Investigating the Bile Acid Mediated Control of *Clostridioides difficile* Infection

Louise Megan von Emloh, BSc MSc

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

September 2023

Abstract

Clostridioides difficile is a spore forming, anaerobic Gram-positive bacterium. If ingested, spores can cause *C. difficile* associated disease (CDAD), with symptoms ranging from diarrhoea to pseudomembranous colitis. Germination of spores in the gut is both induced and inhibited by primary and secondary hepatic bile acids, respectively, and the biotransformation between the two structures by 7α-dehydroxylation, encoded by bile acid inducible (*bai*) genes, is therefore key in the onset of CDAD. Colonic bacteria capable of 7α-dehydroxylation are limited and are mainly from the Clostridia family. These species, particularly *Clostridium scindens*, have been implicated in bile acid-mediated colonisation resistance against *C. difficile*. Unfortunately, research into *C. scindens* has been hindered by its genetic intractability.

This study investigated the use of genome editing tools in the native 7α-dehydroxylating species *Peptacetobacter hiranonis*, with the aim of generating loss-of-function mutations in *bai* genes. Gene transfer into *P. hiranonis* was established and current editing tools were assessed, but their lack of function required an expansion of the genetic toolkit in this novel organism. Reporter assays were established and utilised to harness the functionality of native promoters, in addition to the development of inducible promoters through the use of riboswitches. Whilst a reproducible CRISPR-Cas system was unable to be developed, a *ΔbaiCD* strain was produced and was shown to be incapable of 7α-dehydroxylation. This strain was further studied in *in vitro* competition assays with *C. difficile* to explore bile acid-mediated colonisation resistance.

Clostridium butyricum, a gut commensal, was chosen as a chassis strain to study the 7α-dehydroxylation pathway. The genome editing tools allele-exchange and RiboCas, an inducible CRISPR-Cas9 system, were first established in the strain and used to develop a triple auxotrophic knockout capable of cargo insertion at three loci. Mobilisation of the *bai* genes from *C. scindens* into *C. butyricum* was investigated, with expression driven by promoters characterised for use in *C. butyricum* in this study. Whilst it was not possible to do so for the full pathway, the *bai* genes suggested to be essential for 7α-dehydroxylation were inserted into the *C. butyricum* genome, and gene expression was confirmed. Bile acid-mediated inhibition of *C. difficile*

germination by the strain was not demonstrated, but assessment of its ability to 7α-dehydroxylate is ongoing.

COVID-19 Impact Statement

The COVID-19 pandemic had an impact on this work for several reasons. Firstly, at the onset of the pandemic, with very limited notice, the laboratory spaces at the University of Nottingham were closed for a total of 15 weeks, between March and July 2020. This prevented the undertaking of any practical work. Having only begun the project in October 2020, there was only a limited amount of *in silico* work and data analysis available to make use of the time working from home. It also interrupted the learning of skills and techniques that a first year PhD student would usually undertake.

Upon returning to the laboratory it was not possible to resume normal working standards due to the COVID-19 measures introduced. To reduce the number of colleagues in the lab to accommodate social distancing, bench spaces and offices were reduced and reallocated. A shift pattern was introduced between July 2020 and October 2021, alternating morning and afternoon working on a weekly basis, with no option for out of hours working. On a personal level, this severely restricted the time available to carry out practical work, impacting time-sensitive experiments such as growth curves and conjugations, for example, and reducing the flexibility required when culturing bacteria. Moreover, as a first-year student the isolation and social distancing made it more difficult to access the expertise of colleagues to carry out training in laboratory techniques. The restricted working also impacted this project because of its effect on the day-to-day running of the laboratory; many services such as autoclave runs and sample collections were restricted to the morning only, making them impossible to access on the afternoon shift and causing delays to work.

The wider effects of the pandemic, as well as the impacts of Brexit, also caused disruption to this project, particularly in the supply chains. Shipping of reagents and consumable suffered severe delays, in addition to complete shortages of some items. The Sanger Sequencing services were also widely affected, with shipping and customs issues resulting in the delivery of results increasing from a 24-hour turnaround to up to a week. There were also occasions where samples got lost completely.

Whilst the thesis plan was reviewed and adapted to accommodate its impact, the COVID-19 pandemic and associated issues prevented an efficient use of time in the first two years of this project, and reduced the level of experimental output.

Acknowledgements

I am grateful to many people who have helped me complete this challenging yet rewarding PhD journey.

To my supervisors Nigel Minton and Klaus Winzer for giving me the opportunity to study for this PhD and their mentorship throughout. I would also like to thank ATUM and the Swiss National Science Foundation (SNSF) for providing the funding for this research, as well as our Sinergia project colleagues for their collaboration and input.

A massive thank you to Patrick for teaching me everything there is to know in the lab, providing endless guidance and patience, and always being on hand to reassuringly tackle any problem. And for helping make the COVID shifts slightly more bearable.

To everyone in the SBRC, in B27 and B39, for always lending an ear and so readily offering help and advice when I have asked for it.

To my amazing friends in B50, Temi, Lauren, Liam, Margaux and Danielle, thank you for being there to turn to for advice, support and laughter. We (just about) kept each sane through the highs and lows of doing a PhD during a pandemic. We are lucky that we put up with each other, as I don't think anyone else would.

I will always be grateful for the unwavering belief and encouragement of my family and friends which has supported me through my career so far.

To my (now) husband Dan. For the late-night cooking after long days in the lab, lockdown shift pick-ups, and for listening to my complaints about badly behaved bacteria when you didn't have a clue what I was talking about. It's safe to say I couldn't have done it without you.

Table of Contents

List of Figures

List of Tables

Abbreviations

5-FUMP	5-fluorouridine monophosphate
5'-UTR	5'-untranslated region
ACE	Allele Coupled Exchange
AE	Allelic Exchange
ANI	Average nucleotide identity
ANOVA	Analysis of Variance
ATCC	American Type Culture Collection
BA	Bile Acid
bai	Bile Acid Inducible
BE	essential *bai* genes
BHIS	Brain Heart Infusion Supplemented
BHISS	Brain Heart Infusion Salt-Supplemented
BLAST	Basic Local Alignment Search Tool
BM	Bookmark
bp	Base pairs
BREX	Bacteriophage Exclusion
BSH	Bile Salt Hydrolases
CA	Cholic Acid
Cas	CRISPR associated
CAT	Chloramphenicol Acetyltransferase
CDAD	*Clostridioides difficile* Associated Disease
CDCA	Chenodeoxycholic Acid
CDI	Clostridioides difficile Infection
CDT	Clostridioides difficile Binary Toxin
CFU	Colony Forming Unit(s)

CoA	Coenzyme A
CRISPR	Clustered Regularly Interspaced Palindromic Repeat
dam	Deoxyadenosine methyltransferase
DCA	Deoxycholic Acid
dcm	Deoxycytosine methyltransferase
DMSO	Dimethylsulfoxide
DSM	German Collection of Microorganisms
EDTA	Ethylenediaminetetraacetic acid
FAST	Fluorescence-Activating and Absorption-Shifting Tag
FBS	Fetal Bovine Serum
FOA	Fluoroorotic Acid
HSDH	Hydroxysteroid dehydrogenase
HTH	helix-turn-helix
Kb	Kilobase pair
KO	Knock-out (gene deletion)
LAC	LAC system from C. perfringens (PbgaL and bgaR)
LB	Lysogeny Broth
LCA	Lithocholic Acid
LHA	Left Homology Arm
MCS	Multiple Cloning Site
MM	Minimal media
mMM	Modified minimal media
NCTC	National Collection of Type Cultures
NEB	New England Biolabs
OD	Optical density
OMP	5’-monophosphate
OTU	Operational Taxonomic Unit

PacBio	Pacific Biosciences
PaLoc	Pathogenicity Locus
PAM	Protospacer Adjacent Motif
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
pMTL	Plasmid Microbial Technology Laboratory
RBS	Ribosomal binding site
RCM	Reinforced Clostridial Media
REase	Restriction Endonuclease
RHA	Right Homology Arm
RM	Restriction Methylation
sgRNA	single-guide RNA
SMRT	Single Molecule Real Time
SNP	Single Nucleotide Polymorphism
SOE	Splicing by Overlap Extension
ss	single stranded
TCA	Taurocholic Acid
TSS	Transcriptional start site
WT	Wildtype

Chapter 1 Introduction

1.1 *Clostridioides difficile*

1.1.1 Clostridioides difficile *Phylogeny and Epidemiology*

Clostridioides difficile is a spore forming, anaerobic Gram-positive bacterium, first isolated in 1935 from the stools of healthy infants (Hall, 1935). It is a leading cause of nosocomial infection worldwide, with *Clostridioides difficile* infection (CDI) posing a global public health challenge (Cooper, Jump, & Chopra, 2016; Sartelli et al., 2019). In England between April 2021 and March 2022 there were 14,249 cases of CDI reported, with 25.2 cases per 100,000 of the population (UK Health Security Agency, 2022). In North America CDI is classified as an urgent threat level to human health; in 2019 there were 202,600 cases of the disease, causing an estimated 11,500 deaths and costing approximately 1 billion dollars (CDC, 2019). Whilst the rapid decrease and stabilisation of both total and hospital-onset cases in both the UK and North America since 2007 suggests that infection interventions had achieved some degree of success, CDI still presents a challenge.

First known as *Bacillus difficilis* due to the difficulties in culturing it, *C. difficile* was originally assigned to *Clostridium* as it displayed the broad phenotypes characteristic of the genus; anaerobic, Gram-positive, spore-forming rods (Lancet Infectious Diseases, 2019). Interestingly, a distinguishing feature of *C. difficile* is its ability to decarboxylate parahydroxyphenylacetic acid and produce *p*-cresol, giving it a characteristic smell of tar or pigs (Hafiz & Oakley, 1976; Smits, Lyras, Lacy, Wilcox, & Kuijper, 2016). However, since the advent of molecular methods such as DNA-rRNA pairing and 16S rRNA analysis revealed the true phylogenetic diversity of the *Clostridium* genus it has undergone, and is still undergoing, updates to its taxonomy (Lawson, Citron, Tyrrell, & Finegold, 2016). In 2015 it was proposed that the original rRNA cluster I represent the true genus *Clostridium*, containing the type species *Clostridium butyricum* and related species (Lawson & Rainey, 2016). However, with *C. difficile* phylogenetically distant from cluster I and located in cluster XI, it required reclassification to a new genus, and phylogenetic analysis demonstrated that the type strain *C. difficile* ATCC 9689 is located in *Peptostreptococcaceae*. The biggest issue with this was one of familiarity; the ‘C’ in *C. difficile* has far reaching use and recognition, particularly but not restricted to the medical profession. Because of this, the name *Clostridioides difficile* was officially adopted, but whether this is practically

the case remains to be seen (Elliott, Androga, Knight, & Riley, 2017; Lancet Infectious Diseases, 2019; Lawson et al., 2016).

Understanding the large-scale epidemiology of CDI can pose a challenge due to the variety of typing methods used, and this has occurred on a largely geographical basis as a result of differing resources and technologies (Elliott et al., 2017). Previously used phenotyping methods such as immunoblotting and serotyping have been replaced with genotyping methods, and band-based typing methods are now the norm. These techniques include: restriction enzyme analysis (REA) which produces many fragments that can be difficult to interpret; pulsed field gel electrophoresis (PFGE) which allows the separation of large fragments; and PCR ribotyping which amplifies the spacer regions in 16S and 23S ribosomal RNA (Elliott et al., 2017; Rupnik, Wilcox, & Gerding, 2009; Smits et al., 2016). Historical preferences for REA and PFGE in North America, and PCR ribotyping in Europe, has resulted in epidemic strains being designated multiple typing monikers (Smits et al., 2016). In the early 2000s there was an emergence of hypervirulent strains of *C. difficile* belonging to the BI/NAP1/027 type in North America and the UK, and these were associated with an increase in antibiotic resistance and disease severity (Burns, Heeg, Cartman, & Minton, 2011; Rajani Thanissery, Winston, & Theriot, 2017). Whilst the rising rates of CDI that followed were attributed to the hypervirulent strains, in 2012 it had become less prevalent in the UK and rates were declining (Wilcox et al., 2012), although it has spread to other countries such as China, Japan and South Korea (Cairns et al., 2017). The emerging strain RT 078 has also been attributed to increased rates, and was the third most common strain in Europe in 2011 (Bauer et al., 2011). Whilst other strains have been associated with other geographical areas, for example RT 017 is consistently found in Asia, the lack of epidemiological data outside of the UK, North America and Australia make documenting *C. difficile* on a global basis difficult (Elliott et al., 2017).

1.1.2 Clostridioides difficile *Infection (CDI)*

C. difficile is spread via the oral-faecal route and CDI occurs as a result of spore transmission, survival of the acidic environment of the stomach, and germination in the intestine (Czepiel et al., 2019). Infection occurs with a range of severity. At the minor end of the clinical spectrum, asymptomatic colonisation of *C. difficile* occurs at

a carriage rate of 0-15% in healthy adults in the general population, and may form a potential reservoir for infection (Furuya-Kanamori et al., 2015). The disease state, however, can manifest from the mild diarrhoea most common to CDI, through to pseudomembranous colitis and toxic megacolon, which can result in sepsis and death (Heinlen & Ballard, 2010; Schäffler & Breitrück, 2018). The death rate of CDI is estimated at 17% (Crogan & Evans, 2007).

Traditionally *C. difficile* is considered a nosocomial pathogen, with reservoirs present in patients and contaminated hospital environments, and hospitalisation is one of the risk factors for development of the disease. Antibiotic usage is the main risk factor for CDI, and old age can also increase the risk; the risk is 8-10 fold higher during almost all classes of antimicrobial therapy, and 5- to 10-fold higher in patients over the age of 65, in addition to an increased risk of mortality. These risk factors are likely to be present in hospitalised individuals and increase the likelihood of infection due to alterations in the gut microbiome (Czepiel et al., 2019; Evans & Safdar, 2015; Heinlen & Ballard, 2010). Previous occurrence of CDI is also a risk factor in itself; whilst the standard risk of recurrence following 8 weeks of treatment cessation is 15-25%, this increases to 40-65% if a patient has previous history of CDI (Kelly, 2012). However, the incidence of CDI in the community is also increasing, despite these populations lacking the traditional risk factors, with estimates as high as 66% of all cases in the UK (Public Health England, 2019). Although this may be in part due to changes in awareness and definitions of community-onset, various factors such as acid suppression medications, asymptomatic carriers, and environmental sources have been suggested to play a role (Gupta & Khanna, 2014).

Toxins

Upon original isolation, despite evidence of animal infections, *C. difficile* was not considered to be pathogenic in humans. In the following decades there were few isolations of the bacterium in humans, and even when isolation was successful its pathogenic role was dismissed (Smith & King, 1962). Moreover, *Staphylococcus aureus* was misidentified as the pathogen responsible for pseudomembranous colitis, an emerging complication of clindamycin treatment (Bartlett JG., 1994; Elliott et al., 2017). It was not until 1978 that the role of *C. difficile* in human disease was recognised, with the organism and its toxins finally implicated in antibiotic associated

pseudomembranous colitis (Bartlett, Chang, Gurwith, Gorbach, & Onderdonk, 1978; Lance George, Goldstein, Sutter, Ludwig, & Finegold, 1978).

Most of the clinical symptoms of toxinogenic CDI can be attributed to one or both of its two main toxins, toxin A and toxin B. Produced by the genes *tcdA* and *tcdB* located in a 19.6 kb pathogenicity locus (PaLoc), TcdA and TcdB belong to the Large Clostridial Glycosylating Toxin (LCGT) family and are single chained proteins at 308 and 270 kDa respectively. Their structure consists of three main domains: an amino-terminal binding domain, a carboxy-terminal catalytic domain, and a putative translocation domain (Genth, Dreger, Huelsenbeck, & Just, 2008; Rupnik et al., 2009). LCGTs render the small GTPases of the Rho and Ras families inactive through glycosylation, resulting in the cytotoxic and proinflammatory effects of the toxins. Their inactivation disrupts the actin skeleton and impairs tight junctions, resulting in fluid accumulation and damage to the epithelium. Toxin A acts mainly in the intestine, whereas toxin B has a broader range of targets (Lyras et al., 2009; Rupnik et al., 2009). The expression of both toxins is dependent on global regulators and environmental conditions, for example nutrients and temperature changes, and *C. difficile* begins to secrete them upon entering a vegetative state in the intestine (Di Bella, Ascenzi, Siarakas, Petrosillo, & di Masi, 2016).

The PaLoc is a highly heterogenous region, and strains of *C. difficile* can be further characterised by specific changes in their PaLoc region, known as toxinotyping. There are a great number of possible genetic variations and toxin types across *tcdA* and *tcdB*, with the reference strain *C. difficile* VPI 10463 generating 27 variant toxinotypes (Rupnik et al., 2009). The majority of toxinotypes are variations of $TcdA^{+}TcdB^{+}$, although a small number have a deletion of *tcdA*. Non-toxinogenic strains ($TcdA^{-}$ $TcdB^{-}$) have a 115bp non-coding sequence in place of the PaLoc. As well as dictating the presence/absence of toxins, toxinotypic genetic variations can also alter toxin properties. Some toxinotypes have been associated with CDI outbreaks, for example toxinotype III (ribotype 027) (Genth et al., 2008; Rupnik et al., 2009).

Some *C. difficile* strains also produce a third toxin, '*Clostridioides difficile* toxin' (CDT), either independently or in addition to toxin A and toxin B, and this is encoded by *cdtA* and *cdtB* located on the CdtLoc (Rodriguez, Van Broeck, Taminiau, Delmée, & Daube, 2016). CDT is a binary, ADP-ribosylating toxin consisting of two separate

components, and in cultured cell lines causes the de-polymerisation of actin fibres through the ADP-ribosylation of G-actin. CDT is suggested to weaken the structure of the actin cytoskeleton, and increase adherence of the bacterium, but it functions independently to the glycosylating activities of toxin A and toxin B (Beer et al., 2018; Rodriguez, Van Broeck, et al., 2016).

