Reveals Glycolysis Gene *gpmA* as Part of the H2O2 Defense Mechanisms in *Escherichia coli'*, *Antioxidants*, 11(10). doi: 10.3390/antiox11102053.

Sandoval, N. R., Venkataramanan, K. P., Groth, T. S. and Papoutsakis, E. T. (2015) 'Whole-genome sequence of an evolved *Clostridium pasteurianum* strain reveals Spo0A deficiency responsible for increased butanol production and superior growth', *Biotechnology for Biofuels*. BioMed Central Ltd., 8(1), p. 227. doi: 10.1186/s13068-015-0408-7.

Scott, J. (2012) Characterisation and functional analysis of the putative agr system in *Clostridium acetobutylicum*.

Sedlar, K., Kolek, J., Gruber, M., Jureckova, K., Branska, B., Csaba, G., Vasylkivska, M., Zimmer, R., Patakova, P. and Provaznik, I. (2019) 'A transcriptional response of *Clostridium beijerinckii* NRRL B-598 to a butanol shock', *Biotechnology for Biofuels*. BioMed Central Ltd., 12(1), pp. 1–16. doi: 10.1186/S13068-019-1584-7/TABLES/5.

Seo, S.-O., Wang, Y., Lu, T., Jin, Y.-S. and Blaschek, H. P. (2017) 'Characterization of a *Clostridium beijerinckii spo0A* mutant and its application for butyl butyrate production', *Biotechnology and Bioengineering*. John Wiley and Sons Inc., 114(1), pp. 106–112. doi: 10.1002/bit.26057.

Seys, F. M., Humphreys, C. M., Tomi-Andrino, C., Li, Q., Millat, T., Yang, S. and Minton, N. P. (2023) 'Base editing enables duplex point mutagenesis in *Clostridium autoethanogenum* at the price of numerous off-target mutations', *Frontiers in Bioengineering and Biotechnology*. Frontiers Media SA, 11, p. 1211197. doi: 10.3389/FBIOE.2023.1211197/BIBTEX.

Shanker, E. and Federle, M. J. (2017) 'Quorum sensing regulation of competence and bacteriocins in *Streptococcus pneumoniae* and mutans', *Genes*. MDPI AG. doi: 10.3390/genes8010015.

Shanmugam, S., Ngo, H. H. and Wu, Y. R. (2020) 'Advanced CRISPR/Cas-based genome editing tools for microbial biofuels

production: A review', *Renewable Energy*. doi: 10.1016/j.renene.2019.10.107.

Shao, L., Hu, S., Yang, Yi, Gu, Y., Chen, J., Yang, Yunliu, Jiang, W. and Yang, S. (2007) 'Targeted gene disruption by use of a group II intron (targetron) vector in *Clostridium acetobutylicum*', *Cell Research*. Cell Res, 17(11), pp. 963–965. doi: 10.1038/cr.2007.91.

Sharma, H. K., Xu, C. and Qin, W. (2019) 'Biological Pretreatment of Lignocellulosic Biomass for Biofuels and Bioproducts: An Overview', *Waste and Biomass Valorization*. Springer Netherlands, 10, pp. 235–251. doi: 10.1007/s12649-017-0059-y.

SHELL (2020) *Biofuels | Shell Global*. Available at: https://www.shell.com/energy-and-innovation/new-energies-new/biofuels.html#iframe=L3dlYmFwcHMvMjAxOV9CaW9mdWVsc19pbnRlcmFjdGl2ZV9tYXAv (Accessed: 1 April 2020).

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 †', *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 74(24), pp. 7709–7714. doi: 10.1128/AEM.01948-08.

Sidote, D. J., Barbieri, C. M., Wu, T. and Stock, A. M. (2008) 'Structure of the Staphylococcus aureus AgrA LytTR Domain Bound to DNA Reveals a Beta Fold with a Novel Mode of Binding', *Structure (London, England: 1993)*. NIH Public Access, 16(5), p. 727. doi:

10.1016/J.STR.2008.02.011.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. and Higgins, D. G. (2011) 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular Systems Biology*, 7. doi: 10.1038/msb.2011.75.

Sims, R. E. H., Mabee, W., Saddler, J. N. and Taylor, M. (2010) 'An overview of second-generation biofuel technologies', *Bioresource Technology*. Elsevier Ltd, 101(6), pp. 1570–1580. doi: 10.1016/j.biortech.2009.11.046.

Solomon, S., Plattner, G. K., Knutti, R. and Friedlingstein, P. (2009) 'Irreversible climate change due to carbon dioxide emissions', *Proceedings of the National Academy of Sciences of the United States of America*, 106(6), pp. 1704–1709. doi: 10.1073/pnas.0812721106.

Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., Lawley, T. D., Sebaihia, M., Quail, M. A., Rose, G., Gerding, D. N., Gibert, M., Popoff, M. R., Parkhill, J., Dougan, G. and Wren, B. W. (2009) 'Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provide insight into the evolution of a hypervirulent bacterium', *Genome Biology*, 10(9), p. R102. doi: 10.1186/gb-2009-10-9-r102.

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*', *Molecular Microbiology*. NIH Public Access, 80(3), pp. 641–654. doi: 10.1111/j.1365-2958.2011.07608.x.

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*', *Applied and Environmental Microbiology*. doi: 10.1128/AEM.06376-11.

Strauch, M., Webb, V., Spiegelman, G. and Hoch, J. A. (1990) 'The SpoOA protein of *Bacillus subtilis* is a repressor of the abrB gene', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 87(5), pp. 1801–1805. doi: 10.1073/pnas.87.5.1801.

Sturme, M. H. J., Kleerebezem, M., Nakayama, J., Akkermans, A. D. L., Vaughan, E. E. and De Vos, W. M. (2002) 'Cell to cell communication by autoinducing peptides in gram-positive bacteria', *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 81, pp. 233–243. doi: 10.1023/A:1020522919555.

Sun, C., Zhang, S., Xin, F., Shanmugam, S. and Wu, Y. R. (2018) 'Genomic comparison of *Clostridium* species with the potential of utilizing red algal biomass for biobutanol production', *Biotechnology for Biofuels*. BioMed Central Ltd., 11(1). doi: 10.1186/s13068-018-1044-9.

Sun, G., Yang, M., Jiang, L. and Huang, M. (2021) 'Regulation of pro-σK activation: a key checkpoint in *Bacillus subtilis* sporulation',

Environmental Microbiology. John Wiley & Sons, Ltd, 23(5), pp. 2366–2373. doi: 10.1111/1462-2920.15415.

Sun, Z., Chen, Y., Yang, C., Yang, S., Gu, Y. and Jiang, W. (2015) 'A novel three-component system-based regulatory model for d-xylose sensing and transport in *Clostridium beijerinckii*', *Molecular Microbiology*. Blackwell Publishing Ltd, 95(4), pp. 576–589. doi: 10.1111/mmi.12894.