Diagnosis

Diagnosis of CDI can be difficult as there is no single test for the disease. The presence of *C. difficile* and toxins is not necessarily indicative of CDI, as asymptomatic CDI can occur in the presence of toxins (Furuya-Kanamori et al., 2015). Therefore, CDI diagnosis is usually defined by the presence of clinical symptoms in conjunction with a laboratory test. Symptomatic diarrhoea is characterised by three or more unformed stools within a 24 h period, and CDI is confirmed by: a positive assay for toxin A and/or B; or identification of toxinogenic *C. difficile* in the stool sample; or pseudomembranous colitis observed in an endoscopy (Schäffler & Breitrück, 2018). To detect toxigenicity enzyme immunoassays are common, although they have lower sensitivity and are unable to detect A^-B^+ strains. Because of this, the gold standard is a cytotoxicity neutralisation assay which can detect both the presence of toxins, mainly toxin B, and the ability to produce toxins. At present there are no routine kits available to detect strains that produce only CDT (Elliott et al., 2017; Rupnik et al., 2009).

Treatment

Rapid diagnosis and treatment are important in slowing the progression of CDI. It is recommended that the first step following diagnosis be stopping the antibiotic responsible for infection if possible, particularly high-risk agents such as clindamycin and fluoroquinolones (Heinlen & Ballard, 2010). Standard treatment of CDI then involves antibiotic treatment with vancomycin and metronidazole. Previously metronidazole was the first line drug in CDI with vancomycin reserved for more serious cases, but it has since been demonstrated that vancomycin is superior to metronidazole in treating CDI (Czepiel et al., 2019; Nelson, Suda, & Evans, 2017). Fidaxomicin, a narrow spectrum antibiotic against Gram-positive pathogens, is a newer treatment available since 2011, and has shown comparable or better efficacy than vancomycin. It is highly efficacious against *C. difficile* and has been

demonstrated to have little impact on the native gut microbiome (Czepiel et al., 2019). Because of this it is now seen as a mainstay of CDI treatment along with vancomycin in the US, although hasn't been universally adopted (McDonald et al., 2018).

The two main issues with CDI treatment are the recurrence of the disease and treatment of severe cases. If CDI persists following initial treatment various antibiotic regimens can be trialled with different doses, frequencies and administration methods for varying severities, all the while trying to maintain a balance of normal gut flora. However, if the disease continues to progress and all other options have been exhausted, removal of the colon may sometimes be the only measure to prevent death (Czepiel et al., 2019) Moreover, even if treatment is successful, recurrence of the disease is an ongoing problem. In one study, after an initial treatment of fidaxomicin and vancomycin 15% and 25% of patients suffered a recurrence respectively (Louie et al., 2011), and in patients with previous recurrence the rates were 20% and 35% respectively after 4 weeks (Cornely, Miller, Louie, Crook, & Gorbach, 2012). In cases of multiple recurrence, faecal microbial transplant (FMT) has shown promise, for example it was able to treat 81% of multiple recurring CDI cases in a randomised controlled trial (Van Nood et al., 2013). However, currently it remains a non-standardised procedure and there is little evidence for the long-term effects of altering a patient's microbiota.

1.1.3 Sporulation of Clostridioides difficile

The Trouble with Spores

Spore formation is vital for *C. difficile* pathogenesis; as an obligate anaerobe vegetative cells are unable to survive in the aerobic environment, and under certain stimuli metabolically dormant spores are formed that can survive extreme conditions. Being aerotolerant, spores act as the infectious vehicle of CDI, contributing to transmission in both healthcare and community settings (Abt, McKenney, & Pamer, 2016; Bhattacharjee, McAllister, & Sorg, 2016). The ubiquitous nature of spores in the environment, combined with their long-term persistence, further increases the difficulties of treating CDI and preventing its recurrence and transfer. Spores have been identified in an array of environmental locations including raw meats and vegetables, soil and water samples, in addition to livestock and household animals

(Rodriguez-Palacios, Borgmann, Kline, & LeJeune, 2013; Rodriguez, Taminiau, Van Broeck, Delmée, & Daube, 2016). Moreover, their ability to persist on surfaces for months and withstand heat, antimicrobials and common disinfectants contributes to high rates of dissemination in hospitals (Barra-Carrasco & Paredes-Sabja, 2014). Persistence of spores in the GI tract even following treatment further contributes to CDI recurrence, allowing germination to the vegetative form (Zhu, Sorg, & Sun, 2018).

Formation and Structure of the Endospore

Sporulation is common to approximately 200 species within the Firmicutes phylum, covering 25 genera. The mechanism of sporulation has been extensively studied in the model species *Bacillus subtilis*, and whilst understanding of the mechanism in *C. difficile* still requires development, many of the essential components and morphological sequences are conserved between the two organisms (Henriques & Moran, 2007).

The structure of the mature spore is largely conserved among all spore-forming bacteria (Figure 1.1b). The structure consists of 3 main layers: the central core, the cortex layer, and an outer coat. The central core contains the supercoiled bacterial genome, has a low water content to maintain dormancy, and is rich is calcium-dipicolinic acid (Ca-DPA) to provide heat resistance properties (Kochan, Foley, et al., 2018). The core is surrounded by the inner membrane and the germ cell wall, with both functioning in spore germination. The inner membrane contains many vital elements for germination such as germination receptors and the germ cell wall, a lipid bilayer covered with peptidoglycan, goes on to serve as the vegetative cell wall. Surrounding the germ cell wall, the cortex is a thick layer of modified peptidoglycan that allows resistance to high temperatures. Lastly, the outer multiprotein coat protects against enzymatic breakdown of peptidoglycan (Talukdar, Olguín-Araneda, Alnoman, Paredes-Sabja, & Sarker, 2015; Zhu et al., 2018). In some *C. difficile* strains, there is a further layer surrounding the coat known as the exosporium, and this facilitates interactions between spore and the environment (Ball et al., 2008; Talukdar et al., 2015).

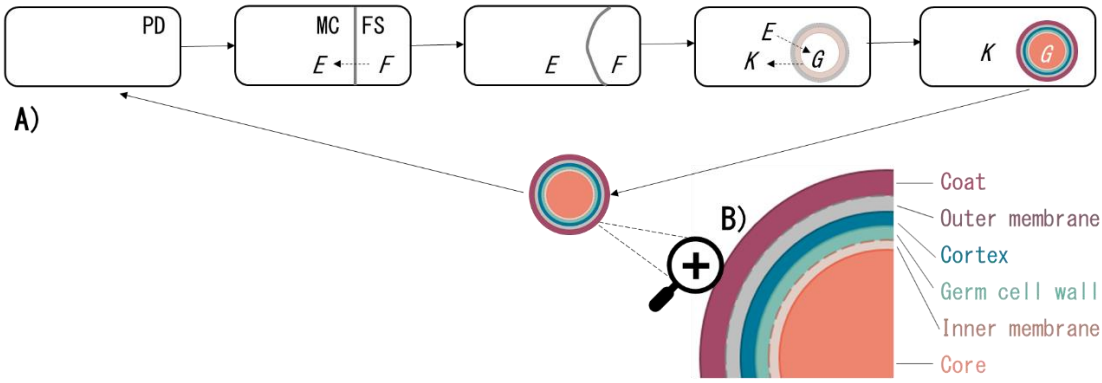


Figure 1.1 A) Morphological stages of sporulation in *B. subtilis*. Sporulation from the pre-divisional cell (PD) through to the release of a mature spore to the environment, with location and activity of sigma factors shown – SigF (F), SigE (E), SigG (G) and SigK (K). Asymmetric division of the predivisional cell occurs to form the mother cell (MC) and forespore (FS). The mother cell envelopes and engulfs the forespore and it becomes a free protoplast in the mother cell cytoplasm, consisting of the inner and outer membranes. The subsequent spore layers are then generated to produce a mature spore which is released during sporulation. **B) Endospore structure.** Including the coat layer present in some spores of *C. difficile*.

The conserved process of sporulation, outlined in Figure 1.1a, can be broken down into seven stages and involves two main structures, the forespore and the mother cell. These respectively represent the smaller and larger compartments of the sporangium, with the germline forespore eventually becoming the mature spore, and the mother cell providing temporary nourishment during the process until it self terminates (Henriques & Moran, 2007). Following growth of the vegetative pre-divisional cell during stage 0, asymmetric cell division occurs to form the prespore and mother cell during stage I and II. By the end of this process all the original DNA has been pumped into the prespore via SpoIIIE, a DNA translocase. During stage III, the mother cell membrane envelopes and engulfs the prespore to create the forespore, a free protoplast in the cytoplasm of the mother cell. Between the inner and outer membranes of the forespore, the peptidoglycan germ cell wall and cortex are then subsequently synthesised during stage IV, and then the external spore coat in stage V and modifications to withstand UV and heat in stage VI. In the final stage VII, the mature spore is released into the environment during sporulation as a result of mother cell

lysis, and exists in a dormant state (Saujet, Pereira, Henriques, & Martin-Verstraete, 2014; Talukdar et al., 2015).

Initiation and Regulation of Sporulation

The physiological changes of sporulation are achieved through separate and compartmentalised transcriptional changes within the forespore and mother cell, best characterised in *B. subtilis* (Edwards & Mcbride, 2014). The shift from vegetative growth to sporulation is controlled by Spo0A, a master regulatory protein which upregulates expression of sporulation-specific genes (Edwards & Mcbride, 2014; Pereira et al., 2013). Activation of Spo0A via a phosphorelay pathway leads to transcriptional control of gene expression in the forespore and mother cell through four sigma factors specific for sporulation (Edwards & Mcbride, 2014; Fimlaid & Shen, 2015). The sigma factors are sequentially activated in an alternating manner. Firstly, σ^F in the forespore and σ^E in mother cell are activated by Spo0A and control early development. These then activate and are replaced by σ^G in the forespore and σ^K in mother cell, once the forespore has been engulfed. This sigma factor cascade, in combination with other regulatory sporulation proteins, allows concerted waves of forespore and mother cell gene expression, allowing tight coordination of each morphological stage of spore development (Pereira et al., 2013; Talukdar et al., 2015).

Spo0A is highly conserved; its operons in *B. subtilis* and *C. difficile* share significant overlap, and it has some functional similarities in *C. difficile* (Edwards & Mcbride, 2014). In *C. difficile*, it has also been demonstrated as a master regulator by functional genomics, allowing various adaptations to host interactions through global transcriptional regulation of sporulation, virulence and metabolic phenotypes (Pettit et al., 2014). Moreover, *spo0A* mutants are able to cause CDI symptoms but are unable to persist within mice and allow transmission of the disease, demonstrating key roles of the protein (Deakin et al., 2012). Inactivation of *spo0A* also prevents spore formation, as in *B. subtilis* (Underwood et al., 2009). However, unlike *B. subtilis*, the full sporulation pathway in *C. difficile* has not been elucidated and the regulatory mechanism is mostly unknown. Spo0A has been shown to be directly phosphorylated and is hypothesised to be activated by a two component system instead of a multicomponent phosphorelay (Underwood et al., 2009)

1.1.4 Models of Spore Germination

Germination in the B. subtilis *model*

Germination is the process by which a dormant spore returns to active vegetative growth (Francis, Allen, & Sorg, 2015). As with sporulation, most knowledge of the mechanisms of germination are based on the model organism *B. subtilis*, and the process shares features amongst many spore forming organisms (Figure 1.2). Commonly, germination is triggered through the sensing of specific small molecules (germinants) that demonstrate the return to favourable environmental conditions, and these germinants are detected at their respective germination receptors (Setlow, Wang, & Li, 2017). *B. subtilis* germinants include either L-alanine or a mixture of asparagine, glucose, fructose and potassium ions, and they respectively bind to the receptors GerAA-AB-AC and GerB/GerK located in the spore inner membrane (Burns, Heap, & Minton, 2010; Setlow, 2014). Germinant binding activates the release of monovalent cations (H^+, Na^+, K^+) and Ca-DPA from the core, likely through the SpoVA channel. At this stage the spore is committed to germinate; germination occurs even if physiological germinants are removed (Francis et al., 2015; Setlow et al., 2017). The release of Ca-DPA leads to the activation of spore cortex lytic enzymes (SCLE) SleB and CwlJ, which hydrolyse the peptidoglycan of the cortex. This allows full rehydration of the core to that of a vegetative cell, and full metabolism and function is recovered, allowing outgrowth (Burns et al., 2010; Zhu et al., 2018).

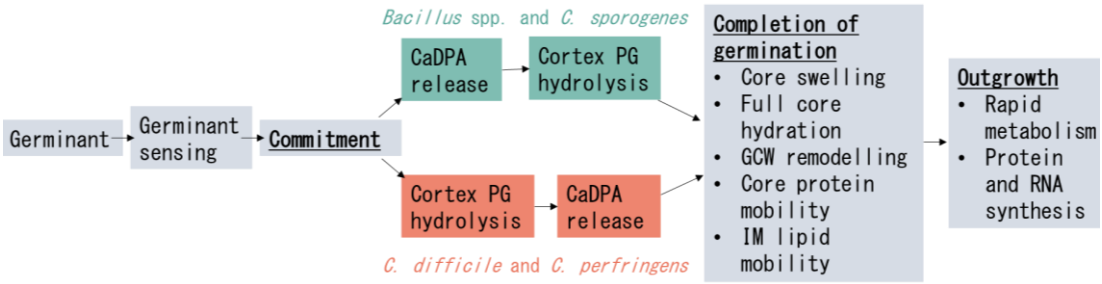


Figure 1.2 Spore germination. Outline of the stages of spore germination in spore forming bacteria, including some species differences. Figure adapted from Setlow et al (2017).

Germination of C. difficile

Germination is a vital step in the pathway from spore ingestion to CDI development, with 511 genes significantly up or down regulated (Dembek, Stabler, Witney, Wren, & Fairweather, 2013). Once *C. difficile* spores have endured the acidic environment of the stomach, germination allows the transition to a metabolically active, vegetative state in which outgrowth, colonisation and toxin production can occur. Spores germinate in the small intestine and colonise the large intestine, the site of CDI (Kochan, Shoshiev, et al., 2018).

The pathway to germination in *C. difficile* has several differences to that in *B. subtilis*. In contrast to *B. subtilis*, it has become evident that *C. difficile* germinates in response to bile acids, specifically the primary bile acid cholate and its derivatives, with glycine acting as a cogerminant (Howerton, Ramirez, & Abel-Santos, 2011; Sorg & Sonenshein, 2008). Opposingly, chenodeoxycholate, another primary bile acid, has been demonstrated to inhibit germination (K. H. Wilson, 1983). The complex and pivotal role of bile acids will be the focus of the remainder of this review. Moreover, concordant with differing germinants, *C. difficile* does not contain an ortholog for the germinant receptor GerA utilised by *B. subtilis* (Kochan, Foley, et al., 2018). Instead, Francis et al. (2013) have identified the pseudoprotease CspC as the bile acid germinant receptor, and suggest *C. difficile* is likely to utilise a unique mechanism of germinant recognition. Further differences have also been identified downstream of germinant recognition. Similar to *Clostridium perfringens*, the Csp proteins and SCLE (spore cortex lytic enzymes) are involved in cortex degradation (Paredes-Sabja, Shen, & Sorg, 2014). Whilst *C. difficile* encodes a homolog to SleB present in *B. subtilis*, cortex hydrolysis is in fact carried out by the SCLE SleC (Gutelius, Hokeness, Logan, & Reid, 2014), and this is thought to be activated by CspB which is cleaved from an original CspBA fusion protein (Adams, Eckenroth, Putnam, Doublié, & Shen, 2013). Finally, as has also been suggested in *C. perfringens*, it appears that cortex hydrolysis precedes Ca-DPA release as opposed to the opposite in *B. subtilis* (Setlow et al., 2017). Whilst advances have been made in characterising certain stages of the *C. difficile* germination pathway as summarised here, the complete mechanism remains to be elucidated.

1.1.5 Resisting Clostridioides difficile *colonisation*

Following transmission of *C. difficile* spores there are many factors that can challenge successful infection in humans, and an adequate immune response is usually sufficient to eliminate the infection, or lead to asymptomatic carriage (Smits et al., 2016). Both the innate and adaptive arms of the immune system are activated. The initial innate response is to prevent bacterial or toxin binding through the production of the intestinal epithelial mucus barrier, host antimicrobial peptides, and chemical attenuation of toxins (Solomon, 2013). Following this, toxins stimulate the release of proinflammatory molecules such as cytokines as well as the recruitment and activation of immune cells (Madan & Petri, 2012; Solomon, 2013). Adaptive immunity against CDI has mainly been studied with regards to the antibody-mediated response, and up to 60% of healthy adults present detectable antibody levels from an original disease challenge (Sánchez-Hurtado et al., 2008). The immunoglobulins IgA, IgM and IgG are important in mediating immunity to *C. difficile*, and high levels have been linked to improved infection and survival rates (Madan & Petri, 2012; Solomon, 2013).

A further defence of the host innate immune response is derived from the microbial flora. Protection by the microbiome is demonstrated by loss of colonisation resistance to CDI as a result of long term antibiotic use and long-lasting community changes in the gut microbiome (Antonopoulos et al., 2009); antibiotic usage is the main risk factor for CDI (Czepiel et al., 2019). For example, in a mouse model by Buffie et al. (2012), mice administered *C. difficile* spores with a native flora did not develop CDI symptoms, whereas even a single dose of clindamycin reduced the normal microbial taxa by 90% with a 40-50% mortality rate. Other mouse models using different combinations of antibiotic treatments, and therefore inducing different microbial perturbations, have further suggested that different microbial communities confer varying levels of colonisation resistance and protection against CDI (Buffie et al., 2012; Koenigsknecht et al., 2015; Reeves et al., 2011; Schubert, Sinani, & Schloss, 2015). Whilst CDI susceptibility has not been attributed to one microbial community alone, associations have been made with a decrease in overall diversity in the gut, an increase in Proteobacteria, and a decrease in Bacteriodetes (Buffie et al., 2012; Theriot et al., 2014; Winston & Theriot, 2016).

Studying the gut microbiota of CDI patients has also demonstrated its probable role in colonisation resistance against *C. difficile* and shares similarities to observations in mouse models. Commonly, there is a reduced microbial richness and diversity in CDI patients compared to healthy controls (Antharam et al., 2013; Gu et al., 2015; L. Zhang et al., 2015). A reduction has been seen in the Firmicutes (Antharam et al., 2013; L. Zhang et al., 2015), specifically the usually abundant *Ruminococcaceae* and *Lachnospiraceae* (Antharam et al., 2013), as well as a reduction in Bacteroidetes and an increase in Proteobacteria (L. Zhang et al., 2015). Moreover, studying faecal samples prior to, not during, CDI development have also demonstrated similar changes, with a reduction in microbial diversity and Bacteroidetes, as well as a reduction in Clostridiales Incertae Sedis XI (Vincent et al., 2013).

It is evident that the microbiome plays a complex role in protection from CDI, with suggestions that colonisation resistance against *C. difficile* occurs from multiple bacterial communities and not one single population or community (Schubert et al., 2015; Theriot & Young, 2015), but also that CDI disease progression is likely to be attributed to specific taxa (Crobach et al., 2018; Leslie, Vendrov, Jenior, & Young, 2019). Identifying and studying the specific taxa involved in CDI protection, and understanding the bacterial mechanisms that allow for this protection, will allow developments in CDI prevention and treatment. In addition to more generalised colonisation resistance mechanisms, microbially derived bile acids and the species responsible for their synthesis have been shown to play a defensive role against *C. difficile* (Theriot, Bowman, & Young, 2016; Winston & Theriot, 2016). This arises due to the role of bile acids in *C. difficile* germination and growth and will be the focus of this thesis.

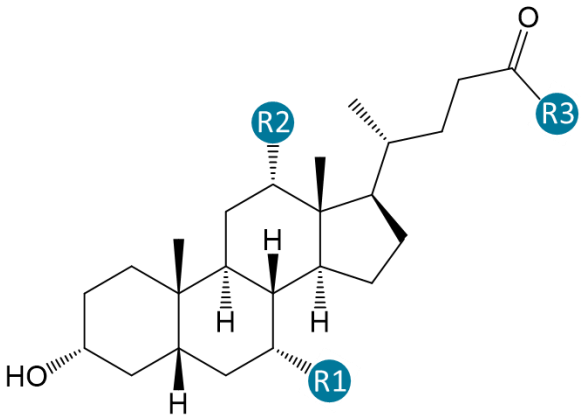
1.2 Bile Acids and *Clostridioides difficile* Infection

1.2.1 Bile Acids

The Structure and Function of Bile Acids

Produced in the hepatocytes of the liver, bile acids are derived from cholesterol by the action of 17 enzymes, and are amphipathic molecules based on a hydroxyl containing sterol nucleus with a carboxyl side chain and various hydrophilic substituents, shown

in Figure 1.3 (Houten, Watanabe, & Auwerx, 2006; Urdaneta & Casadesús, 2017). In humans bile acids are produced at a rate of 350 mg per day and can exist in four states: primary or secondary; and conjugated or unconjugated to taurine or glycine (Urdaneta & Casadesús, 2017). The decrease in pK_a arising due to conjugation results in ionisation of bile acids at physiological pH, and association with Na^+ and K^+, when they are then termed bile salts (Ridlon, Kang, & Hylemon, 2006). Primary bile acids are the direct products of bile acid synthesis from cholesterol, with the major entities in humans being cholic acid (CA) and chenodeoxycholic acid (CDCA). The major secondary bile acids in humans are deoxycholic acid (DCA) and lithocholic acid (LCA), derived from cholic acid and chenodeoxycholic acid respectively, and are produced by intestinal bacterial transformation (Staels & Fonseca, 2009). The human bile acid pool consists mainly of cholic acid, chenodeoxycholic acid and deoxycholic acid in an approximate ratio of 4:4:2 (Pandak & Kakiyama, 2019).



Bile Acid	Abbr.	Primary/ Secondary (P/S)	R1 (position 7)	R2 (position 12)	R3 (position 24)
Glycocholate	GCA	P	OH (α)	OH (α)	$NHCH_2COO^-$
Taurocholate	TCA	P	OH (α)	OH (α)	$NHCH_2CH_2SO_3^-$
Cholate	CA	P	OH (α)	OH (α)	OH
Chenodeoxycholate	CDCA	P	OH (α)	H	OH
Glycodeoxycholate	GDCA	S	H	OH (α)	$NHCH_2COO^-$
Taurodeoxycholate	TDCA	S	H	OH (α)	$NHCH_2CH_2SO_3^-$
Deoxycholate	DCA	S	H	OH (α)	OH
Lithocholate	LCA	S	H	H	OH

Figure 1.3 Structures of common primary and secondary bile acids. Bile acid steroidal nucleus with functional groups (R1, R2, R3) listed.

The main role of bile acids in humans is to enable the digestion and absorption of lipids in the small intestine. Bile acids are secreted as a constituent of bile into the duodenum on ingestion of a meal (Houten et al., 2006) and their amphipathic nature and planar conformation allow them to act as physiological surfactants for solubilisation of dietary fats (Enright, Griffin, Gahan, & Joyce, 2018). Bile acids also play a major role in cholesterol homeostasis, representing the primary pathway for cholesterol catabolism, accounting for 50% of its daily turnover (Staels & Fonseca, 2009). In addition to their role as detergents, they have also been shown to act as signalling molecules, with systemic endocrine functions. Downstream signalling has been implicated in a diverse range of metabolic functions including inflammation and glucose metabolism (Houten et al., 2006; Ticho, Malhotra, Dudeja, Gill, & Alrefai, 2019).

Bile Salt Metabolism and Distribution

Following secretion into the duodenum, bile acids are reabsorbed from the distal ileum and 95% undergo the process of enterohepatic recirculation, maintaining the bile acid pool. The portal circulation delivers them back to the liver where they are removed by hepatic active transport, and re-secreted into the bile (Ridlon et al., 2006; Staels & Fonseca, 2009). The remaining 5% not reabsorbed is largely explained by actions of microbial enzymes, and these enzymes are also pivotal in the further transformations that occur when this 5% escapes recirculation and enters the large intestine (Enright et al., 2018).

The prevention of bile acid reabsorption can mainly be attributed to microbial bile salt hydrolases (BSHs), and these are present in all major bacterial divisions of the gut, albeit with differing levels of activity even at the species level (Jones, Begley, Hill, Gahan, & Marchesi, 2008). BSHs hydrolyse the amide bond between the steroid nucleus of the bile acid and glycine or taurine, resulting in deconjugation and production of a free bile acid. Deconjugated bile acids are no longer reabsorbed as they cannot undergo active transport (Enright et al., 2018; Urdaneta & Casadesús, 2017). The deconjugation of primary bile acids is of utmost importance as it allows their subsequent conversion into secondary bile acids. Passing into the large intestine, they undergo transformations by a limited number of microbial species, mainly clostridia, that possess 7α-dehydroxylating activity. This transformation is solely

responsible for the production of DCA and LCA and cannot occur with conjugated bile acids due to steric hindrance (Enright et al., 2018; Ridlon et al., 2006; Urdaneta & Casadesús, 2017). The process of converting primary to secondary bile acids is summarised in Figure 1.4.

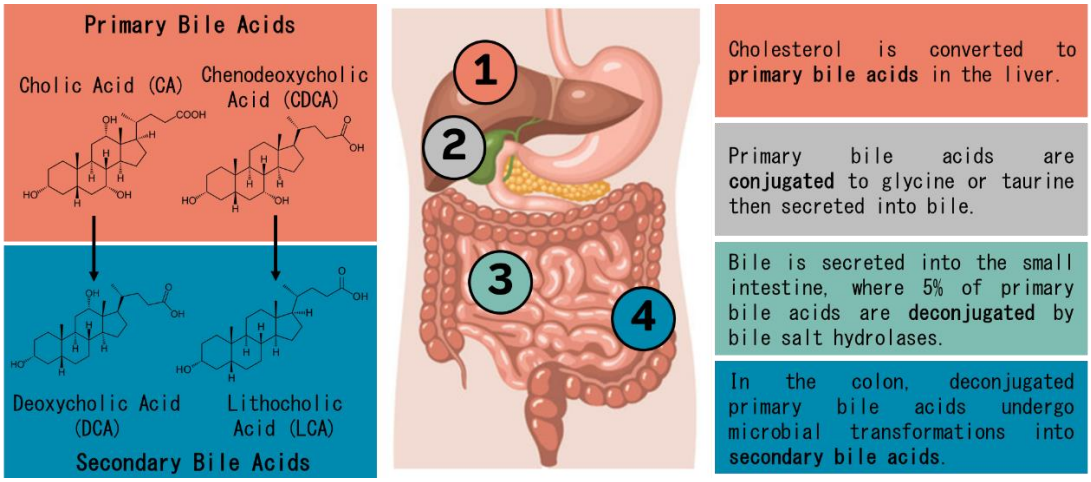


Figure 1.4 Summary of bile acid transformations in the human gut. The process of converting primary bile acids to secondary bile acids.

The importance of all microbial transformations is demonstrated through the changing diversity and concentration of gut microbiota throughout the digestive tract, and its effect on the composition of the bile acid pool, shown in Figure 1.5; there is an inversion of the predominance of conjugated/deconjugated and primary/secondary bile acids from the small intestine to large intestine (Enright et al., 2018). Germ free rats have been shown to have an increase in taurine-conjugated bile acids and a lower bile acid diversity (Swann et al., 2011). The subsequent array of downstream signalling networks involving bile acids, and their dependence on bile acid composition, thus highlights the importance of microbial transformations.

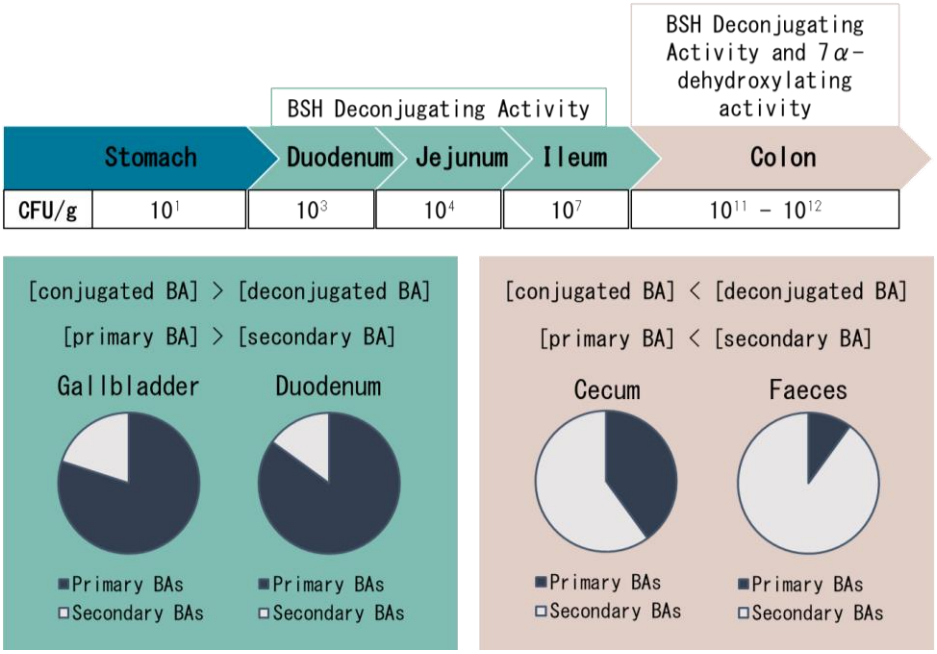


Figure 1.5 Microbial populations and their bile acid transformations along the digestive tract. From the small intestine to colon there is an inversion of the predominance of conjugated/deconjugated and primary/secondary bile acids. Adapted from Enright et al (2016).

1.2.2 The Role of Bile Acids in Clostridioides difficile *Germination*

Stimulation of C. difficile *Germination by Bile Acids*

It was in the early 1980's that the primary bile acid taurocholate was first shown to be important in the germination of *C. difficile*. Initial reports by Raibaud *et al.* (1974) suggested 1% sodium taurocholate augmented outgrowth of faecal clostridial spores, and Wilson *et al.* (1982) demonstrated a 50-fold higher recovery of *C. difficile* spores with 0.1% taurocholate than without. A full characterisation by Sorg and Sonenshein (2008) revealed more detail. Taurocholate at 0.1% was shown to induce a rapid rate of colony formation in CD196, and plating efficiency increased with taurocholate concentration. However, measuring germination using optical density (OD) identified that taurocholate alone could not activate germination, and glycine was identified as a co-germinant using a defined medium, confirmed in CD196 where together they enhanced germination efficiency. Only cholate derivatives taurocholate, glycocholate,

and cholate allowed germination and colony formation when primary bile acids were tested, in addition to the secondary bile acid deoxycholate (Sorg & Sonenshein, 2008). It has since been determined that *C. difficile* spores have an apparent Km for taurocholate of 2.8 mM, a physiologically relevant concentration (Sorg & Sonenshein, 2010). Understanding of the mechanism was further developed by a structure-activity study by Howerton *et al* (2011). Taurocholate and glycine analogues revealed that the germination machinery recognises amino acids and taurocholate through multiple interactions. For example, the 12-hydroxyl group of taurocholate is necessary but not sufficient to activate germination, and only shorter alkyl amino sulfonic acid side chains are recognised by spores (Howerton et al., 2011).

In addition to this *in vitro* work using synthetic bile acids, work with mouse intestinal and cecal extracts has also implicated primary bile acids as a likely germination factor. Extracts that could stimulate colony formation had a higher ratio of primary to secondary bile acids than extracts that could not, and a bile salt sequestrant ameliorated this stimulation (Giel, Sorg, Sonenshein, & Zhu, 2010). Moreover, taurocholate incubated with clindamycin-treated cecal contents, and therefore less likely to have undergone transformations into secondary bile acids, had a higher CFU recovery (Giel et al., 2010).

Whilst taurocholate and glycine have been demonstrated to play an important role in activating germination, diversity in this response has been demonstrated in different clinical isolates. A study of 15 *C. difficile* isolates showed a varied extent and rate of germination response to 0.1% taurocholate, and this was also seen within grouped BI/NAP/027 isolates (Heeg, Burns, Cartman, & Minton, 2012). Moore *et al.* (2013) have also demonstrated different germination efficiencies between clinical isolates, and implicated strains with higher efficiencies in severe CDI. Both studies of clinical isolates also demonstrated germination in the absence of taurocholate, but this also differed between strains (Heeg et al., 2012; Moore et al., 2013). These studies suggest that the response to germinants is complex and may be linked to disease severity and treatment success in different strains of *C. difficile*.

Considering that the suggested germinants for *C. difficile* differ to that of *B. subtilis*, it is not surprising that homology searches for a *ger* receptor in *C. difficile* have been unsuccessful (Sebaihia et al., 2006). An alternative germination receptor has been

identified, the non-catalytically active pseudoprotease CspC (Francis et al., 2013), which on bile acid binding activates the CspB serine protease through an unknown mechanism, resulting in the activation of the SleC cortex hydrolase and degradation of the cortex (Fimlaid et al., 2015). Following a screen to identify *C. difficile* germination-null phenotypes using chemical mutagenesis, 8 out of 10 of the corresponding SNPs were located in *cspC*. Site directed mutagenesis of *cspC* using TargeTron technology generated mutants unable to initiate germination in response to taurocholic acid, and a further single point mutation G457R generated an altered germinant specificity. These results demonstrate that CspC plays both a role in germination and that it functions as a receptor for bile acid germinants (Francis et al., 2013).

Inhibition of C. difficile *Germination by Bile Acids*

The other primary bile acid chenodeoxycholate has been shown to be involved in the inhibition of *C. difficile* spore germination. Chenodeoxycholate has been shown to block taurocholate and cholic acid induced colony formation in a concentration dependent manner, as well as inhibiting taurocholate induced germination, suggesting competitive inhibition (Sorg & Sonenshein, 2009). However, similarly to taurocholate, it has been shown that chenodeoxycholate does not inhibit the germination of every *C. difficile* isolate (Heeg et al., 2012). There are currently no structure-function analyses available to identify which structural entities are important for this to occur.

Secondary bile acids have also been shown to act as germination inhibitors, with lithocholic acid and the minor secondary bile acid ursodeoxycholic acid exhibiting inhibitory effects of taurocholate induced germination (Sorg & Sonenshein, 2010). Thanissery *et al* (2017) also tested seven gut derived secondary bile acids at physiological concentrations for the inhibition of taurocholate induced germination in a range of *C. difficile* strains. Whilst strain specific, their results confirmed inhibition by LCA and UDCA and highlighted inhibition by deoxycholic acid and ω-muricholic acid, a murine bile acid (Rajani Thanissery et al., 2017). Reduction of vegetative growth by secondary bile acids has also been shown, for DCA (Sorg & Sonenshein, 2008; Rajani Thanissery et al., 2017; K. H. Wilson, 1983), LCA, ω-MCA and UDCA (Rajani Thanissery et al., 2017), and chenodeoxycholate (Sorg & Sonenshein, 2008).