Tan, Y., Liu, Z. Y., Liu, Z., Zheng, H. J. and Li, F. L. (2015) 'Comparative transcriptome analysis between csrA-disruption *Clostridium acetobutylicum* and its parent strain', *Molecular BioSystems*. doi: 10.1039/c4mb00600c.

Thoendel, M. and Horswill, A. R. (2009) 'Identification of *Staphylococcus aureus* AgrD residues required for autoinducing peptide biosynthesis', *Journal of Biological Chemistry*, 284(33), pp. 21828–21838. doi: 10.1074/jbc.M109.031757.

Thoendel, M. and Horswill, A. R. (2013) 'Random mutagenesis and topology analysis of the autoinducing peptide biosynthesis proteins in *Staphylococcus aureus*', *Molecular Microbiology*. NIH Public Access, 87(2), pp. 318–337. doi: 10.1111/mmi.12100.

Thoendel, M., Kavanaugh, J. S., Flack, C. E. and Horswill, A. R. (2011) 'Peptide signaling in the Staphylococci', *Chemical Reviews*. American Chemical Society, 111(1), pp. 117–151. doi: 10.1021/cr100370n.

Torcato, I. M., Kasal, M. R., Brito, P. H., Miller, S. T. and Xavier, K. B.

(2019) ‘Identification of novel autoinducer-2 receptors in *Clostridia* reveals plasticity in the binding site of the LsrB receptor family’, *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc., 294(12), pp. 4450–4463. doi: 10.1074/jbc.RA118.006938.

Trindade, W. R. da S. and Santos, R. G. dos (2017) ‘Review on the characteristics of butanol, its production and use as fuel in internal combustion engines’, *Renewable and Sustainable Energy Reviews*. Elsevier Ltd, 69, pp. 642–651. doi: 10.1016/j.rser.2016.11.213.

Vidal, J. E., Ma, M., Saputo, J., Garcia, J., Uzal, F. A. and McClane, B. A. (2012) ‘Evidence that the Agr-like quorum sensing system regulates the toxin production, cytotoxicity and pathogenicity of *Clostridium perfringens* type C isolate CN3685’, *Molecular Microbiology*. doi: 10.1111/j.1365-2958.2011.07925.x.

Walter, K. A., Bennett, G. N. and Papoutsakis, E. T. (1992) ‘Molecular characterization of two *Clostridium acetobutylicum* ATCC 824 butanol dehydrogenase isozyme genes’, *Journal of Bacteriology*, 174(22), pp. 7149–7158. doi: 10.1128/jb.174.22.7149-7158.1992.

Wang, Y., Li, X., Mao, Y. and Blaschek, H. P. (2011) ‘Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq’, *BMC Genomics*, 12. doi: 10.1186/1471-2164-12-479.

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, p. 102. doi: 10.1186/1471-2164-13-102.

Wang, Y., Li, X., Milne, C. B., Janssen, H., Lin, W., Phan, G., Hu, H., Jin, Y. S., Price, N. D. and Blascheka, H. P. (2013) 'Development of a gene knockout system using mobile group II introns (Targetron) and genetic disruption of acid production pathways in *Clostridium beijerinckii*', *Applied and Environmental Microbiology*. American Society for Microbiology (ASM), 79(19), pp. 5853–5863. doi: 10.1128/AEM.00971-13.

Wang, Y., Zhang, Z. T., Seo, S. O., Lynn, P., Lu, T., Jin, Y. S. and Blaschek, H. P. (2016) 'Gene transcription repression in Clostridium beijerinckii using CRISPR-dCas9', *Biotechnology and Bioengineering*, 113(12). doi: 10.1002/bit.26020.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics', *Nature reviews. Genetics*. Nat Rev Genet, 10(1), pp. 57–63. doi: 10.1038/NRG2484.

Wen, Z., Lu, M., Ledesma-Amaro, R., Li, Q., Jin, M. and Yang, S. (2020) 'TargeTron Technology Applicable in Solventogenic Clostridia: Revisiting 12 Years' Advances', *Biotechnology Journal*. Wiley-VCH Verlag, 15(1). doi: 10.1002/biot.201900284.

Wen, Z., Minton, N. P., Zhang, Y., Li, Q., Liu, J., Jiang, Y. and Yang, S. (2017) 'Enhanced solvent production by metabolic engineering of a

twin-clostridial consortium', *Metabolic Engineering*. Academic Press Inc., 39, pp. 38–48. doi: 10.1016/j.ymben.2016.10.013.

Wiesenborn, D. P., Rudolph, F. B. and Papoutsakis, E. T. (1989) 'Coenzyme A transferase from *Clostridium acetobutylicum* ATCC 824 and its role in the uptake of acids.', *Applied and environmental microbiology*, 55(2). doi: 10.1128/aem.55.2.323-329.1989.

Wietzke, M. and Bahl, H. (2012) 'The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in *Clostridium acetobutylicum*', *Applied Microbiology and Biotechnology*. doi: 10.1007/s00253-012-4112-2.

Winzer, K., Hardie, K. R., Burgess, N., Doherty, N., Kirke, D., Holden, M. T. G., Linforth, R., Cornell, K. A., Taylor, A. J., Hill, P. J. and Williams, P. (2002) 'LuxS: Its role in central metabolism and the in vitro synthesis of 4-hydroxy-5-methyl-3(2H)-furanone', *Microbiology*. Society for General Microbiology, 148(4), pp. 909–922. doi: 10.1099/00221287-148-4-909.

Woolley, R. C. and Morris, J. G. (1990) 'Stability of solvent production by *Clostridium acetobutylicum* in continuous culture: strain differences', *Journal of Applied Bacteriology*, 69(5), pp. 718–728. doi: 10.1111/j.1365-2672.1990.tb01569.x.

Wu, Y. R., Li, Y., Yang, K. L. and He, J. (2012) 'Draft genome sequence of butanol-acetone-producing *Clostridium beijerinckii* strain G117', *Journal of Bacteriology*, 194(19), pp. 5470–5471. doi:

10.1128/JB.01139-12.

Xia, P. F., Casini, I., Schulz, S., Klask, C. M., Angenent, L. T. and Molitor, B. (2020) 'Reprogramming Acetogenic Bacteria with CRISPR-Targeted Base Editing via Deamination', *ACS Synthetic Biology*. American Chemical Society, 9(8), pp. 2162–2171. doi: 10.1021/ACSSYNBIO.0C00226/ASSET/IMAGES/MEDIUM/SB0C00226_M006.GIF.