This provides partial insight into the inhibitory effects of individual secondary bile acids on *C. difficile*. In addition to the possibility of its clinical use, the synthetic bile salt analogue CamSA (cholic amide *m*-sulfonic acid), may help further mechanistic understanding of the role of bile acids in germination as it has been identified as a strong competitive inhibitor of taurocholate induced *C. difficile* germination (Howerton et al., 2011). It was able to prevent CDI infection in a concentration dependent manner in spore but not vegetative challenged mice with two different strains of *C. difficile* (Howerton, Patra, & Abel-Santos, 2013).

The importance of the inhibitory role played by secondary bile acids is demonstrated through the decrease in CDI resistance when microbes responsible for the conversion of primary bile acids to secondary bile acids are removed. By using antibiotics to create distinct murine microbial and bile acid environments, *C. difficile* spores were shown to germinate and grow in most ileal and cecal content that had significant alterations to the microbiome and were depleted of secondary bile acids, and the respective antibiotics used have been associated with risk of CDI in humans, for example cephalosporin and clindamycin. Specific bacterial members were also positively correlated with secondary bile acids in the cecum and CDI resistance, specifically the Lachnospiraceae and Ruminococcaceae families (Theriot et al., 2016). Similarly, using antibiotic exposed mice Buffie *et al.* (2015) demonstrated that a high secondary bile acid relative abundance was highly correlated with CDI resistance, as was the abundance of the gene family responsible for secondary bile acid synthesis.

7α-dehydroxylation and the bai *operon*

The multistep process of 7α-dehydroxylation, carried out by microbes, converts the primary bile acids CA and CDCA into the secondary bile acids DCA and LCA respectively, through the removal of the 7-hydroxy group, and occurs in the large intestine (Ridlon, Kang, & Hylemon, 2010). The extent of this microbial metabolism is demonstrated through comparison of bile acid composition in the gallbladder and faeces, with CA and CDCA decreasing from 35% to 2% each, and DCA and LCA increasing from 25% and 1% to 34% and 29% respectively (Ridlon et al., 2006). The microbiome is pivotal in these transformations; germ free mice have been shown to lack secondary bile acids (Narushima et al., 2006). However, only a limited number of species have been shown to carry out 7α-dehydroxylation, with the pathway

estimated to be present in only 0.0001% of total colonic flora (Ridlon et al., 2006). These species are largely members of the genus *Clostridium*, with the main species characterised being *Clostridium scindens* and *Clostridium hylemonae* in cluster XIVa, and *Peptacetobacter hiranonis* in cluster XI (Ridlon & Hylemon, 2012). The pathway involves a net reduction of two electrons, and the ability of these bacteria to use bile acids as electron acceptors is hypothesised to be a beneficial niche (Ridlon et al., 2006).

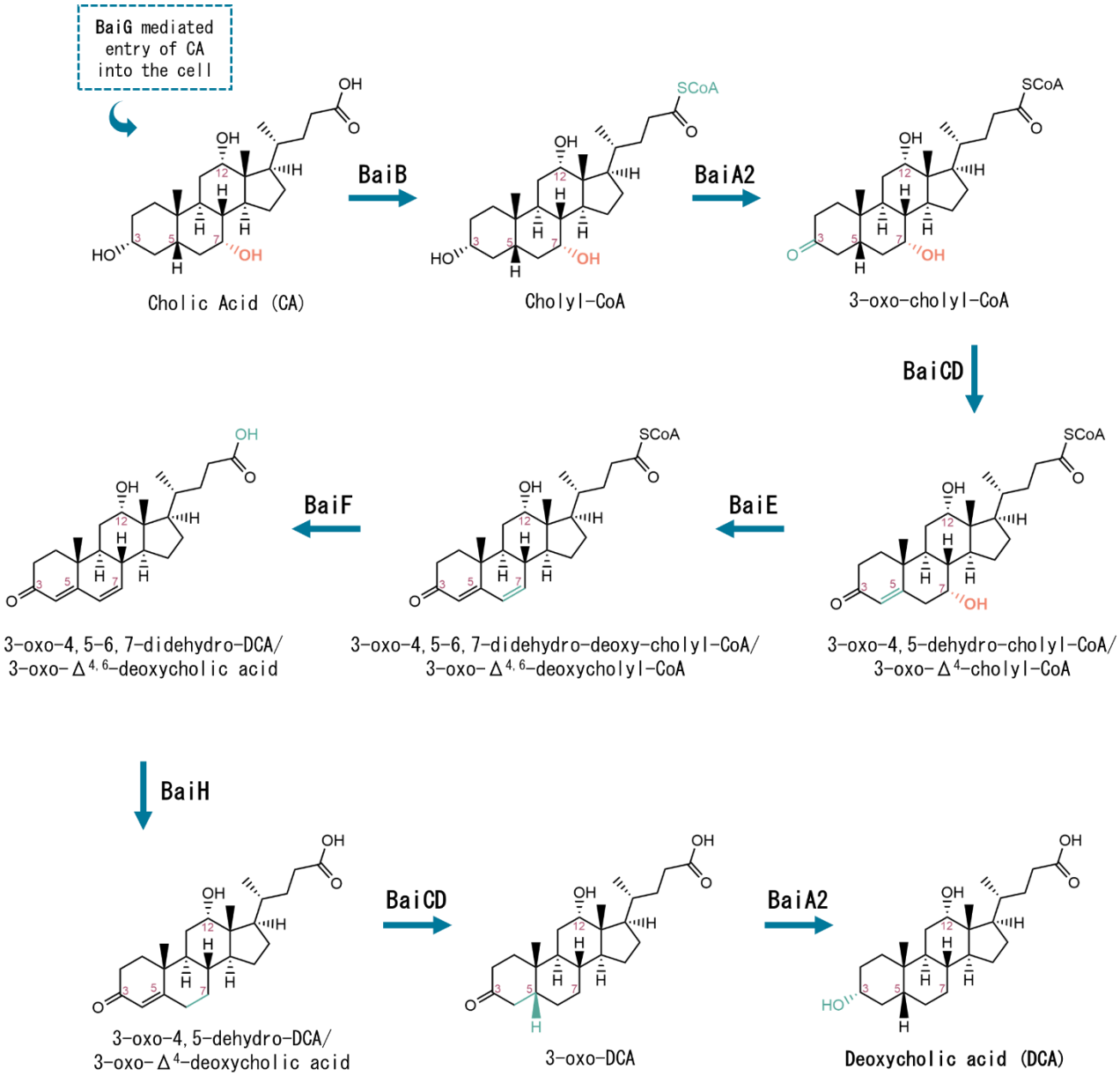


Figure 1.6 The 7α-dehydroxylation pathway. The proposed pathway for the conversion of cholic acid (CA) to deoxycholic acid (DCA) via 7α-dehydroxylation. The steps are the same for the conversion of chenodeoxycholic acid (CDCA) to lithocholic acid (LCA) as they only differ in structure at the C12 group. Based on the pathways proposed by Funabashi et el., (2020) and Lee et al., (2022). Opposing structure names are both shown (A/B). The sequential reduction of 3-oxo-Δ^4-DCA/LCA to allo-DCA/LCA, the 'flat' 3D structures, is not shown as the enzymes involved are currently unknown.

The biochemical pathway for 7α-dehydroxylation involves enzymes encoded by bile acid inducible (*bai*) genes and has mainly been characterised from work in *C. scindens* (Bhowmik et al., 2016). The current model of the 7α-dehydroxylation pathway is shown in Figure 1.6, is most well characterised for the conversion of CA/CDCA to DCA/LCA, respectively, and can be split into an oxidative arm and reductive arm. The first step is the transport of free primary bile acids into the bacterial cell by an H^+-dependent active transporter, encoded by *baiG* (Darrell H Mallonee & Hylemon, 1996b). Once inside, the bile acid undergoes ATP-dependent thioesterification to coenzyme A (CoA) by a CoA-ligase, encoded by *baiB* (Darrell H Mallonee, Adams, & Hylemon, 1992), and then enters the oxidative phase of the pathway. The bile acid-CoA thioester is converted to the intermediates 3-oxo-7α-hydroxy bile acid-CoA and 3-oxo-Δ4−7α-hydroxy bile acid-CoA, by the actions of the *baiA* dehydrogenase (Bhowmik et al., 2014; D H Mallonee, Lijewski, & Hylemon, 1995) and *baiCD* oxidoreductase (D. J. Kang, Ridlon, Moore, Barnes, & Hylemon, 2008) respectively. The latter product then undergoes 7α-dehydration by *baiE* which is the irreversible and rate limiting step, converting it to a 3-oxo-Δ4,6−7α-hydroxy bile acid-CoA intermediate (Dawson, Mallonee, Bjorkhem, & B, 1996; Ridlon & Hylemon, 2012), which then enters the reductive arm of the pathway. It is only recently that the enzymes of the reductive arm have been elucidated (Funabashi et al., 2020; Lee et al., 2022). The Co-A transferase BaiF is responsible for the removal of the CoA moiety, forming a highly-oxidised intermediate of DCA (Ridlon & Hylemon, 2012), and the second and third reductive steps were surprisingly shown to be catalysed by *baiH* and *baiCD*, revealed to be homologous enzymes in the Fe-S flavoenzyme superfamily. This generates a 3-oxo-Δ^4-intermediate then 3-oxo-DCA. The last step, reduction to DCA, is carried out by BaiA2 (Funabashi et al., 2020). The reductive arm for the pathway generating allo-DCA and allo-LCA is still unclear.

The *bai* operon encoding the *bai* genes was first discovered in *C. scindens*, with Northern Blot analysis identifying a large (>10kb) CA-inducible mRNA transcript, in addition to a smaller transcript of 1.5 kb. The *bai* regulon was then identified, encoding at least 10 open reading frames, and the function of individual genes has been identified through subcloning in *Escherichia coli* and characterisation of native or recombinant enzymes. Currently genetic knockouts of the major *bai* operon have not been reported (Ridlon, Harris, Bhowmik, Kang, & Hylemon, 2016; Ridlon et al.,

2006). Characterisation of the *bai* operon has been carried out in species such as *C. scindens* ATCC 35704, *C. scindens* VPI 12708, *C. hylemonae* TN271, *P. hiranonis* TO931, with some differences observed as shown in Figure 1.7 (Ridlon et al., 2016). 7α-dehydroxylating activity has also been identified in *Clostridium sordellii*, *Paraclostridium bifermentans* (formerly *Clostridium bifermentans*), and *Clostridium leptum* (Maki Kitahara, Takamine, Imamura, & Benno, 2000).

In addition to the genes involved in the main 7α-dehydroxylation pathway, there are further *bai* genes that have been identified and characterised, shown in Figure 1.7. The multi gene operon *baiJKL(M)* has been identified in *C. hylemonae* and *C. scindens* VPI 12708, encoding: BaiK, a bile acid CoA transferase; BaiJ, a predicted flavin-dependent oxidoreductase; BaiL, a putative hydroxysteroid dehydrogenase; and BaiM, and a putative transporter (Ridlon & Hylemon, 2012). Currently their role in bile-metabolism is not known, particularly given that the operon is not present in all major 7α-dehydroxylation strains. Other *bai* genes of interest include *baiN*, a flavoprotein predicted to catalyse similar sequential reactions to *baiH* and *baiCD* (Funabashi et al., 2020; Harris et al., 2018). BaiJ and BaiP have also been identified and proposed as 5α- reductases, suggested to catalyse the formation of allo-bile acids (Lee et al., 2022). Whilst the complete 7α-dehydroxylation pathway has now been proposed, it appears that there may be some redundancy among enzymes, and further work is required to elucidate the role of all Bai proteins.

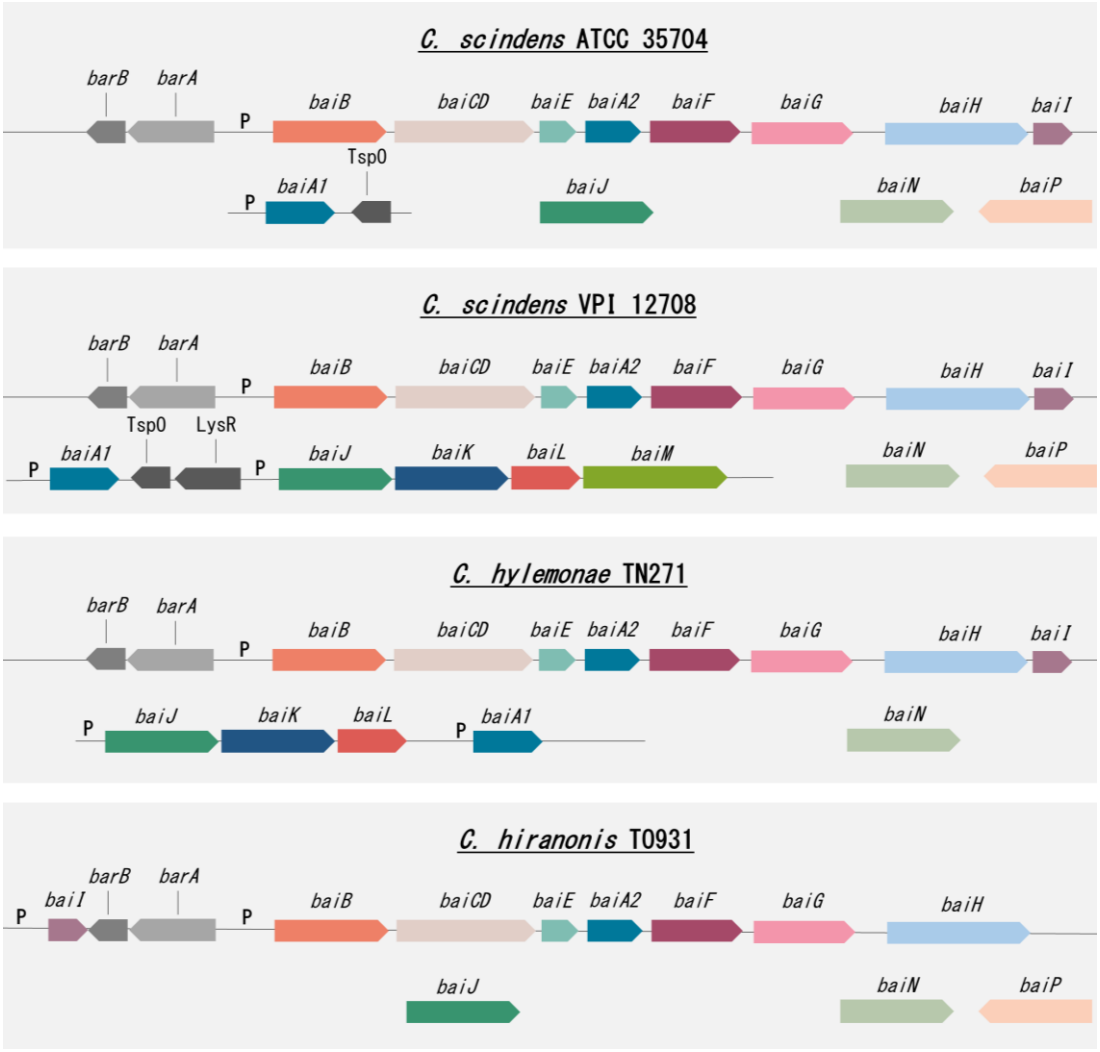


Figure 1.7 Comparison of *bai* operon structure in classical 7α-dehydroxylating organisms. Arrangement(s) of genes contributing to bile acid metabolism and conversion of primary to secondary bile acids by 7α-dehydroxylation. Image reproduced from (Lee et al., 2022).

Colonisation Resistance by C. scindens

It was not until the pivotal work of Buffie et al. (2015) that *C. scindens* was identified as a probable single species mediator of colonisation resistance against *C. difficile*, through its bile acid modulating activity. Their paper has expanded upon the work of previous studies that have demonstrated colonisation resistance of 7α-dehydroxylating bacteria in general, discussed in 1.2.2. This principle was combined with mouse models, clinical studies, metagenomic analyses and mathematical modelling,

culminating in the identification of *C. scindens* (Buffie et al., 2015). Initial work in mice demonstrated that clindamycin, ampicillin and enrofloxacin induced varying susceptibility to CDI and also distinct changes to the intestinal microbiota composition. Correlating resistance with individual bacterial species abundances identified 11 bacterial operational taxonomic units (OTUs) comprising mainly of *Clostridium* cluster XIVa, with *C. scindens* displaying the highest correlation. To investigate these correlations in humans, mathematical modelling was used to predict the microbiotia composition of antibiotic treated hematologic patients and again identified *C. scindens* as the strongest CDI inhibiting OTU. To evaluate causality, *C. scindens* alone and as part of a set of 4 inhibiting OTUs was transferred to antibiotic treated animals and both ameliorated CDI, with the abundance of *C. scindens* correlating with resistance. Analysis of the original antibiotic treated mice demonstrated that the recovery of secondary bile acids, and the abundance of *bai* genes, but not BSH genes, correlated with CDI resistance. Mice treated with *C. scindens* had restored levels of DCA and LCA, but exhibited no significant changes in primary bile acids. Finally, *ex vivo* work demonstrated that the inhibition of CDI in intestinal content by *C. scindens* was prevented when the bile acid sequestrant cholestyramine was used (Buffie et al., 2015). Their work strongly implicates *C. scindens* as a single species in conferring resistance to CDI by utilising bile acids to synthesise metabolites that inhibit *C. difficile*. This could be by inhibition of germination, growth, or a combination of the two.