Xin, F., Chen, T., Jiang, Y., Dong, W., Zhang, W., Zhang, M., Wu, H., Ma, J. and Jiang, M. (2017) 'Strategies for improved isopropanol-butanol production by a *Clostridium* strain from glucose and hemicellulose through consolidated bioprocessing', *Biotechnology for Biofuels*. BioMed Central Ltd., 10(1). doi: 10.1186/s13068-017-0805-1.

Xin, X., Cheng, C., Du, G., Chen, L. and Xue, C. (2020) 'Metabolic Engineering of Histidine Kinases in *Clostridium beijerinckii* for Enhanced Butanol Production', *Frontiers in Bioengineering and Biotechnology*. Frontiers, 8, p. 214. doi: 10.3389/fbioe.2020.00214.

Xu, T., Li, Yongchao, Shi, Z., Hemme, C. L., Li, Yuan, Zhu, Y., Van Nostrand, J. D., He, Z. and Zhou, J. (2015) 'Efficient genome editing in clostridium cellulolyticum via CRISPR-Cas9 nickase', *Applied and Environmental Microbiology*. American Society for Microbiology, 81(13), pp. 4423–4431. doi: 10.1128/AEM.00873-15/SUPPL_FILE/ZAM999116362SO1.PDF.

Xue, C. and Cheng, C. (2019) 'Butanol production by *Clostridium*', in.

Elsevier, pp. 35–77. doi: 10.1016/bs.aibe.2018.12.001.

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', *Applied Microbiology and Biotechnology*. doi: 10.1007/s00253-016-7638-x.

Yachdav, G., Wilzbach, S., Rauscher, B., Sheridan, R., Sillitoe, I., Procter, J., Lewis, S. E., Rost, B. and Goldberg, T. (2016) 'MSAViewer: Interactive JavaScript visualization of multiple sequence alignments', *Bioinformatics*. Oxford University Press, 32(22), pp. 3501–3503. doi: 10.1093/bioinformatics/btw474.

Yakovlev, A. V., Shalygin, M. G., Matson, S. M., Khotimskiy, V. S. and Teplyakov, V. V. (2013) 'Separation of diluted butanol-water solutions via vapor phase by organophilic membranes based on high permeable polyacetylenes', *Journal of Membrane Science*. Elsevier, 434, pp. 99–105. doi: 10.1016/j.memsci.2013.01.061.

Yamaguchi, Y. and Inouye, M. (2009) 'mRNA Interferases, Sequence-Specific Endoribonucleases from the Toxin-Antitoxin Systems', *Progress in Molecular Biology and Translational Science*, pp. 467–500. doi: 10.1016/S0079-6603(08)00812-X.

Yan, R.-T., Zhu, C.-X., Golemboski, C. and Chen, J.-S. (1988) 'Expression of Solvent-Forming Enzymes and Onset of Solvent Production in Batch Cultures of *Clostridium beijerinckii* ("*Clostridium butylicum*")', *Applied and Environmental Microbiology*, 54(3). doi:

10.1128/aem.54.3.642-648.1988.

Yang, Y., Nie, X., Jiang, Y., Yang, C., Gu, Y. and Jiang, W. (2018) ‘Metabolic regulation in solventogenic *clostridia*: regulators, mechanisms and engineering’, *Biotechnology Advances*. Elsevier Inc., 36(4), pp. 905–914. doi: 10.1016/j.biotechadv.2018.02.012.

Yu, Q., Lepp, D., Gohari, I. M., Wu, T., Zhou, H., Yin, X., Yu, H., Prescott, J. F., Nie, S. P., Xie, M. Y. and Gong, J. (2017) ‘The Agr-like quorum sensing system is required for pathogenesis of necrotic enteritis caused by *Clostridium perfringens* in poultry’, *Infection and Immunity*. American Society for Microbiology, 85(6). doi: 10.1128/IAI.00975-16.

Yusufaly, T. I. and Boedicker, J. Q. (2017) ‘Mapping quorum sensing onto neural networks to understand collective decision making in heterogeneous microbial communities’, *Physical Biology*. Institute of Physics Publishing, 14(4). doi: 10.1088/1478-3975/aa7c1e.

Zhang, C., Li, T. and He, J. (2018) ‘Characterization and genome analysis of a butanol-isopropanol-producing *Clostridium beijerinckii* strain BGS1 06 Biological Sciences 0605 Microbiology 06 Biological Sciences 0604 Genetics’, *Biotechnology for Biofuels*. BioMed Central Ltd., 11(1). doi: 10.1186/s13068-018-1274-x.

Zhang, L., Lin, J. and Ji, G. (2004) ‘Membrane Anchoring of the AgrD N-terminal Amphipathic Region Is Required for Its Processing to Produce a Quorum-sensing Pheromone in *Staphylococcus aureus*’,

Journal of Biological Chemistry, 279(19), pp. 19448–19456. doi: 10.1074/jbc.M311349200.

Zhang, Y., Zhang, J., Hoeflich, K. P., Ikura, M., Qing, G. and Inouye, M. (2003) 'MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in *Escherichia coli*', *Molecular Cell*. Cell Press, 12(4), pp. 913–923. doi: 10.1016/S1097-2765(03)00402-7.

Zhong, J. and Lambowitz, A. M. (2003) 'Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription', *EMBO Journal*. European Molecular Biology Organization, 22(17), pp. 4555–4565. doi: 10.1093/emboj/cdg433.

Zuleta-Correa, A., Chinn, M. S. and Bruno-Bárcena, J. M. (2022) 'Application of raw industrial sweetpotato hydrolysates for butanol production by *Clostridium beijerinckii* NCIMB 8052', *Biomass Conversion and Biorefinery*, 1, p. 3. doi: 10.1007/s13399-022-03101-z.

8: Appendix

8.1 Growth characterisation and solvent profile of *C. beijerinckii* NCIMB 8052 WT in CBM broth

Three individual colonies from the WT spore stock were used to separately inoculate 10 ml CBM broth and each diluted to 10^{-5} as pre-cultures, these were left to grow overnight. The overnight cultures were then used to inoculate triplicate 60 ml CBM broth cultures for the main growth experiment.

Figure 8.1 shows the growth curve for all three WT isolates was highly similar. The exponential stage of growth was observed between 0 and 12 hours of growth. The average growth rates between 3 and 10 hours, calculated using the OD_{600} measurements, were 0.41 h^{-1}, 0.39 h^{-1} and 0.42 h^{-1} for isolates A, B and C respectively.

The OD_{600} was observed to plateau by 24 hours at an average of 5.57. From here, the culture OD_{600} remain stable as they are in stationary phase.