Following this work many groups have sought to characterise the protective effects of *C. scindens* against CDI. Studer *et al.* (2016) used a novel gnotobiotic mouse model to test the impact of introduction of *C. scindens* into the mouse microbiome. A defined intestinal microbial consortium was used and confirmed to lack colonisation resistance against *C. difficile* and the ability to 7α-dehydroxylate. The mice lacked most secondary bile acids, but four out of six were fully or partially restored with colonisation of *C. scindens*. *C. difficile* infection was then carried out 4 days after *C. scindens* colonisation. Compared to controls, *C. scindens* colonisation decreased early *C. difficile* large intestinal colonisation and pathogenesis, with only minor changes in the native microbial consortium. Full colonisation was not prevented, but this may require a more complex native microbiota. This work further associated *C. scindens* and its production of LCA and DCA with CDI resistance *in vivo*, although full proof

of causality is prevented due to the inability to generate genetic mutants (Studer et al., 2016).

Further work in mice with the defined intestinal microbial consortium investigated the colonisation dynamics of *C. scindens* using labelled cells and demonstrated that the large intestine is a niche for *C. scindens* and is the location for 7α-dehydroxylation of CA to DCA *in vivo*. The authors also identified other transformations carried out by the bacterium, for example oxidation of other hydroxy groups, identifying a major transient product in 7α-dehydroxylation (Marion et al., 2019). Transcriptomic analyses by Devendran et al. (2019) have also defined the response to bile acids. The *bai* operon was highly expressed in response to CA but not DCA, consistent with other studies. Novel genes were also identified that were upregulated by CA, but their function, and possible function in cholic acid metabolism, is currently undefined. The authors also suggest a link between nutritional requirements and 7α-dehydroxylation, for example tryptophan (Devendran et al., 2019). Interestingly Kang *et al.* (2019) suggest that production of tryptophan-derived antibiotics by *C. scindens* offers a partial explanation for its protection. The newly discovered 1-acetyl-β-carboline and turbomycin A inhibited growth of *C. difficile*, as well as other gut bacteria such as *Staphylococcus aureus* and *E. coli*, with inhibition enhanced by DCA and LCA, but not CA. The suggested mechanism of action for the inhibition is inhibition of cell division (Kang et al., 2019). This antibiotic production is another factor to be considered with regards to *C. scindens* protection, with no further work carried out at present.

Colonisation Resistance by Other 7α-dehydroxylating bacteria

Despite identification of other organisms that encode the *bai* operon and carry out 7α-dehydroxylation, unlike *C. scindens* there is little direct evidence for their involvement in colonisation resistance against *C. difficile* in the human gut. *P. hiranonis* has been implicated in CDI resistance in a canine model (Stone et al., 2019) and has the second highest activity of the 7α-dehydroxylating Clostridia (Doerner, Takamine, Lavoie, Mallonee, & Hylemon, 1997), so may present an alternative opportunity to further understanding of bile acid-mediated CDI resistance; this will be explored later in this thesis. *P. hylemonae* was the first 'low activity' 7α-dehydroxylating strain in which a *bai* operon was identified (Ridlon et al., 2010), and has contributed to the

understanding of the 7α-dehydroxylation metabolic pathway (Funabashi et al., 2020). Both of these species have undergone preliminary *in vitro* characterisations, for example transcriptomic analyses by Ridlon et al. (2020) demonstrated differential expression of 197 and 118 genes for *P. hiranonis* and *P. hylemonae* respectively when grown in the presence of CA, however research into their individual colonisation resistance against *C. difficile* is largely limited to *in vitro* work by Reed et al. (2020). When grown in the presence of 2.5 mM of CA, neither the supernatant of *P. hiranonis* nor *P. hylemonae* resulted in the inhibition of *C. difficile* growth *in vitro*¸ unlike that of *C. scindens*. This is likely explained by the higher rate of 7α-dehydroxylation by *C. scindens*, however it is a surprising result, particularly as there was no DCA production detected by *P. hylemonae*. This perhaps highlights the limitations of the assay used, with only one set of conditions tested, and suggests further work is needed. Reed et al. (2020) also showed that *C. difficile* was able to outcompete *P. hiranonis* and *P. hylemonae* in co-culture experiments, but this was not tested in the presence of CA.

Research into other 7α-dehydroxylating species is also limited. *P. bifermentans* has been demonstrated as a single species mediator of protection from lethal *C. difficile* disease in a murine host, but the authors suggest this is likely due to its modulation of nutrient availability rather than its 7α-dehydroxylation activity (Girinathan et al., 2021). This has not been explored further. Moreover, whilst a *bai* operon has been identified in *C. sordellii* and *C. leptum* there has been no further study of either species with regards to 7α-dehydroxylation.

Overall, these organisms may assist in future understanding of CDI resistance by gut microbes, but at present their precise roles in the colon are unclear; their potential roles are proposed mainly due to their similarities to *C. scindens* and based on limited *in vitro* results. Further characterisation would be required to establish if they are involved in colonisation resistance, and what mechanisms are involved.

1.3 Genetic Modification of Clostridia

1.3.1 ClosTron

Historically the genetic manipulation of clostridia was limited, but advancement came in the form of insertional mutagenesis technologies. For gene targeting and disruption these technologies utilise mobile group II introns, site-specific retroelements that are found in bacterial genomes (Karberg et al., 2001). These catalytic RNAs create insertions in DNA via the process of retrohoming, in which they are exised from RNA transcripts and inserted into a new target site. The high efficiency and specificity of this process, coupled with the ability to retarget to any desired DNA site easily, has led to the exploitation of group II introns in genetic tools, namely Targetron and ClosTron (Karberg et al., 2001; Sarah A. Kuehne & Minton, 2012)

The ClosTron is a derivative of the Sigma Aldrich Targetron system with the advantageous development to allow for selection of successful insertion. ClosTron is a universal, clostridial gene knock-out system based on the mobile group II intron from the *ltrB* gene of *Lactococcus lactis* (Ll.ltrB). The *ltrB* is localised to the pMTL plasmid backbone, with other requisite ClosTron components (Figure 1.8). The Intron-Encoded Protein (IEP) mediates the insertion of the intron, encoded by *ltrA*, and can be manipulated easily for target insertion into any region within the genome. Intron insertion is selected for by acquisition of erythromycin resistance, as a retrotransposition activated marker (RAM) based on the *ermB* gene is used. Whilst advantageous, this does mean that subsequent mutations cannot be generated due to the same marker being required (Heap, Pennington, Cartman, Carter, & Minton, 2007; Sarah A. Kuehne & Minton, 2012). ClosTron technology has been used to generate mutants in a wide range of clostridia, for example histidine kinase mutants in *Clostridium beijerinckii* (Humphreys, Debebe, Diggle, & Winzer, 2023), a *dltD* mutant in *C. butyricum* (Wydau-Dematteis et al., 2015), and sigma factor mutants in *C. difficile* (Pereira et al., 2013).

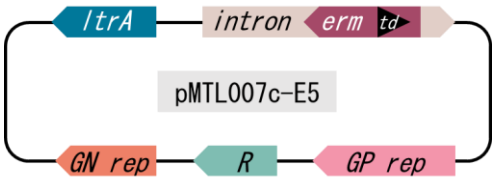


Figure 1.8 Schematic of ClosTron vector pMTL007c-E5. ClosTron specific components include: ***ltrA*** = gene encoding the Intron-Encoded Protein; **intron** = *ltrB* gene from *Lactococcus lactis*, including the retrotransposition- activated marker (RAM) (***erm*** + **td**). Insertion of the *ermB* gene inactivated by the phage *td* group I intron allows selection of group II intron insertion through acquisition of erythromycin resistance. Carried on pMTL84151 backbone where components include: **GN rep** = Gram-negative replicon (ColE1 + *traJ*); **R** = Resistance marker (Cm^R); **GP rep** = Gram-positive replicon (pCD6).

1.3.2 Allele Exchange

Until recently, the two main methods used for the introduction of DNA into clostridial genomes were based either on retargeting of group II introns (eg. the ClosTron) or on recombination to bring about allelic exchange (AE). The utility of the latter is enhanced by the provision of selection markers that allow the isolation of the desired mutants in which recombination has taken place. For many years, these were counter selection markers, such as *pyrE*, but in more recent years have been widely replaced by CRISPR-Cas. The former has found particular use in a special form of AE termed Allele-Couple Exchange (ACE) which, through the use of asymmetric homology arms, allows the rapid, selectable insertion of DNA into the genome at defined loci without the use of counter selection markers (Heap et al., 2012).

The *pyrE* locus has been widely exploited in many different clostridia for both the development of gene KO systems and as a locus for the insertion of application specific modules (reviewed by Minton et al. (2016)). Its exploitation relies on the dual and opposite phenotypes conferred by the mutant and wildtype allele, and the implementation of ACE technology. A functional *pyrE* gene is required for uracil production. It encodes orotate phosphoribosyl transferase, which converts the pyrimidine intermediate orotic acid into orotidine 5'-monophosphate (OMP). This reaction allows for selection of a *pyrE* mutant; fluoroorotic acid (FOA), an analogue

of orotic acid, is toxic in a wildtype *pyrE* strain as it is eventually converted to the toxic 5-fluorouridine monophosphate (5-FUMP). Inactivation of *pyrE* via a gene truncation, however, prevents 5-FUMP accumulation and confers a resistant phenotype. Moreover, the inactivation of uracil production and requirement for exogenous uracil of the *pyrE* mutant also allows selection for complementation back to wildtype; the auxotrophic mutant is restored to prototrophy and will again grow on media lacking uracil. Procedures based on exploiting *pyrE* can be applied to all clostridia as they contain the requisite pyrimidine pathway that allows generation of the auxotrophic mutant. Moreover, in theory this principle can also be applied at other loci that allow an auxotroph to be generated.

A *pyrE* deletion mutant can be used to select for insertion of application specific elements by concomitant gene repair and restoration of prototrophy. The ACE complementation vectors used comprise: (i) a replication defective Gram-positive replicon; (ii) an antibiotic resistance gene marker; (iii) a 300 bp left homology arm (LHA) encompassing the 3' end of *pyrE*, and; (iv) a longer, 1200 bp right homology arm (RHA) from immediately downstream of *pyrE*. The longer RHA directs the first recombination event which leads to plasmid integration. Cells in which integration has occurred grow visibly faster (larger colonies) on the selective media used that is supplemented with antibiotic, and uracil, as every cell carries a copy of the plasmid-encoded antibiotic resistance gene. This is not the case with cells in which the antibiotic resistance gene is localised to the replication defective ACE vector, which therefore grow slower. The faster growing, single crossover integrants may then be plated on minimal media lacking both antibiotic and uracil. This selects for cells in which the LHA-mediated, second recombination has occurred that restores the *pyrE* locus to WT (Minton et al., 2016).

ACE can be used to make genome alterations in clostridia, providing a *ΔpyrE* strain has been generated. This has been exemplified in *C. difficile* where in frame deletions were generated for three separate genes, *spo0A, cwp84, and mtlD* (Ng et al., 2013). Counter selection for gene modification was achieved in the presence of FOA using a heterologous *pyrE* allele from *C. sporogenes,* avoiding recombination with the native *pyrE*. The *ΔpyrE* strain can also be used to select for insertion of cargo into the

chromosome downstream of *pyrE*, with insertion concomitant with a return to prototrophy. This can be utilised for complementation of an inactivated gene, avoiding the use of multicopy plasmids, or overexpression of the inactivated gene, demonstrated by the P_{fdx} driven expression of the *mtlD* gene in the *C. difficile ΔmltD* strain (Ng et al., 2013). ACE can also allow the insertion and stable expression of other application specific modules at *pyrE*, for example the insertion in *C. sporogenes* of a cassette to express the nitroreductase from *Neisseria meningitidis* as a cancer therapy delivery vehicle (Heap et al., 2014). This application will be explored in this dissertation.

1.3.3 CRISPR-Cas

The development of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and Cas (CRISPR-associated) systems has offered new methods of genetic modification, and this has been utilised in clostridia as a selection tool. Following introduction of the desired editing template, the CRISPR-Cas system is able to counter select against cells that have not undergone homologous recombination and allele exchange. This allows a high frequency of mutants in the population and improved editing efficiency.

The most commonly used system is the Class II CRISPR-Cas9 from *Streptococcus pyogenes*, which contains a single effector protein in the form of Cas9. CRISPR arrays are introduced into target organisms via a vector, and the Cas9 nuclease is targeted to the DNA sequence of interest by a chimeric guide RNA (gRNA). The system is designed to target wildtype sequences that are only present if homologous recombination of the editing template has not occurred. As Cas9 contains two nucleolytic domains it introduces a lethal double strand break at the target site, thus removing wildtype cells from the population (Maikova, Kreis, Boutserin, Severinov, & Soutourina, 2019; S. Wang et al., 2018). CRISPR-Cas9 has been used to generate several different mutants in clostridia, for example deletion of *spo0A* and insertion of a GFP gene (S. Wang et al., 2018), mutants defective in selenoprotein synthesis (McAllister, Bouillaut, Kahn, Self, & Sorg, 2017) and generation of an erythromycin-sensitive strain of *C. difficile* (Ingle et al., 2019).

Whilst CRISPR-Cas9 systems have had many successes, the obstacles of low transformation efficiencies and excessive Cas9 toxicity can be experienced. To overcome these, Cañadas et al. (2019) have generated a modified system for clostridia, known as RiboCas. The expression of *cas9* is controlled by a theophylline inducible P_{fdx} promoter, reliant on the positioning of a riboswitch downstream of the promoter. Without theophylline the riboswitch forms a stem-loop structure that prevents ribosome binding, but when theophylline binds it causes a conformational change and allows ribosome binding and translation. RiboCas is a highly efficient tool that can improve the performance of CRISPR in clostridia, and this is attributed to the riboswitch allowing homologous recombination to occur before Cas9 activity, in addition to the control of Cas9 cellular toxicity through control of its expression levels (Cañadas et al., 2019).

1.4 Objectives and thesis organisation

The overall aim of this thesis was to improve the understanding of the 7α-dehydroxylation pathway by providing genetic tools to allow the study of different combinations of *bai* genes, in order to further elucidate enzyme functions and redundancies. In doing so it was also hoped to provide an insight into the proposed mechanism of bile-acid mediated colonisation resistance against *C. difficile* observed in 7α-dehydroxylating bacteria. Given the previous difficulties in genetically modifying 7α-dehydroxylating bacteria, it was decided to approach this goal from two different angles, and this formed the two main projects presented: the study of native 7α-dehydroxylation in *P. hiranonis*; and the development of a *C. butyricum* strain for expression of *bai* genes.

Chapter 3 describes the establishment of DNA transfer in *P. hiranonis*, the use of existing tools to attempt to generate gene knockouts, and the optimisation of tools for use in *P. hiranonis*. It also explores bile-acid mediated resistance of *P. hiranonis* against *C. difficile* in the form of germination and vegetative growth.

Chapter 4 outlines the establishment and optimisation of DNA transfer in *C. butyricum* and the development and characterisation of auxotrophic knockout strains for use with ACE technologies. Promoter activities are also characterised, providing tools for heterologous gene expression.

Chapter 5 utilises the *C. butyricum* knockout strains as a chassis for expression of *C. scindens bai* genes and investigates the mobilisation of 7α-dehydroxylation activity.

Finally, Chapter 6 considers the impact of the two projects as a collective and discusses the future directions of this field.

Chapter 2 Materials and Methods

2.1 Bacterial Strains and Plasmids

2.1.1 Bacterial Strains

All strains of bacteria used in this study are summarised in Table 2.1.

Table 2.1 Bacterial strains

Strain	Description	Source
E. coli TOP10	MC1061 derivative. General cloning, plasmid propagation and storage: F− *mcrA* Δ*(mrr–hsdRMS– mcrBC) φ80lacZΔM15* Δ*lacX74 recA1deoR araD139* Δ*(ara– leu)7697 galU galK rpsL (StrR) endA1 nupG*	Invitrogen
E. coli CA434	Conjugal donor strain. *E. coli* HB101 carrying R702 (an IncP plasmid with ori-T mediated transfer functions): *E. coli HB101* [F- *mcrB mrr hsdS20(rB- mB-) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20(Smr) glnV44λ-*] *R702*	Williams et al. (1990)
E. coli sExpress	Conjugal donor strain with increased efficiency. Lacks Dcm methylation, allowing circumvention of Type IV restriction systems: *fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb- 210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10, R702*	Woods et al. (2019)
E. coli XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacIqZΔM15* Tn10 (Te^{tr})].	Agilent
C. scindens ATCC 35704	Type strain	Jason Ridlon, Virginia Commonwealth University
C. difficile R20291	PCR-ribotype 027 (Stoke Mandeville, UK)	Anaerobe Reference Laboratory, Cardiff
C. difficile 630 NCTC 13307	PCR-ribotype 012 (Zurich, Switzerland)	NCTC, Public Health England
P. hiranonis DSM 13275	Type strain	Solenne Marion (EPFL, Lausanne, Switzerland)
P. hiranonis ΔbaiCD	*baiCD* in-frame deletion mutant created using RiboCas	This study
C. butyricum NCTC 7423	Type strain	Mike Young (Aberystwyth)
C. butyricum Δpglx	*pglx* in-frame deletion mutant created using RiboCas	This study
C. butyricum Δpglx::pglx	*pglx* mutant repaired to wildtype using RiboCas	This study
C. butyricum ΔpyrE	*pyrE* in-frame deletion mutant created using RiboCas	This study
C. butyricum ΔpurD	*purD* in-frame deletion mutant created using RiboCas	This study
C. butyricum ΔhisI	*hisI* in-frame deletion mutant created using RiboCas	This study
C. butyricum ΔpyrE::pyrE	*pyrE* mutant repaired to wildtype using ACE	This study

C. butyricum ΔpurD::purD	*purD* mutant repaired to wildtype using ACE	This study
C. butyricum ΔhisI::hisI	*hisI* mutant repaired to wildtype using ACE	This study
C. butyricum ΔpyrEΔpurD	*C. butyricum ΔpyrE* with *purD* in-frame deletion mutant created using RiboCas	This study
C. butyricum ΔpyrEΔpurDΔhisI	*C. butyricum ΔpyrEΔpurD* with *hisI* in-frame deletion mutant created using RiboCas	This study
C. butyricum::baiG	*C. butyricum ΔpyrE* with *baiG* (*C. scindens*) supplied by ACE at *pyrE*. *pyrE* repaired to wildtype. Abbr. to TR strain in text.	This study
C. butyricum::baiG-baiB	*C. butyricum ΔpyrEΔpurD* with *baiG* (*C. scindens*) supplied by ACE at *pyrE*. *baiB* (*C. scindens*) supplied by ACE *at purD*. Both *pyrE* and *purD* repaired to wildtype during ACE. Abbr. to BE1 in text.	This study
C. butyricum::baiG-baiB-baiCD-baiE	*C. butyricum ΔpyrEΔpurD* with *baiG* (*C. scindens*) supplied by ACE at *pyrE*. *baiB, baiCD* and *baiE* (*C. scindens*) supplied by ACE at *purD*. Both *pyrE* and *purD* repaired to wildtype during ACE. Abbr. to BE3 in text.	This study
C. butyricum::baiG-baiB-baiCD-baiE-baiJ	*C. butyricum ΔpyrEΔpurD* with *baiG* (*C. scindens*) supplied by ACE at *pyrE*. *baiB, baiCD* and *baiE* (*C. scindens*) supplied by ACE at *purD*. baiJ (*C. scindens*) supplied by RiboCas at *hisI*. Both *pyrE* and *purD* repaired to wildtype during ACE. Abbr. to BE4 in text.	This study

2.1.2 Plasmids

A detailed list of all plasmid constructs and strains used in this study can be found in Appendix A.1.