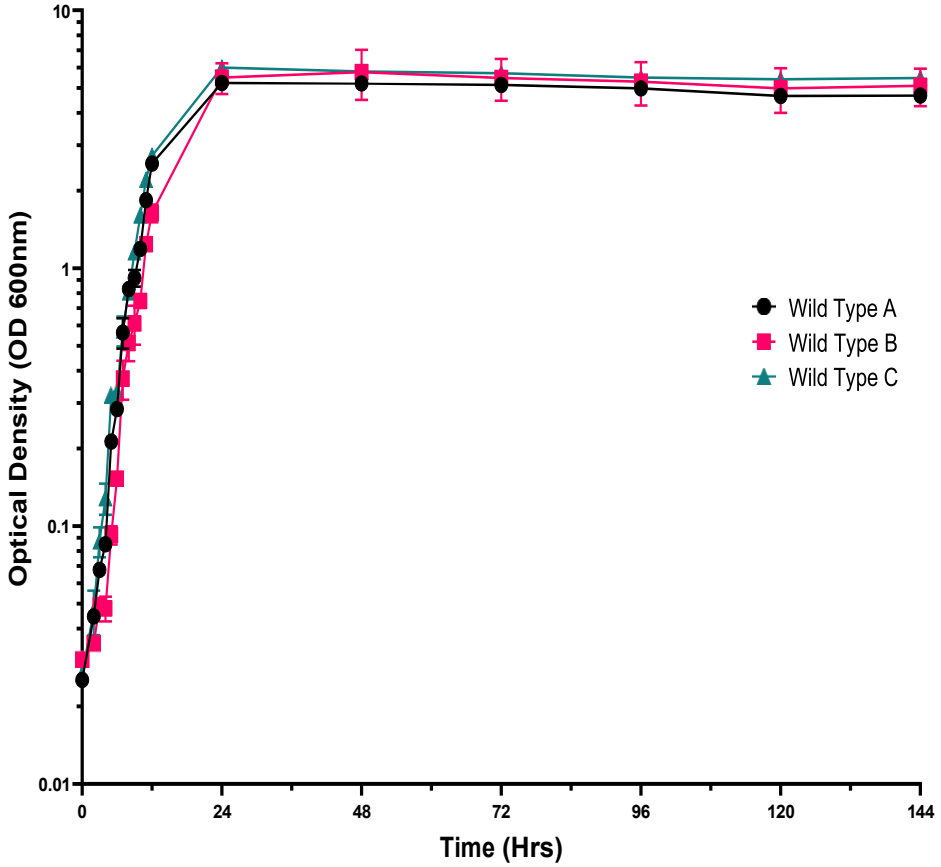


Figure 8.1 Growth of three independent *C. beijerinckii* NCIMB 8052 wild type isolates.

Single colony isolates WT A, B and C were obtained from germination of a single spore stock on CGM agar and used to create independent CBM starter cultures dilution series. Following overnight growth each was used to inoculate triplicate cultures for the above growth experiment. Data points represent means with standard error of the mean (SEM).

Figure 8.2 shows the fermentation product profiles for the three isolates were also highly similar. Glucose uptake was very little between 0 to 12 hours, in line with the amount of acids and solvents produced. The production of the acids acetate and butyrate was observed to begin increasing from the 6 hour time point, where it was first measured at concentrations of 2.4 mM and 1.7 mM, respectively. This outweighed the mean concentration of the solvents acetone, butanol and ethanol which, at 6 hours, had only reached 1.2 mM, 0.03 mM and 0.3 mM respectively. Glucose uptake drastically increased between 12 and 24 hours after inoculation, during which the later exponential phase of growth was occurring in **Figure 8.1**. Between 6 and 12 hours an active increase in the production of butanol, ethanol and acetone was detected, suggesting that the solventogenic phase was established between these time points. Butanol was observed to be present in the highest amount, with the peak concentration reaching a mean of 74 mM at 72 hours in isolate B. The acetone concentration peak was observed at 48 hours and did not exceed 15 mM. This coincides with the apparent reduced uptake of glucose by 48 and 72 hours, where the concentration of glucose left in the sample begins to plateau at around 200 mM. This suggests that metabolism from this point is very low. The concentration of solvents decreased in later stages, likely due to the switch from solventogenic to the sporogenic phase, meaning they were no longer being metabolised and likely lost due to higher volatility.

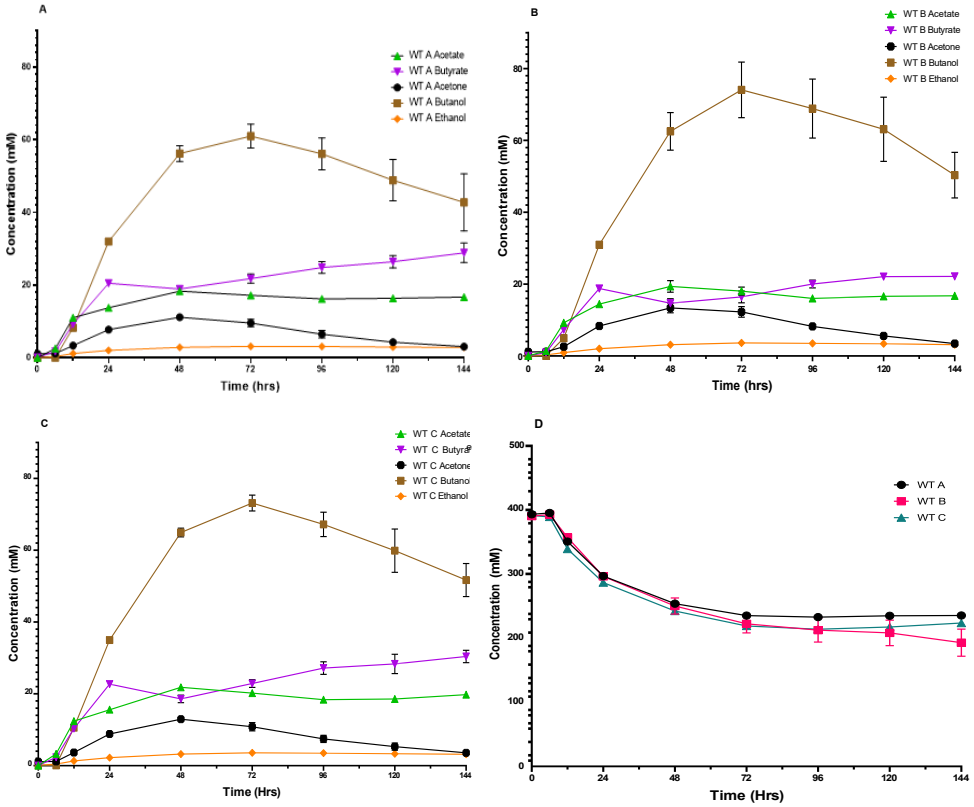


Figure 8.2 Fermentation product profiles for three independent *C. beijerinckii* NCIMB 8052 wild type isolates.

Butyrate, acetone, butanol, and ethanol concentrations as determined by gas chromatography for (A) WT A, (B) WT B and (C) WT C. High performance liquid chromatography was used to measure glucose concentration (D). The values plotted are the means with standard error of the mean (SEM) from three technical triplicates.

8.1.1 Granulose accumulation in the WT isolates

The accumulation of the starch-like polysaccharide granulose is observed in each of the three WT isolates. This was done as described previously. Granulose accumulation was indicated if the cell lawn showed a black colouration, whereas no granulose accumulation was indicated if the growth remained unstained. **Figure 8.3** shows the granulose assay carried out for the three WT isolates.

Figure 8.3 shows there was very little variation between all three of the WT isolates tested for granulose accumulation. All indicated high levels of granulose accumulation, demonstrated by the black colouration, on the outer area of the thick bacterial growth. However, these results also indicated that no or very little granulose accumulation occurred at the centre of these thick growth areas. This could be explained due to availability of nutrients. The build-up of granulose in solventogenic clostridia occurs just before the onset of sporulation and solventogenic clostridia are known to sporulate under nutrient excess (Gottschal & Morris, 1981; Long et al., 1984)The nutrients available at the centre of thick growth is likely to become depleted more quickly than those on the outer area of growth and therefore fewer nutrients are available to cells growing in the centre, meaning sporulation and granulose production here would be reduced.

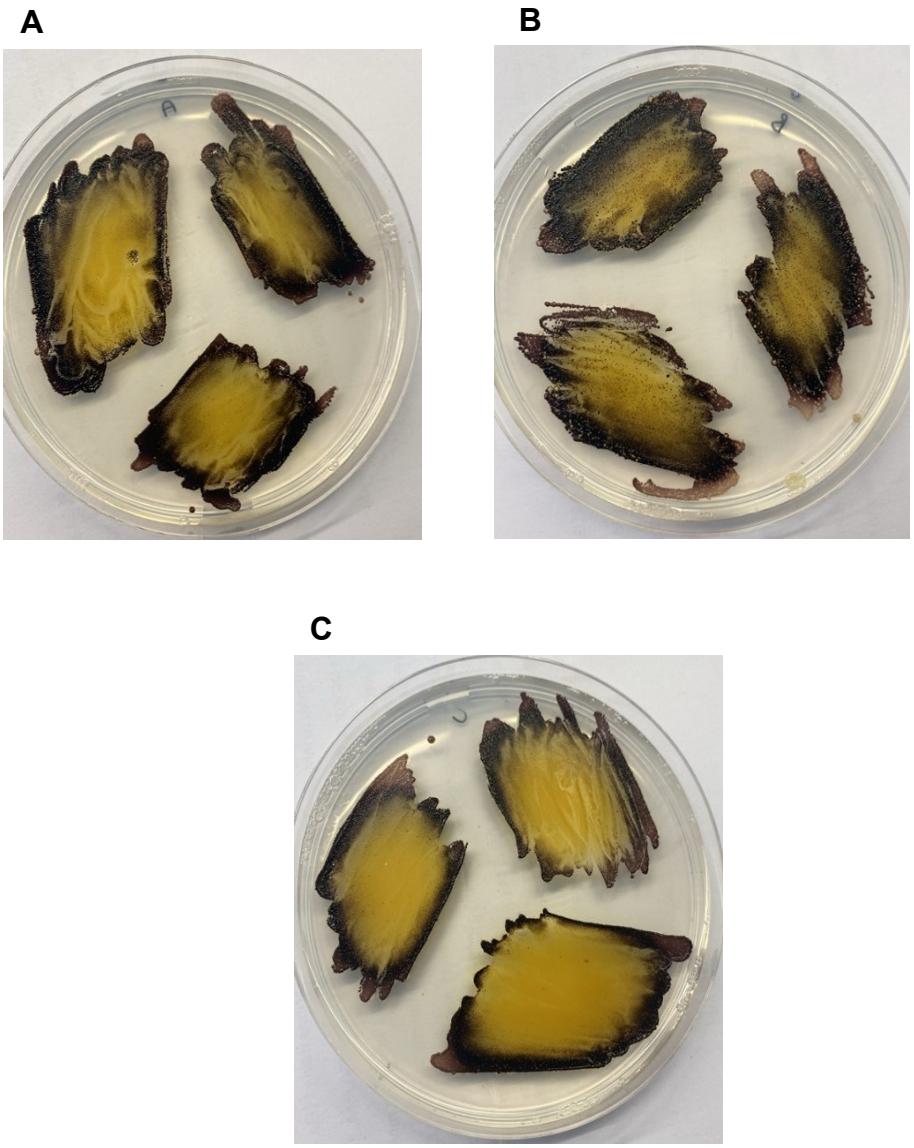


Figure 8.3 Granulose assay for three independent *C. beijerinckii* NCIMB 8052 wild type isolates.

The three WT isolates, indicated by A, B and C, were streaked in technical triplicate and allowed to grow for 48 hours. The accumulation of granulose was observed through staining with Lugol's iodine solution resulting in black colouration.

8.1.2 Sporulation in the *C. beijerinckii* NCIMB 8052 WT

To determine the number of heat resistant spores across the period of growth for each WT isolate, 100 µl of culture was heat shocked at 80°C for 10 minutes. This was then used to make a 1:10 serial dilution series and each was plated onto CGM agar. After 24 hours the CFU/ ml was calculated for the original culture.

The sporulation profiles for all three WT isolates appeared to be highly similar in **Figure 8.4**. A high number of viable heat resistant spores were observed for all three after 24 hours of growth in CBM broth. This coincides with the increased concentration of solvents after 24 hours in **Figure 8.2**, such as butanol and acetone present in the growth media, which alongside reduced pH, would initiate sporulation under nutrient excess. The heat resistant CFU/ml was shown to reach between 10^7 and 10^8 for all isolates.