2.2 Culture Media and Growth Conditions

2.2.1 Culture Media

Bacterial culture media is listed in Table 2.2. Media sterilisation was carried out by autoclaving at 121°C at 100kPa above atmospheric pressure for 15 minutes.

Table 2.2 Culture Media

Media	Target Organism	Components
Lysogeny Broth (LB)	*E. coli* strains	10 g tryptone, 10 g NaCl, 5 g yeast extract in 1 L dH_2O.
		For solid plate media, 15 g of Oxoid agar No. 1 is added.
Brain Heart Infusion Supplemented (BHIS)	*C. difficile* strains	37 g brain heart infusion (Oxoid, CM1135), 5 g yeast extract, 1 g L-cysteine in 1 L dH_2O.
		For solid plate media, brain heart infusion is replaced with 47 g of brain heart infusion agar (Oxoid, CM1136).
Brain Heart Infusion Salt Supplemented (BHISS)	*P. hiranonis* strains	As for BHIS, with 2 g fructose, 40 mL BHISS salt solution (detailed below).
		Salt solution: 0.2 g $CaCl_2$, 0.2 g $MgSO_4$, 1 g K_2HPO_4, 1 g KH_2PO_4, 10 g $NaHCO_3$, 2 g NaCl in 1 L dH_2O.

Reinforced Clostridial Media (RCM)	*C. butyricum* strains	38 g of reinforced clostridial medium (Oxoid, CM0149) in 1 L dH_2O.
		For solid plate media, 52.5 g reinforced clostridial agar (Oxoid, CM0151) in 1 L dH_2O.
2xYTg	*C. butyricum* strains	16 g tryptone, 5 g NaCl, 10 g yeast extract in 1 L dH_2O with 2% glucose.
		For solid plate media 15 g of Oxoid agar No. 1 is added.
Minimal Media (MM)	*C. butyricum* strains	For 1 L total volume solid media: 55 mL dH_2O, 100 mL amino acid solution (5X), 50 ml salt solution (10X), 25 mL glucose solution (20% w/v), 10 mL trace salt solution (50X), 5 mL $FeSO_4.7H_2O$ solution (100X), 5 mL vitamin solution (100X) and 500 mL agar (20%).
		Amino acid solution (5X): 10 g Cas-amino acids, 0.5 g tryptophan and 0.5 g cysteine in 200 mL dH_2O. Salt solution (10X): 1.8 g KH_2PO_4, 10 g Na_2HPO_4, 1.8 g NaCl and 10 g $NaHCO_3$ in 200 mL dH_2O. Trace salt solution (50X): 260 mg $CaCl_2.2H_2O$, 200 mg $MgCl_2.6H_2O$, 100 mg $MnCl_2.4H_2O$, 400 mg $(NH_4)_2SO_4$ and10 mg $CoCl_2.6H_2O$ in 200 mL dH_2O. $FeSO_4.7H_2O$ solution (100X): 20 mg $FeSO_4.7H_2O$ in 50 mL dH_2O. Vitamin solution (100X): 20 mg Ca-D-pantothenate, 20 mg pyridoxine and 20 mg d-biotin in 200 mL dH_2O.

2.2.2 Media Supplements

Media supplements were prepared as stock solutions and filter sterilised using 0.2 µm Minisart NML syringe filters (Sartorius). Solutions were stored in the recommended conditions for no longer than 6 weeks. Antibiotic working concentrations are listed in Table 2.3.

Table 2.3 Supplements used in this study

Supplement	Stock concentration (mg/mL)	Solvent	Working concentration for *E. coli* (µg/mL)	Working concentration for clostridium (µg/ml)
Chloramphenicol	25	100% EtOH	12.5 (broth), 25 (plates)	-
Thiamphenicol	15	50% EtOH	-	15
D-cycloserine	50	dH_2O	-	250
Erythromycin	50	100% EtOH	500	10
Spectinomycin	250	dH_2O	250	250
Tetracycline	5	70% EtOH	10	10
Cefoxitin	50	dH_2O	-	8
Anhydrous tetracycline	2	70% EtOH	-	0.06
Theophylline	48	DMSO	-	0.1-10 mM
Lactose	100 mM	dH_2O	-	0.1-10 mM
Uracil	1	dH_2O	-	5

Histidine	20	dH_2O	-	20
Adenine	2.7	0.05 mM NaOH	-	135
Guanine	3	0.05 mM NaOH	-	151
Hypoxanthine	2.7	0.05 mM NaOH	-	136
Thiamine	1.5	dH_2O	-	1.5
5-Fluoroorotic acid (5-FOA)	100	DMSO	-	3000
Sodium taurocholate	100	dH_2O	-	100
Cholic acid sodium salt	4.3	dH_2O	-	0.25-10 mM
Deoxycholic acid sodium salt	1.25	dH_2O	-	0.15–3 mM

2.2.3 Growth Conditions

E. coli strains were cultured aerobically at 37°C. Liquid cultures were cultured with horizontal shaking at 200 G.

Clostridia strains were cultured anaerobically at 37°C in an MG1000 Mark II anaerobic workstation (Don Whitley Scientific Ltd), with an internal atmosphere of N_2:CO_2:H_2 at 80:10:10 (v/v/v). All culture media were pre-reduced for a minimum of 4 hours for solid media and 12 hours for liquid media.

2.2.4 Storage Conditions

To make frozen stocks for storage bacterial strains were grown on appropriate solid media and growth harvested with a 10 µL inoculation loop.

E. coli strains were resuspended and stored in Microbank Long Term Bacterial Storage tubes (Pro-Lab Diagnostics). Clostridial strains were resuspended in 1 mL of appropriate liquid broth with 10% (v/v) glycerol, and stored in screw cap tubes.

All bacterial stocks were stored at -80°C.

2.3 Microbiological Methods

2.3.1 Preparation of Chemically Competent E. coli

Overnight cultures of *E. coli* were prepared and subsequently added to pre-warmed LB broth at a ratio of 1:200 mL. Cultures were incubated at 37°C under appropriate

growth conditions until an OD_{600} of 0.5-0.7 was reached. For optimised handling the culture was divided into two 40 mL falcon tubes, and cells were incubated on ice for 15 minutes. Cells were then harvested by centrifugation at 2,000 x g for 10 minutes at 4°C. The pellets were washed and combined in 20 mL of ice cold $CaCl_2$ (100 mM), incubated on ice for 30 minutes, then centrifuged under the same conditions. The pellet was then resuspended in 2 mL ice cold $CaCl_2$ (100 mM) + 15 % glycerol, and aliquoted and stored in 50 µL volumes at -80 °C.

2.3.2 Conjugal Transfer of Plasmids

Overnight cultures of *E. coli* donor cells were prepared in 5 mL of LB broth with appropriate antibiotic selection until stationary phase was reached. The recipient clostridia was incubated anaerobically overnight in appropriate liquid media in a 10 fold dilution series. A 1 mL sample of *E. coli* overnight culture was centrifuged at 2,000 x g for 1 minute and washed twice with 400 µL of PBS to remove antibiotic. The pellet was then transferred to the anaerobic cabinet and resuspended in 200 µL of the recipient culture, using the most dilute overnight culture displaying growth. The resultant mixture was then plated onto an agar plate of appropriate media in individual 20 µL spots and incubated for 24 hours. The growth was harvested and suspended in 500 µL liquid broth before being spread onto agar plates supplemented with D-cycloserine and appropriate antibiotic selection for the conjugal plasmid. The plates were incubated until distinct transconjugant colonies appeared. Colonies were re-streaked to purify on the same selective media.

2.3.3 Conjugation Efficiency Assay

For each plasmid to be assayed a conjugation, as outlined in 2.3.3, was carried out in triplicate. The growth of the conjugal mixture, harvested after 24 h and resuspended in 500 µL liquid broth, was serially diluted in the same broth. The dilution series was plated onto D-cycloserine, with 20 µL in triplicate per dilution, to calculate the CFU/mL of the conjugal acceptor. The neat mixture and 10^{-1} were plated onto D-cycloserine and appropriate antibiotic selection for the conjugal plasmid to calculate the CFU/mL of transconjugants. Conjugation efficiency is the percentage of transconjugants from the total conjugal donor.

2.3.4 Bacterial Growth

Where OD_{600} readings were taken of liquid cultures >1 mL, these were read in a BioMate 3 spectrophotometer (Thermo Scientific) using sample volumes of 1 mL. Dilutions were made as required.

Unless otherwise stated, microplate assays were carried out in clear flat bottomed 96-well plates (Costar, Corning®), with OD_{600} readings taken using the CLARIOstar plate reader (BMG Labtech).

2.3.5 Bacterial Growth Curves

Clostridia strains grown on solid media, of not more than 2 days old, were used to prepare a 10-fold dilution series of overnight cultures by inoculating appropriate liquid media. The overnight culture that was most dilute with growth was used to inoculate 50 mL of pre-reduced liquid media to a starting OD_{600} of 0.05. Cultures were incubated anaerobically, and optical density readings were taken at a wavelength of 600 nm every hour over a 12 h period, in addition to a reading at 24 h.

2.3.6 Minimum Inhibitory Concentration Assay

A broth dilution method was used to ascertain minimum inhibitory concentrations in clostridia. A 96 well microplate (Greiner Bio-One) was set up with antibiotics diluted in appropriate broth media, in a range appropriate to the standard Clostridia working concentration, and pre-reduced for 3 h. The antibiotic supplemented broth was then inoculated with independent overnight cultures, and incubated anaerobically for 24 h. OD_{600} readings of the microplate were then taken. The MIC value was calculated as the lowest concentration at which no growth was observed.

2.3.7 Plasmid Interference Assay

To assess the transfer efficiency of putative PAM sequences a conjugation efficiency assay was carried out according to 2.3.3. A plasmid containing the putative PAM upstream of the protospacer sequence was compared to the plasmid control, containing 5 nucleotides of direct repeat upstream of the protospacer sequence.

2.3.8 FAST promoter assay

Strains harbouring the relevant promoter/FAST plasmid were plated onto selection and grown. An overnight culture in relevant liquid media with selection was then used to inoculate fresh liquid media with selection. Where appropriate inducers were added to the culture either from 0 h or at a determined time point. Cultures were grown until mid-exponential phase. Samples taken were normalised to an OD_{600} of 1 by dilution in PBS, or to the lowest OD_{600} of all samples taken. Samples were centrifuged at 12,000 x g for 1 min and washed in 1 mL PBS and this was carried out twice before resuspending the final pellet in 250 µL of PBS.

The assay requires the TFAmber fluorophore (The Twinkle Factory Ltd, 499558-250) resuspended in DMSO to yield a 5 mM stock solution. This was used to generate a working solution of 10 µM by dilution in DMSO.

In a black walled, clear and flat-bottomed microplate (Corning™, 10530753) the prepared sample was diluted 1:1 with the fluorophore and measured at an excitation wavelength of 499 nm and emission at 558 nm following 5 seconds of agitation.

2.3.9 Bile Acid Analysis

Bile acid analysis was performed by collaborators at the École Polytechnique Fédérale de Lausanne (EPFL), according to the following protocol.

Bile acid extraction

Both the WT and *ΔbaiCD* strains of *P. hiranonis* were cultured in BHISS liquid media supplemented with 100 µM CA at 0 h. Samples were vacuum dried overnight (evaporated in a SpeedVac, RVC 2-33 CDplus Infrarot, Martin Christ, Germany). The dried cells were collected and approximately 450 mg of 0.5 mm zirconium beads were added to each tube. 1200 µL of MeOH/H2O (2/1) + 0.1% formic acid was used as extraction solvent. Samples were homogenised in a Precellys 24 Tissue Homogenizer (Bertin Instruments, Montigny-le-Bretonneux, France) at 5,000 xg 2x 20” beat and 20” rest. The homogenized samples were centrifuged at 21,000 x g, for 15 min, at 4°C. 10 µL from each supernatant and 100 µL from each calibration standard were transferred into individual wells of 2 mL 96-well plate. 50 µL of an ISTD solution

(CA-d4, CDCA-d4, DCA-d4 and LCA-d4, each at 2 µM in methanol) was pipetted in each well. Immediately after the addition of ISTD, 600 µL of 0.2% formic acid in H_2O was added to each sample or calibration standard level. The 96-well plate was shaken with an orbital shaker at 300 rpm and centrifuged at 1500 x g, 5 min, 4°C.

The contents of the 96-well plate were extracted by solid phase extraction with an Oasis HLB 96-well uElution plate. The extracted samples were dried in a Biotage® SPE Dry 96 at 20°C and reconstituted with 100 µL of MeOH/H_2O (50/50). The plate was shaken with an orbital shaker at 300 rpm, 5 min and centrifuged at 1500 x g, 5 min, 4°C. The samples were injected on the LC-HRMS system.

Liquid chromatography – mass spectrometry (LC-MS)

The quantitative method was performed on an Agilent ultrahigh-performance liquid chromatography 1290 series coupled in tandem to an Agilent 6530 Accurate-Mass Q-TOF mass spectrometer (https://www.biorxiv.org/content/10.1101/2022.02.15.480494v2.full). The separation was done on a Zorbax Eclipse Plus C18 column (2.1 × 100mm, 1.8 µm) and a guard column Zorbax Eclipse Plus C18 (2.1 × 5mm, 1.8 µm) both provided by Agilent technologies (Santa Clara, CA, USA). The column compartment was kept heated at 50°C. Two different solutions were used as eluents: ammonium acetate [5 mM] in water as mobile phase A and pure acetonitrile as mobile phase B. A constant flow of 0.4 mL/min was maintained over 26 minutes of run time with the following gradient (expressed in eluent B percentage): 0-5.5 min, constant 21.5% B; 5.5-6 min, 21.5-24.5% B; 6-10 min, 24.5-25% B; 10-10.5 min, 25-29% B; 10.5-14.5 min, isocratic 29% B; 14.5-15 min, 29-40% B; 15-18 min, 40-45% B; 18-20.5 min, 45-95% B; 20.5-23 min, constant 95% B; 23-23.1 min, 95-21.5% B; 23.10-26 min, isocratic 21.50% B. The system equilibration was implemented at the end of the gradient for 3 minutes in initial conditions. The autosampler temperature was maintained at 10°C and the injection volume was 5 µL. The ionisation mode was operated in negative mode for the detection using the Dual AJS Jet stream ESI Assembly. The QTOF acquisition settings were configured in 4GHz high-resolution mode (resolution 17000 FWHM at m/z 1000), data storage in profile mode and the high-resolution full MS chromatograms were acquired over the range of m/z 100-1700 at a rate of 3 spectra/s. The mass spectrometer was calibrated in negative mode using ESI-L solution from

Agilent technologies every 6 hours to maintain the best possible mass accuracy. Source parameters were setup as follows: drying gas flow, 8 L/min; gas temperature, 300°C; nebulizer pressure, 35psi; capillary voltage, 3500V; nozzle voltage, 1000V. Data were processed afterwards using the MassHunter Quantitative software and MassHunter Qualitative software to control the mass accuracy for each run. In the quantitative method, 11 bile acids were quantified by calibration curves. The quantification was corrected by addition of internal standards in all samples and calibration levels. Extracted ion chromatograms were generated using a retention time window of ± 1 min and a mass extraction window of ± 40 ppm around the theoretical mass of the targeted bile acid.

2.3.10 Spore preparations

The strain of interest was inoculated into 1 mL of pre-reduced liquid media for an overnight culture. This was then plated onto appropriate solid media in 100 µL aliquots and left to incubate anaerobically for five days. Plates were then removed from the anaerobic workstation and incubated overnight at 4°C. Growth was then harvested by scraping the plate with a 10 µL loop and resuspended in ice cold dH_2O, with 3 plates collected per 1 mL. Following overnight incubation at 4°C the pellet was then resuspended with gentle aspiration and centrifuged (16000 x g, 4°C, 4 minutes). The spore pellet was then washed by removal of supernatant and debris followed by resuspension in 1 mL of ice water and centrifugation repeated. This was repeated at least 10 times until a purity of ~95% was achieved, estimated by visualisation under the upright Eclipse Ci phase-contrast microscope (Nikon). Spore suspensions in dH_2O were stored at 4°C.