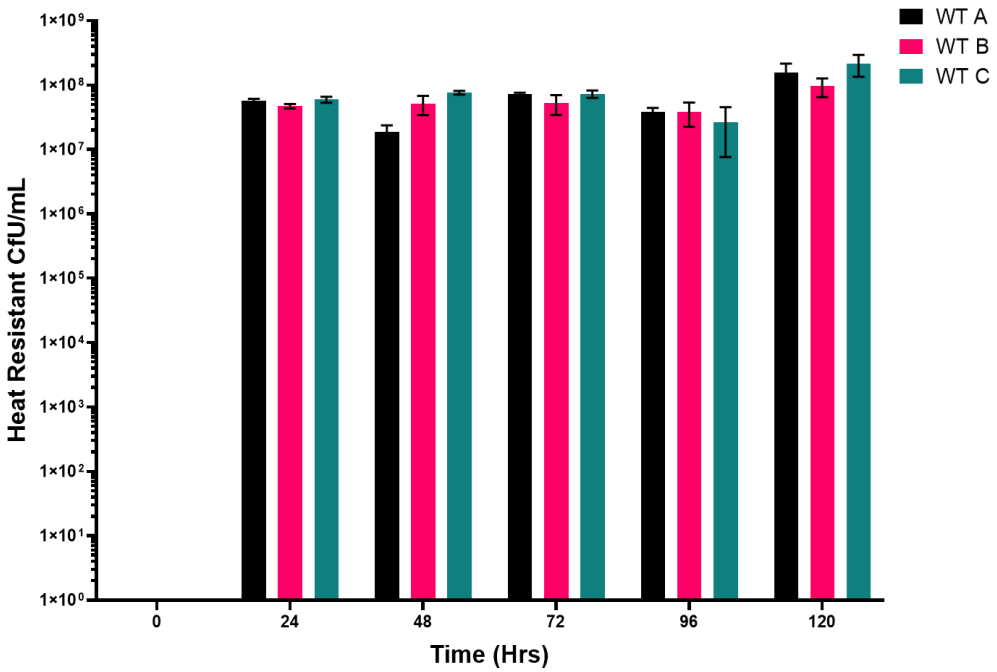


Figure 8.4 The calculated heat resistant CFU/ml *C. beijerinckii* NCIMB 8052 WTA, WTB and WTC isolates.

Heat resistant spore CFU/ml were calculated after 0, 24, 48, 72, 96 and 120 hours of growth in CBM broth (6% glucose). The values plotted are the means with standard error of the mean (SEM) from technical triplicates.

8.1.2 Growth characterisation and solvent profile of *C. beijerinckii* NCIMB 8052 ∆*agrD5* ClosTron mutant

The ∆*agrD5* ClosTron mutant and the WT comparison spore stocks were enumerated and then prepared for growth analysis as described in section 2.6.2.

The growth rates of the ∆*agrD5* ClosTron mutant and WT were highly similar (data not shown). However, the mutant was observed to plateau at a lower OD600 compared to the WT, this was also observed in repeated experiments (not shown). The mutant maintained an OD600 at around 5.5 whereas the WT maintained a higher one at around 6.5.

The accumulation of the storage polysaccharide granulose was analysed in the WT and *agrD5* ClosTron mutant Granulose accumulation was observed by Lugol's iodine staining of 48 hour old cell lawns on CBM agar lacking $CaCO_3$.

Figure 8.5 shows the granulose assay comparing the *agrD5* ClosTron mutant to the WT. After 48 hours the WT shows high levels of granulose accumulation, indicated by the growth staining black/dark brown. This was in contrast to the *agrD5* ClosTron mutant which appeared to produce no granulose in comparison, indicated by the growth staining a light yellow.

Chemical complementation of the AgrD5 derived AIP was attempted by streaking the WT in very close proximity to the *agrD* mutant. In theory, the AIP produced by the WT would diffuse through the agar and be active in the mutant. If this chemical complementation worked, the

mutant's granulose accumulation would be restored, indicated by black colouration. **Figure 8.5** shows that granulose accumulation was not restored in the mutant when grown in proximity to the WT.

The sporulation profiles for the ∆*agrD5* ClosTron mutant and WT comparison are shown in **Figure 8.6.** It was observed that after 120 hours of growth in CBM broth a 2-log reduction, with over 95% confidence, in heat CFU/ml was observed in the mutant compared to the WT. Attempts to chemically complement the endospore formation in the ClosTron *agrD mutant* gene proved difficult (not shown).

The lack of complementation suggested that these effects were not caused by lack of AIP signalling but most likely a results of polar effects or spontaneous mutation elsewhere on the chromosome. Further colony morphology observations suggested that this isolate was degenerate. Therefore, it was necessary to knock out more *agr* genes identified in the same genome using a method which was less likely to cause polar effects in the genome compared to ClosTron mutagenesis. The use of allelic exchange has already been described in previous studies (Scott, 2012; Humphreys, 2019). It was therefore decided that this would be attempted to produce in-frame targeted deletion of multiple *agrD* genes in the newly characterised strain, with ClosTron mutagenesis as an alternative for quick gene disruption.

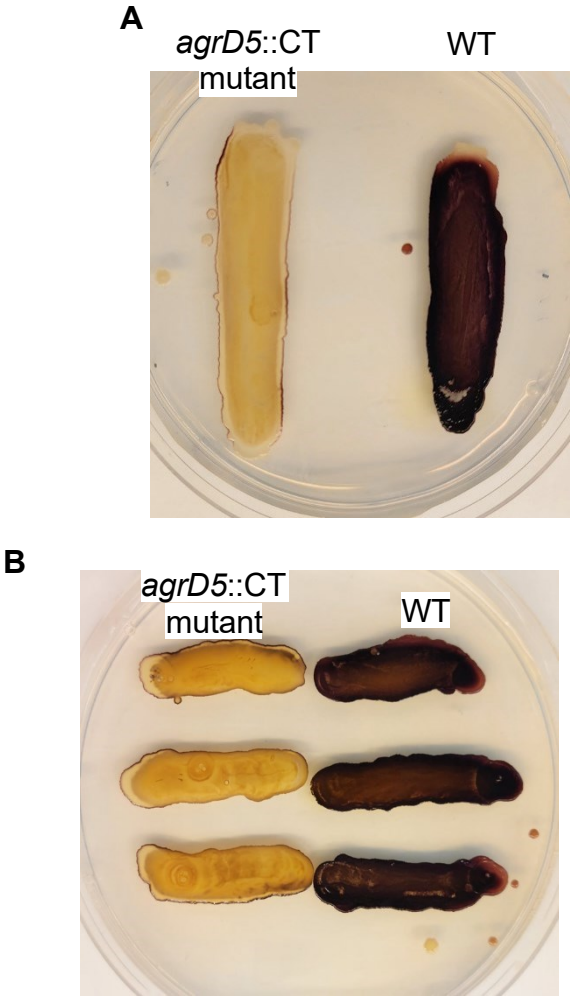


Figure 8.5 Granulose assay comparing the *C. beijerinckii* NCIMB 8052 wild type to the agrD5 ClosTron mutant.

The three WT and *agrD5* ClosTron mutant were heavily streaked onto CBM lacking $CaCO_3$ and allowed to grow for 48 hours. The accumulation of granulose was observed by dropping Lugols iodine solution onto the surface of the growth. The presence of granulose was confirmed by the black colouration observed.

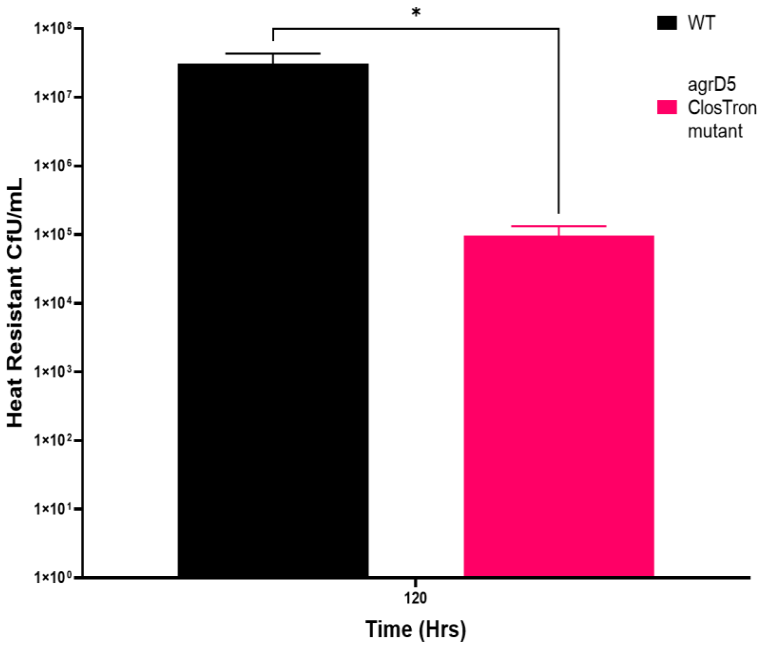


Figure 8.6 Heat resistant CFU/ml in the *C. beijerinckii* NCIMB 8052 *agrD5* ClosTron mutant and WT.

Heat resistant spore CFU/ml were calculated after 120 hours of growth in CBM broth (6% glucose). The values plotted are the means with standard error of the mean (SEM) from technical triplicates ($p \leq 0.05$ (*)).

8.2 Plasmid maps of the main vectors used in this study

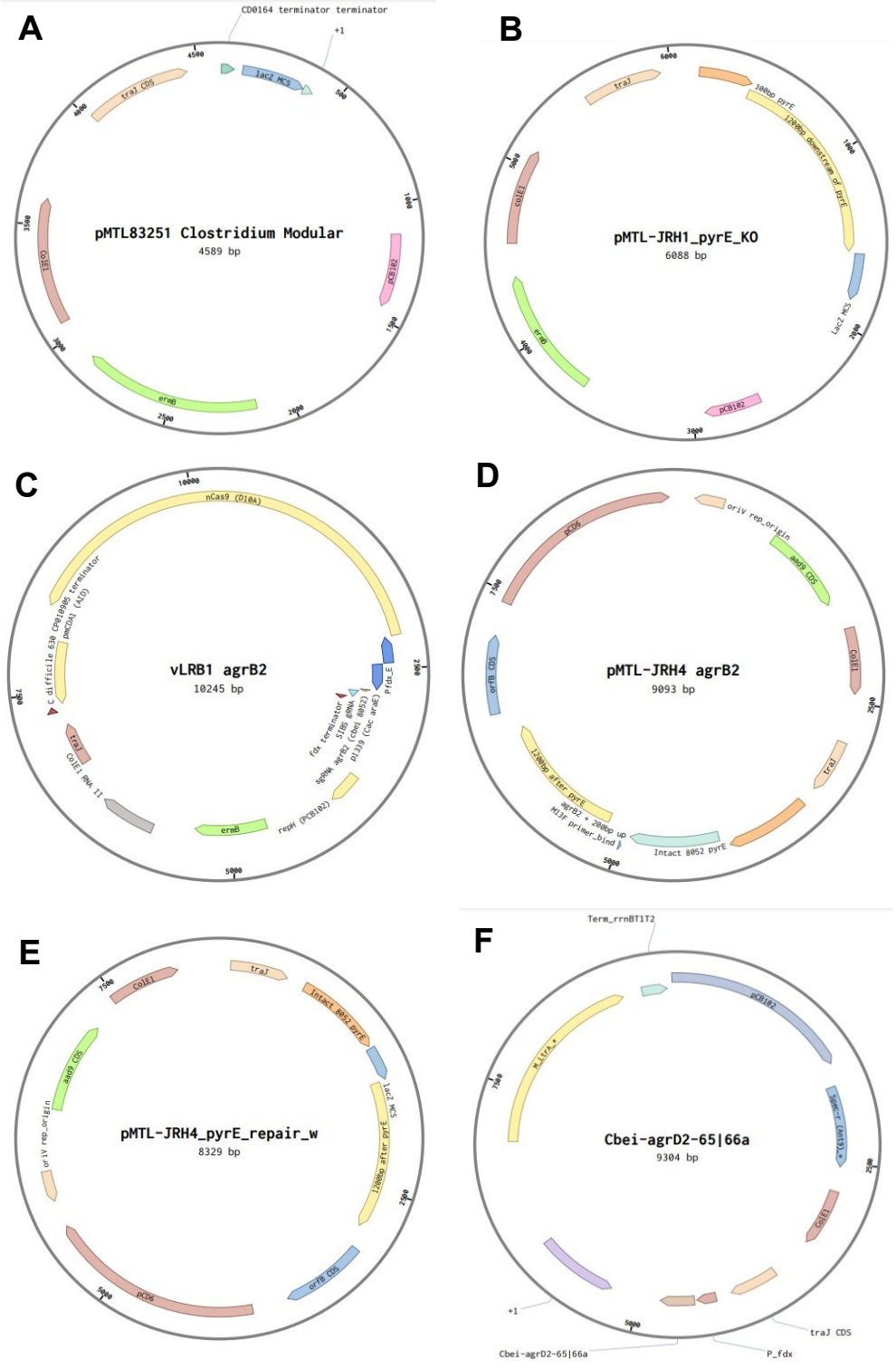


Figure 8.7 Plasmid maps of the main vectors used throughout this project.

(A) pMTL83251 vector, (B) pMTL-JRH1_pyrE_KO vector, (C) vLRB1 agrB2 ko vector, (D) pMTL-JRH4 agrB2 vector, (E) pMTL-JRH4_pyrE_repair vector, (F) Cbei-agrD2-65|66a ClosTron vector.

8.3 RNA-seq data for the *agrB* mutants with differential gene expression compared to that of the WT at either 12 hours or 24 hours.

The full RNA-seq data is available in the public research data of the Nottingham research repository under the DOI: 10.17639/nott.7351.