2.3.11 Supernatant preparation

Strains of interest were grown in appropriate liquid media with and without CA supplementation at the relevant concentration in an overnight culture. Samples were taken and centrifuged (16,000 x g for 5 mins), the supernatant taken and filter sterilised using 0.2 µm Minisart NML syringe filters (Sartorius). Samples were used immediately or stored at 4°C until use.

2.3.12 Spore outgrowth assay

Supernatants were prepared for the strain(s) of interest incubated with and without CA, according to 2.3.11. The supernatants obtained were diluted 1:1 in 2 x concentrated BHIS broth and supplemented with 0.1% sodium taurocholate (TCA). Control samples were also prepared, consisting of: BHIS; BHIS +0.1 % TCA; BHIS +0.1 % TCA + 2 mM DCA. 180 µL of supernatant was added to a 96-well flat-bottomed microplate (Costar, Corning®) and reduced in the anaerobic cabinet for a minimum of 3 hours.

C. difficile spore preparations for the relevant strain(s) were prepared according to 2.3.10, and diluted to an OD_{600} of 0.5 in dH_2O. Spores were heat treated at 65°C for 30 minutes and then 20 µL added to the reduced supernatants for a 1:10 dilution. The microplate was sealed using a plate sealer (R&D Systems) and transferred to a pre-warmed CLARIOstar plate reader (BMG Labtech). The plate was incubated at 37°C and OD_{600} readings were taken every hour for 24 hours, following a 6 second period of orbital shaking.

2.3.13 Germination initiation

Supernatants were prepared for the strain(s) of interest incubated with and without CA, according to 2.3.11. The supernatants obtained were diluted 1:1 in 2 x concentrated BHIS broth and supplemented with 0.2% sodium taurocholate (TCA). Control samples were also prepared, consisting of: BHIS; BHIS +0.2 % TCA; BHIS +0.2 % TCA + 4mM DCA. 75 µL of supernatant was added to a 96-well flat-bottomed microplate (Costar, Corning®).

C. difficile spore preparations for the relevant strain(s) were prepared according to 2.3.10, and diluted to an OD_{600} of 0.5 in dH_2O. Spores were heat treated at 65°C for 30 minutes then centrifuged for 1 min at 12,000 x g. The supernatant was discarded and the pellet suspended in BHIS to an OD_{600} of 1. 75 µL of spores was added to the supernatants for a 1:1 dilution. The microplate was transferred to a CLARIOstar plate reader (BMG Labtech) and OD_{600} readings were taken every minute for 10 mins, then every 10 minutes following a 6 second period of orbital shaking, for a total period of

100 mins. Germination initiation was assessed as an initial drop in OD_{600}, indicating spore rehydration.

2.3.14 Transwell plate assays

The respective BHISS liquid media with and without bile acid supplementation was loaded into the wells of a Transwell™ 12-well plate with permeable polycarbonate membrane inserts (Fisher Scientific). For both spore and vegetative cell assays the centre (upper) well was used for *C. difficile* growth, with 540 µL of BHISS, and the outer (lower) well was used for *P. hiranonis* growth, with 1660 µL of BHISS. The plate was placed into the anaerobic cabinet and reduced for a minimum of 3 hours.

For both assays an overnight culture from a dilution series of *P. hiranonis*, both the WT and *ΔbaiCD* strains, was used to inoculate 40 µL into the centre well, for an approximate OD_{600} of 0.05. For the spore assay, *C. difficile* spore preparations (2.3.10) for the relevant strain were heat treated at 65 °C for 30 minutes, and 60 µL inoculated into the outer well. TCA was also added to all *C. difficile* wells at 0.1%. For the vegetative growth assay an overnight culture from a dilution series of *C. difficile* was used to inoculate 60 µL into the outer well, for an approximate OD_{600} of 0.05. Plates were incubated for 8 or 24 hours with gentle agitation (50 rpm). At the chosen time points, 4 and 8 hours for the spore assay, and 8 and 24 hours for the vegetative cell assay, a 100 µL sample was taken from both the centre and outer wells and diluted 1 in 10 to read in the CLARIOstar plate reader (BMG Labtech).

2.3.15 Growth competition assays

BHISS liquid media with appropriate CA or DCA, in 10 mL aliquots, was incubated in the anaerobic chamber for a minimum of 3 hours. An overnight dilution series of *P. hiranonis* WT, *P. hiranonis ΔbaiCD* and *C. difficile* R20291 was used to inoculate the BHISS, with the least concentrated dilution with growth chosen. *P. hiranonis* and *C. difficile* were then each inoculated to a starting OD_{600} of 0.05, either as mono-or co-cultures. Cultures were incubated for 24 hours. Samples were taken and a dilution series generated in BHISS. Onto pre-reduced BHISS plates 20 µL of each dilution was spotted in triplicate and the plates incubated for 48 hours. The colonies were counted

to calculate the CFU/mL of each species, using the differing morphologies to differentiate between *C. difficile* and *P. hiranonis* in the co-cultures.

2.3.16 Germination assays

Heat treatment of vegetative cells

C. butyricum was inoculated to a starting OD_{600} of 0.05 from an overnight dilution series, and cultured in RCM liquid medium. Samples were taken after 4 h and heat treated at 65°C or 85°C for 20 or 30 mins. A dilution series was generated in RCM, plated on RCM media and the CFU/mL calculated after 24 h.

Germination

A *C. butyricum* spore preparation was prepared according to 2.3.10 and heat treated at 65°C or 85°C for 20 mins. A sample was also taken from a 5-day old culture in RCM liquid media. For both spore preparation and culture sample, a dilution series was generated in RCM and plated on media: RCM, BHIS or RCM + fetal bovine serum (10%; Gibo), and the CFU/mL calculated after 24 h.

2.3.17 Chloramphenicol Acetyltransferase (CAT) Assay

CAT assay

Chloramphenicol Acetyltransferase (CAT) activity was measured using a protocol derived from Shaw (1975). *C. butyricum* strains harbouring the relevant promoter/*catP* plasmid were plated onto selection and grown. An overnight dilution series in RCM with selection was then used to inoculate fresh liquid media with selection. Where appropriate inducers were added to the culture at a determined time point. Cultures were grown until mid-exponential phase. Samples taken were normalised to an OD_{600} of 1, or to the lowest OD_{600} of all samples taken, and centrifuged (17,000 x g for 3 mins). Pellets were stored at -20°C until further processing.

Cell pellets were thawed on ice and washed by resuspension in 1 mL 1.5M NaCl. Samples were centrifuged (21,000 x g for 10 mins at 4°C) and the supernatant removed. The wash step was then repeated. Pellets were then resuspended in 250 µL

of lysis master mix (1x Bugbuster® protein extraction reagent (EMD Millipore™) supplemented 1:100 with Protease Inhibitor Cocktail Set VII (Calbiochem) and incubated at 37°C for 45 mins with constant agitation (220 rpm). Samples were then sonicated for 10 mins in a Bioruptor® (Diagenode) at 4°C, following 5 cycles of high intensity sonication with 30 seconds on and 30 seconds off. Samples were then centrifuged (21,000 x g for 15 mins at 4°C) and the supernatant removed to be kept on ice.

10 µL of sample and 90 µL 100 mM Tris buffer was added to each well in a 96-well microplate, for analysis on a Tecan Infinite® M1000PRO (TECAN) plate reader. An assay master mix was prepared for injection consisting of: 0.005% w/v chloramphenicol (Sigma Aldrich); 0.167mM acetyl-CoA sodium salt (Sigma Aldrich); 0.0833mM 5,5'-dithio-bis-[2-nitrobenzoic acid] (DTNB)(ThermoFisher Scientific); 100mM Tris buffer pH7.8 (Fisher Scientific). The assay was performed at 25°C, with automated injection of 100 µL of master mix to each sample well. Following 4 secs of orbital shaking, the reaction ran for 90 secs with the absorbance change at 412 nm recorded every 5 secs.

Bicinchoninic Acid (BCA) assay

Quantification of protein concentration was carried out for the samples used in the CAT assay, using the Pierce™ BCA Protein Assay Kit (Thermo Scientific) according to manufacturer's instructions detailed below.

1. Add 25 µL of sample and albumin standards (BSA) into a microplate well.
2. Add 200 µL of the working reagent (50:1, Reagent A:B) to each well and place on a plate shaker for 30 secs. Cover plate and incubate at 37°C for 30 minutes.
3. Cool plate to RT. Measure the absorbance at or near 562 nm on a plate reader.
4. Unknown samples concentrations were determined using the standard curve plotted on GraphPad Prism (Dotmatics).

Using the protein concentration quantified, the U/mg protein was calculated according to the following equation:

$$\frac{Units}{mg}protein = \frac{\frac{Units}{ml}enzyme}{mg\ protein/ml\ enzyme}$$

CAT activity

CAT Units per ml was calculated according to the following equation, where df = dilution factor and ε = micromolar extinction coefficient for DTNB at 412 nm (0.0136).

$$Units/ml\ CAT = \frac{(\Delta A_{412}/\text{ minute}\ test - \Delta A_{412}/\text{ minute}\ blank)(volume\ in\ assay)(df))}{(\varepsilon)(volume\ of\ cel\ lysate)}$$

2.3.18 Transport assay – BaiG

Plated stocks of the *C. butyricum* wildtype and TR strains were used to inoculate pre-reduced RCM liquid media, with and without theophylline (5 mM). Cholic acid (5 µM) was added at 5 h and samples taken after a further 1.5 h. Samples of 500 µL were centrifuged (5,000 x g for 5 mins) and 200 µL of supernatant was aqueous filtered. Bile acid analysis was then carried out by Dr David Tooth (University of Nottingham) as outlined below.

Calibration standards of 100uM-195nM cholic acid were prepared. To the samples and standards 800 µL hyodeoxycholic acid in methanol (20 µM) was added, pulse vortexed and stored at -20°C overnight. They were then pulse mixed and centrifuged (16,000 x g for 5 mins) and supernatants analysed by LC-MS.

The analytical conditions used were as follows: Reversed phase HPLC (Waters CORTECS T3 2.7um, 2.1x50mm) used a developed gradient of 20-80% methanol in 10mM ammonium acetate over 10 minutes, at a flow rate of 0.25 mL.min-1. Detection used Single Ion Recording (SIR) of acetate adducted negatively charged molecular ions at 467.2m/z and 451.2m/z for Cholic acid and Hyodeoxycholic acid respectively, using a triple quadrupole mass-spectrometer (Waters Quattro Ultima) equipped with an electrospray ion source. All instrument control, acquisition and data processing used MassLynx (v4.0) software.

2.4 Molecular Methods

2.4.1 Oligonucleotides

A detailed list of all oligonucleotides used in this study can be found in Appendix A.2.

2.4.2 Polymerase Chain Reactions

For proof reading applications, polymerase chain reactions (PCR) were carried out using Q5 High Fidelity 2x PCR Master Mix (NEB). Typically, reactions were carried out in a total volume of 25 µL, with 1.25 µL of each primer (10 mM), 12.5 µL 2x Q5 Master Mix, 9 µL of dH_2O and 1 µL of template DNA where appropriate. Reaction conditions are listed in Table 2.4.

For non-proof reading applications, PCRs were carried out using DreamTaq PCR Master Mix (2x) (Thermo Scientific). Typically, reactions were carried out in a total volume of 25 µL, with 0.2 µL of each primer (10 mM), 12 µL 2x DreamTaq Master Mix, 11.6 µL of dH_2O and 1 µL of template DNA where appropriate. Reaction conditions are listed in Table 2.4.

Colony PCRs were carried out according to the DreamTaq PCR protocol with an additional cycle of 95°C for 15 mins at the beginning.

Table 2.4 PCR conditions

	Q5 touchdown			DreamTaq touchdown		
Reaction Step	**Temperature**	**Time**	**Cycles**	**Temperature**	**Time**	**Cycles**
Initial Denaturation	98 °C	2 mins	1	95 °C	2 mins	1
Denaturation	98 °C	30 secs	10 per annealing temperature	95 °C	30 secs	10 per annealing temperature
Annealing	52°C – 62°C	30 secs		50°C – 60°C	30 secs	
Extension	72 °C	30-40 secs per Kb		72 °C	20 secs per Kb	
Final Extension	72 °C	10 mins	1	72 °C	10 mins	1
Hold	15 °C	-	-	15 °C	-	-

SOEing PCR

Where specified the editing template fragment for insertion into the RiboCas backbone was generated by SOEing PCR. Primers were designed to generate two homology arms, each with a region of overlapping homology. These were generated using the Q5 protocol, then purified. A further SOEing PCR was then carried out, again using

the Q5 protocol, with 1 µL of each homology arm added to the reaction mixture as template DNA.

Primer dimer PCR

Where specified small fragments used in cloning, for example sgRNA fragments, were generated by a primer dimer PCR. The Q5 protocol was used without the touchdown temperatures, the conditions were as follows: denaturation (98°C for 30 secs), annealing (60°C for 30 secs) and extension (72°C for 45 secs), for a total of 30 cycles.

2.4.3 PCR Product Purification

For primer dimer reactions a PCR clean-up was used to ensure a sufficient yield. A small PCR sample was run on a gel to assess PCR success. PCR products were purified using the QIAquick® PCR Purification Kit (Qiagen) according to the manufacturer's instructions, outlined below. All centrifugation steps were carried out at 17,000 x g.

1. Add 5 volumes Buffer PB to 1 volume of the PCR reaction and mix.
2. Apply the sample to a QIAquick column and centrifuge for 60 secs. Discard flow-through.
3. To wash, add 750 µL Buffer PE to the column and centrifuge for 60 secs. Discard the flow-through and repeat centrifugation of column.
4. Place the column into a 1.5 mL microcentrifuge tube, add 12 µL of dH_2O to the membrane, and centrifuge for 60 secs.

2.4.4 Agarose Gel Electrophoresis

DNA was separated on 1% agarose (Sigma) gels in 1x TAE buffer with 0.01% (v/v) SYBR Safe DNA Gel Stain (Thermo Scientific), and electrophoresis was run at 100V for 60 minutes. 6x purple loading dye (NEB) was adding to samples before loading, and subsequently visualised using a Gel Doc XR system (BioRad). A 1 Kb+ DNA ladder was used (NEB).

2.4.5 Extraction of DNA from Agarose Gels

DNA fragments were visualised under blue light and relevant fragments excised using a scalpel. DNA extraction from the agarose gel was carried out using the QIAquick

Gel Extraction Kit (Qiagen) according to manufacturer's instructions detailed below. All centrifugation steps were carried out at 17,000 x g.

1. Add 3 volumes Buffer QG to 1 volume of gel and incubate at 50°C until completely dissolved. Add 1 volume of isopropanol to the sample and mix.

2. Apply the sample to a QIAquick column and centrifuge for 60 secs. Discard flow-through. Add 500 mL Buffer QG to the column and centrifuge for 60 secs. Discard flow-through.

3. To wash, add 750 µL Buffer PE to the column and centrifuge for 60 secs. Discard the flow-through and repeat centrifugation of column.

4. Place the column into a 1.5 mL microcentrifuge tube, add 12 µL of dH_2O to the membrane, and centrifuge for 60 secs.

2.4.6 Plasmid Extraction

Overnight cultures of *E. coli* were prepared in 5 mL of LB broth. Aliquots of 2 mL were centrifuged at 6,800 x g for 3 mins. Plasmids were extracted using the QIAprep® Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions, detailed below. All centrifugation steps were carried out at 17,000 x g.

1. Resuspend pelleted bacterial cells in 250 µL Buffer P1. Add 250 µL Buffer P2 and mix thoroughly by inversion until the solution becomes clear. Add 350 µL Buffer N3 and mix by inverting 4-6 times.

2. Centrifuge for 10 mins. Apply supernatant to the QIAprep spin column, centrifuge for 60 secs and discard flow-through.

3. To wash, add 500 µL Buffer PB to the column and centrifuge for 60 secs. Discard the flow-through. Add 750 µL Buffer PE to the column and centrifuge for 60 secs. Discard the flow-through and repeat centrifugation of column.

4. Place the column into a 1.5 mL microcentrifuge tube, add 40 µL of dH_2O to the membrane, and centrifuge for 60 secs.

2.4.7 Restriction Digest

DNA was cleaved at chosen sites using the relevant restriction endonuclease (NEB) according to manufacturer's instructions. All enzymes were stored at -20°C before and after use. Typically, reactions contained 1µg DNA, 1 µL restriction enzyme(s), 2 µL buffer, with dH2O to a final volume of 20 µL. Reactions were incubated for 2 hours minimum at 37 °C, before visualisation on an agarose gel.

For vector digestion dephosphorylation was carried out to prevent re-ligation. After incubation, 1 µL of Antarctic phosphatase enzyme (NEB) and 2.3 µL of Antarctic phosphatase buffer (NEB) was added to the 20 µL reaction. The reaction was then incubated for a further 45 minutes at 37 °C, before visualisation on an agarose gel.

2.4.8 Ligation

Ligation of digested fragments and vectors with compatible ends was carried out using T4 DNA ligase (NEB) and T4 ligase buffer (NEB) according to manufacturer's instructions. Reaction ratios of 3:1 were used and volumes were calculated using the NEB Ligation Calculator (https://nebiocalculator.neb.com/#!/ligation) with 100 ng of vector. Along with vector and insert, reactions contained 1 µL T4 DNA Ligase, 2 µL T4 ligase buffer and dH2O to a final volume of 20 µL. Reactions were incubated overnight on ice water.