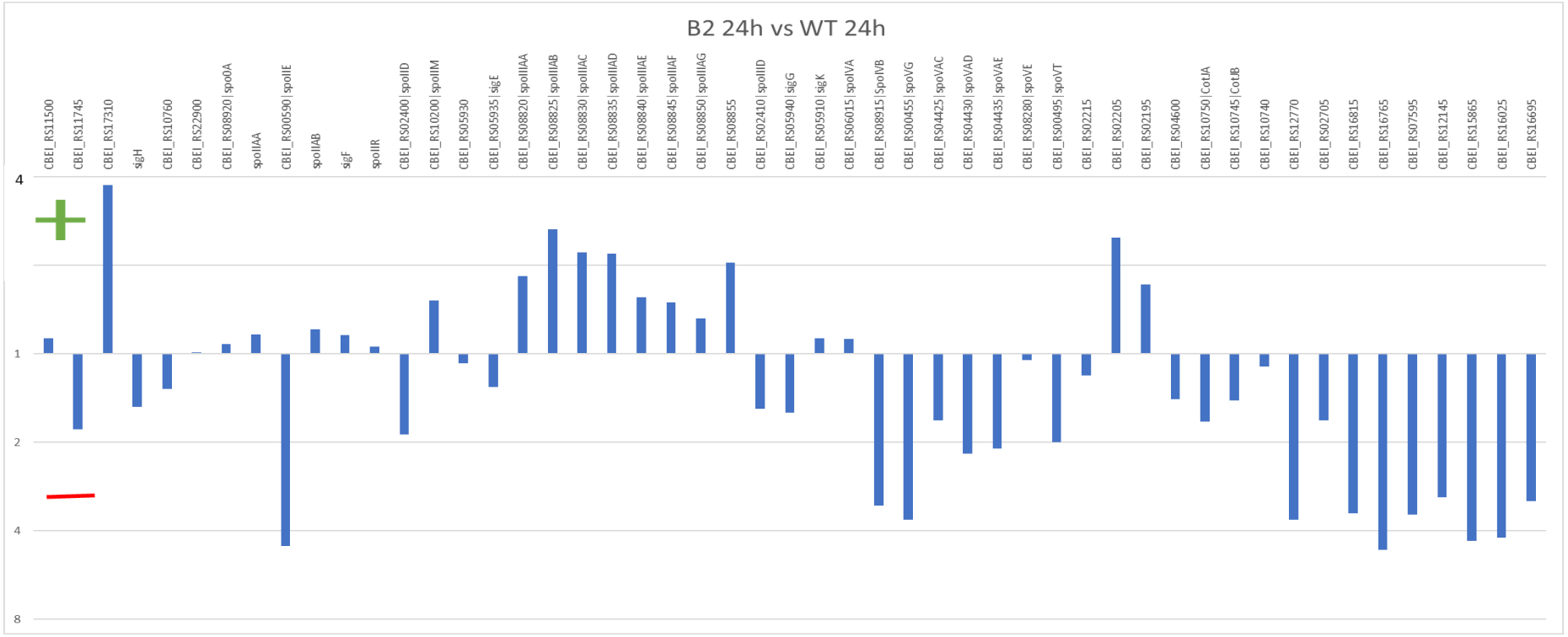


Figure 8.8 Change in key spore-formation gene expression in the *agrB2* mutant compared to the WT at 24 hours of growth.

The genes identified to be involved in *C. beijerinckii* sporulation were placed in order of early to late expression during the sporulation cycle, using Diallo et al. (2021) as a reference. The Fold change in gene expression, *agrB2* mutant versus WT, from the RNA-seq data was plotted for each gene. The green + indicated upregulation in the *agrB2* mutant compared to the WT whilst the red – sign represents downregulation in the mutant.

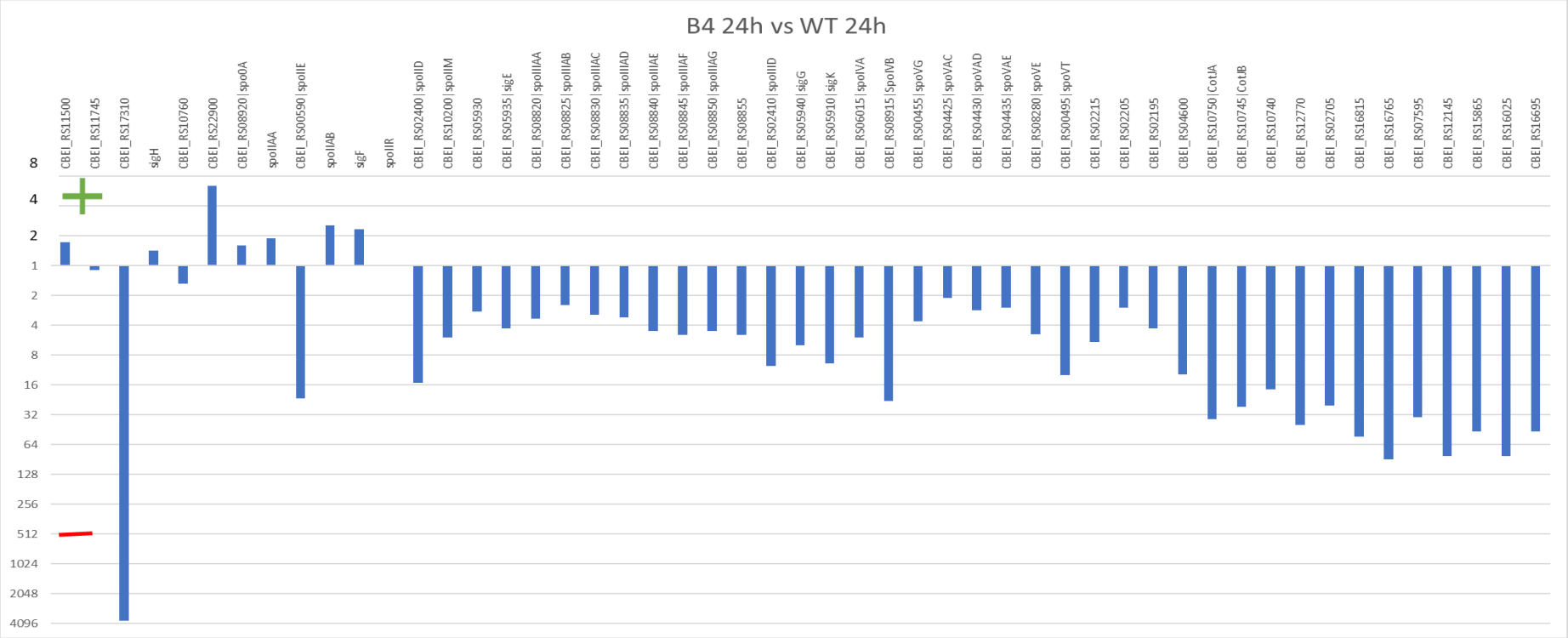


Figure 8.9 Change in key spore-formation gene expression in the *agrB4* mutant compared to the WT at 24 hours of growth.

The genes identified to be involved in C. *beijerinckii* sporulation were placed in order of early to late expression during the sporulation cycle, using Diallo et al. (2021) as a reference. The Fold change in gene expression, *agrB4* mutant versus WT, from the RNA-seq data was plotted for each gene. The green + indicated upregulation in the *agrB4* mutant compared to the WT whilst the red – sign represents downregulation in the mutant.