2.4.9 Hifi Assembly

For vector assembly

Primers for HiFi assembly were designed with the NEBuilder online tool (https://nebuilder.neb.com/) to generate fragments with a 20 bp 5' overlap. Vector backbones were digested with appropriate restriction enzymes and HiFi mixtures were prepared according to the manufacturer's instructions (Table 2.5).

Table 2.5 Hifi Reactions

	Recommended Volumes	
	2-3 Fragment Assembly	**4-6 Fragment Assembly**
Recommended DNA Molar Ratio	vector:insert = 1:2	vector:insert = 1:1
Total Amount of Fragments	0.03–0.2 pmols (X µL)	0.2–0.5 pmols (X µL)
NEBuilder Hifi DNA Assembly Master Mix	10 µL	10 µL
dH_2O	10 – X µL	10 – X µL
Total volume	20 µL	20 µL

Reaction mixtures were incubated at 50°C for 60 mins before transformation into *E. coli* or stored at -20°C.

Bridging double-stranded DNA with a single-stranded DNA oligo

The modified NEB Hifi assembly protocol was used for insertion of small fragments into a linearised vector using a single-stranded DNA oligo, according to the manufacturer's instructions (Table 2.6). The single-stranded oligo was designed with 25-30 nucleotide overlaps with the vector on each end, and diluted to a final concentration of 0.2 µM.

Table 2.6 Hifi assembly with single-stranded oligo

Component	20 µL Reaction	Final Conc. Or Amount
ssDNA Oligo (0.2 pmol/µL)	5 µL	1 pmol
Vector (0.005 pmol/µL)*	1 µL	0.005 pmol
Nuclease-free Water	4 µL	
NEBuilder HiFi DNA Assembly Master Mix 2X	10 µL	1X

Reaction mixtures were incubated at 50°C for 60 mins before transformation into *E. coli* or stored at -20°C.

2.4.10 Geneblocks

Where double-stranded gene fragments were externally synthesised, gBlocks™ were ordered from Integrated DNA Technologies (IDT), Leuven, Belgium, and used according to manufacturer's instructions.

2.4.11 Assembly of Mutagenesis Vectors

ClosTron

ClosTron vectors for gene knockouts were designed for use according to Heap et al., (2007) and Kuehne & Minton (2012). Introns for ClosTron mutagenesis were identified using the tool at www.ClosTron.com, which uses the Perutka algorithm. Introns were selected based on a central location within the gene of interest, and a high score identified by the tool. Final plasmid construction was then achieved by insertion of the selected intron sequences into pMTL007c-E5, by ATUM, California, USA.

RiboCas

RiboCas vectors for gene knockouts or insertions were designed for use according to Cañadas et al. (2019), requiring insertion of an editing template and sgRNA into the RiboCas vector backbone (pRECas1_MCS). Final vectors were either assembled using a three-way ligation of editing template (generated by SOEing PCR of two homology arms) and sgRNA into a digested backbone, or by hifi assembly of two homology arms and sgRNA into a digested backbone.

For gene deletions the editing template was designed to generate a deletion of the whole gene excluding the first two and final three codons. To allow CRISPR-mediated complementation a bookmark sequence was inserted by inclusion on the oligonucleotide.

To identify the 20 nucleotide SEED sequence of the sgRNA, Benchling's CRISPR tool was used (Benchling.com). This identifies 20 nucleotide sequences immediately 5' to a 5'-NGG-3' PAM (protospacer associated motif), either on the positive or negative strand, and the top-scoring sequences were chosen. The sgRNAs were generated using a primer-dimer PCR reaction (2.4.2).

2.4.12 Transformation of Chemically Competent E. coli

Chemically competent *E. coli* cells were thawed on ice before the addition of 1 – 10 ng of plasmid DNA. Cells were incubated on ice for 30 mins, heat shocked at 42 °C for 30 secs, and incubated on ice for 5 minutes. Cells were recovered in 450 mL S.O.C. (Super Optimal broth with Catabolite repression) medium (Thermo Fisher) and incubated at 37 °C with agitation for at least 1 h, prior to plating on warmed LB agar plates supplemented with the appropriate antibiotic.

2.4.13 Plasmid loss

The strain harbouring the plasmid of interest was plated onto appropriate media without selection. Colonies were used to inoculate 1 mL of pre-reduced liquid media without selection and incubated overnight. The culture was serially diluted in the same liquid media and three dilutions chosen to spread 100 µL onto three plates without selection. Following incubation to produce single colonies, approximately 20 were

individually picked using a 1 µL loop and patch-plated, first onto media without selection and then onto media with selection. Growth of each colony was compared with and without antibiotic selection for identification of loss of antibiotic resistance.

2.4.14 Plasmid Stability Assay

The strain harbouring the plasmid of interest was cultured overnight with antibiotic selection in appropriate liquid media, and then sub-cultured at 1% (v/v) into media without selection. New subcultures were made every 12 hours into pre-reduced liquid media without selection. Each subculture was serially diluted in media and plated on plates with and without thiamphenicol selection. The proportion of plasmid retained (R) was calculated by the CFU/mL with selection/ CFU/mL without selection, and the plasmid stability per generation (P) calculated as $\sqrt[n]{R}$, where n is the number of generations. R was then calculated as function of time (t), where $R=P^{(0.554t)}$ and plotted.

2.4.15 Extraction of Chromosomal DNA

To prepare genomic DNA samples the phenol-chloroform extraction method was used. The clostridial strain of interest was inoculated into a 10 mL overnight culture dilution series. A 2 mL sample was taken from most dilute culture with growth, and centrifuged (16,000 x g for 2 minutes). The pellet was resuspended in 180 µL of 40 mg/mL lysozyme solution (lysozyme from chicken egg white (Sigma-Aldrich) in PBS) and incubated at 37°C for 90 minutes with occasional agitation. 20 µL of RNase A solution (Promega) was added and incubated at room temperature for 30 minutes. Following this 25 µL of proteinase K solution (Merck), 85 µL of ddH_2O and 110 µL of 10% (w/v) SDS solution (Promega) were added and incubated at 65°C for 30 minutes with occasional agitation. The sample was then transferred to a phase-lock tube (5Prime) and an equal volume added of phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v) saturated with 10mM Tris (Thermo Scientific). After mixing by inversion it was then centrifuged (16,000 x g for 5 minutes), the top phase transferred to a new phase lock tube and the wash repeated twice more. Following the third and final wash the top phase was transferred to an ice cold solution of sodium acetate (3 M, Sigma-Aldrich):ethanol (100%, -80°C) at a ratio of 0.1:2 (v/v). This was incubated at -80°C for 60 minutes to precipitate DNA, then centrifuged at 4°C (16,000 x g for 15 minutes). The supernatant was then removed, the pellet washed in 70% ethanol and

centrifuged again at 4°C (16,000 x g for 3 minutes). The supernatant was removed, centrifugation repeated and final supernatant removed before the pellet was air-dried at room temperature for 45 minutes. The pellet was then resuspended in 30 µL of 100mM Tris buffer (pH 7.8).

2.4.16 DNA Quantification

DNA concentration and A260/280 ratio was measured at 260 nm using a NanoDrop Lite Spectrophotometer (Thermo Scientific). Measurements were taken against a blank of elution buffer, commonly dH_2O.

The concentration of genomic DNA preparations intended for sequencing was measured using the Qubit 4 Fluorometer (Invitrogen) according to the manufacturer's instructions.

2.4.17 Extraction of RNA

A 10 mL overnight dilution series of the *C. butyricum* strain(s) of interest was used to inoculate RCM liquid broth. After 3 hours, a 2 mL sample was taken from the culture, added to 4 mL of RNA protect (Qiagen), and incubated at 5 mins at room temperature. The sample was then centrifuged (5,000 x g for 10 minutes at 4°C), the supernatant discarded and pellet stored at -80°C until further processing.

RNA extraction

Using the FastRNA PRO Blue kit (Fisher Scientific), following thawing on ice pellets were resuspended in 1 mL RNA Pro solution, added to RNA pro matrix tubes and homogenised (6,400 rpm for 45 secs) in the Precellys 24 homogeniser (Bertin Instruments). Samples were then centrifuged (21,000 x g for 15 mins at 4°C) and the supernatant removed to a clean microcentrifuge tube to incubate at room temperature for 5 mins. 300 µL of ice-cold chloroform was added to each sample, the sample vortexed for 6 secs and then incubated at room temperature for 5 mins. Samples were centrifuged (21,000 x g for 15 mins at 4°C) and the upper phase transferred to a clean microcentrifuge tube containing 500 µL of ethanol. The samples were mixed by inversion and incubated at 80°C for 1-2 hours to precipitate the RNA. Samples were then centrifuged (21,000 x g for 15 mins at 4°C) and the pellet washed in 500 µL of

ethanol, before centrifugation (21,000 x g for 15 mins at 4°C). The pellet was then air dried at room temperature for 30 mins before resuspension in 50 µl of DEPC water and storage at -80°C.

RNA Clean-Up

Using the Ambion TURBO DNA-free Treatment (Thermo Fisher), 5 µL of Turbo DNase buffer and 2 µL of Turbo DNase was added to each sample and mixed. Samples were incubated at 37°C for 30 mins with occasional agitation before processing with the RNeasy Mini Kit (Qiagen). To the sample 350 µL of the RLT 2-mercaptoethanol mixture and 200 µl of ethanol was added, applied to an RNeasy column and centrifuged (13,000 x g for 15 secs), before the flow-through was discarded. 500 µL of RPE wash buffer was applied to the column, centrifuged (13,000 x g for 15 secs) and the flow-through discarded. The wash step was then repeated. The column was then dried by centrifugation (13,000 x g for 2 mins) before insertion into an RNase free tubes. To elute RNA 17 µL of DEPC water was applied to the column, left to stand for 3 mins, then centrifuged (13,000 x g for 1 mins). This step was repeated two further times to give a final elution volume of approximately 50 µL before storage at -80°C.

2.4.18 Generation of cDNA

RNA samples were thawed on ice. The Omniscript RT Kit (Qiagen) was used and reaction mixtures prepared according to the manufacturer's instructions (Table 2.7). Reaction volumes of 20 µL were used, with 1 µg of RNA added to each reaction.

Table 2.7 Omniscript reactions

	Volume (µL)
10x buffer RT	2
dNTPs	2
R hexamers	2
RNase inhibitor (diluted)	1
Omniscript	1
H_2O	Up to 20 µL
RNA	1 µg

The reactions were incubated at 37°C for 30 mins before storage at -20°C.

2.5 Sequencing

2.5.1 Sanger Sequencing

Plasmid DNA and PCR products were sequenced using Sanger sequencing performed by Eurofins Genomics, Ebersberg, Germany, or Genewiz, Azenta Life Sciences, Essex, UK.

2.5.2 Illumina Sequencing

Whole genome sequencing was conducted by DeepSeq (University of Nottingham), using Illumina MiSeq.

2.5.3 Pacific Biosciences Single Molecule Real-Time Sequencing

Prepared genomic DNA samples were sent to the University of Liverpool for Pacific Biosciences (PacBio) Single Molecule Real-Time (SMRT) sequencing. Unfortunately, data was unable to be produced due to sample quality.

2.5.4 Oxford Nanopore Sequencing

Oxford nanopore sequencing was carried out by Ruth Cornock (University of Nottingham, UK) using the FLO-MIN106 flow cell and Rapid Barcoding Kit (both by Oxford Nanopore Technologies). Sequences were assembled by Epi2me software (https://epi2me.nanoporetech.com/) (Oxford Nanopore Technologies).

2.6 Bioinformatics and Data Analysis

2.6.1 Data Visualisation and Statistical Analysis

All data was prepared using GraphPad Prism 10. Statistical analyses were carried out with: a Student's t-test; a one-way ANOVA with Dunnett's or Tukey's multiple comparisons tests; or a two-way ANOVA with Sidak's multiple comparison test. Calculated P-values <0.05 were deemed statistically significant.

2.6.2 Plasmid Maps

Plasmid maps were designed and edited using SnapGene Viewer (Dotmatics) and/or the Benchling online resource (www.benchling.com).

2.6.3 Sequencing Data Analysis

Sequencing data was analysed using the Benchling online resource (www.benchling.com).

2.6.4 Sequence Alignments

For alignments of one or more sequences to a reference sequence for SNP identification, the Benchling alignment tool was used (www.benchling.com) with standard parameters.

For multiple sequence alignment for sequence comparison (stated in results), the ClustalW tool was used (https://www.genome.jp/tools-bin/clustalw) with standard parameters.

2.6.5 Whole Genome Sequencing Analysis

CLC Genomics Workbench v22 (Qiagen) was used for analysis of Illumina reads. Paired end reads were mapped to the appropriate reference genome, stated alongside results, using the resequencing analysis map reads to reference tool with the following settings: mismatch cost = 2, insertion cost = 3, deletion cost = 3, length fraction = 0.5, similarity fraction = 0.8. Genomic variants were identified using the resequencing analysis basic variant detection tool, using quality parameters: neighbourhood radius = 5, maximum gap and mismatch count = 2, minimum neighbourhood quality = 15, minimum central quality = 20; and significance parameters: minimum coverage = 10, minimum variant frequency = 70, require presence in both forward and reverse reads = yes.

2.6.6 CRISPR prediction

The online CRISPR-Cas Finder tool (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index) was used to identify native CRISPR systems using a search for the genome sequence of interest.

The CRISPR Target online software and its associated databases (http://crispr.otago.ac.nz/CRISPRTarget/crispr_analysis.html) was used to search for the possible origins of the spacer sequences identified, in order to identify protospacer adjacent motif (PAM) sequences.

2.6.7 Python Script

For prediction of PAMs, the python prediction software developed by Poulalier-Delavelle, Baker, Millard, Winzer, & Minton (2023) was used with standard parameters. The Python script was sent upon request.

CRISPR system data generated by CRISPRFinder was inputted into the script, and used to set the array direction. The script submits the spacer list to BLASTn (National Library of Medicine, National Center for Biotechnology Information) (global mismatch <20%, no more than one mismatch in the first seven nucleotides). Hits are submitted to PHASTER (Arndt, Marcu, Liang, & Wishart, 2018) to update the phage attribute. If at least two of the hits were in phage regions, a consensus sequence was created for potential PAMs.

2.6.8 Promoter and Terminator Prediction

For prediction of bacterial promoters, the BPROM software (V. Solovyev, 2011) was used (http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb) to analyse chosen nucleotide sequences.

For prediction of bacterial terminators, the ARNold software was used (http://rssf.i2bc.paris-saclay.fr/toolbox/arnold/index.php#Results) to analyse chosen nucleotide sequences.

Chapter 3 Characterisation of the endogenous 7α-dehydroxylating species *Peptacetobacter hiranonis* DSM 13275 and the development of tools for knockout generation

3.1 Introduction

3.1.1 Peptacetobacter hiranonis *background*

Peptacetobacter hiranonis was first isolated from human faeces in 2001, with 16sRNA sequencing and DNA-DNA hybridisation with *P. bifermentans* and *C. sordellii* revealing the new species in clostridia cluster XI (Maki Kitahara, Takamine, Imamura, & Benno, 2001), followed by a confirmation of presence in the native flora of healthy humans (M. Kitahara, Sakamoto, & Benno, 2001). *P. hiranonis* was only recently reclassified into the novel genus *Peptacetobacter gen. nov*, from the genus *Clostridium* (X. J. Chen et al., 2020). Nearly half of the 16 genera in the family *Peptostreptococcaceae* came from misclassified taxa, and many members of the family have been reclassified to novel genera in recent years, including the former *Clostridium difficile*. The taxonomic position of *P. hiranonis* was brought into question following the characterisation of a novel isolate, ZHW00191. Based on phenotypic, chemotaxonomic and genetic analysis of this isolate it was proposed as the type strain of the type species *Peptacetobacter hominis*, forming a new genus. With *C. hiranonis* sharing a 95.3% similarity, its reclassification to the new genus was proposed (X. J. Chen et al., 2020).

P. hiranonis is a Gram-positive, rod-shaped obligate anaerobe that does not form spores. There is a limited body of work on this organism, but its focus has been its role in bile acid modifications in the gut. It is one of the limited number of gut microbes that can convert primary to secondary bile acids by 7α-dehydroxylation, and it is the only known bacterium capable of also deconjugating bile salts, due to the presence of bile salt hydrolases (Narushima et al., 2006). Along with *C. scindens, P. hiranonis* has the second highest activity of the 7α-dehydroxylating Clostridia (Doerner, Takamine, Lavoie, Mallonee, & Hylemon, 1997). The core oxidative *bai* genes of *P. hiranonis* are organised into a similarly structured operon (Figure 3.1) to that of the model 7α-dehydroxylating species *C. scindens* VPI 12708. The *baiB-baiH* genes form the main operon, downstream of the promoter region that is putatively regulated by upstream transcriptional regulators *barB* and *barA.* However, the *baiI* gene is not adjacently situated downstream of *baiH*, as in *C. scindens*, but rather resides downstream of *barB* in the reverse orientation. Further differences include the

presence of *baiJ*, similarly in a separate chromosomal region, only one copy of *baiA*, and absence of a *baiJKLM* operon (Ridlon et al., 2020).

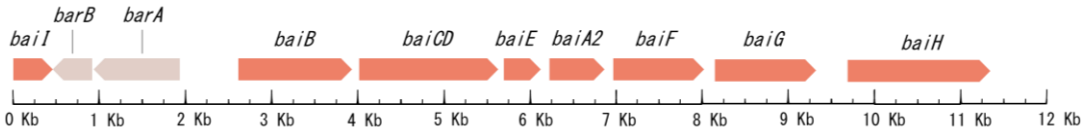


Figure 3.1 The *bai* operon of *P. hiranonis*. Bile acid inducible (*bai*) genes present in the *bai* operon, responsible for the 7α-dehydroxylation pathway and conversion of primary to secondary bile acids. Genes *baiB*-*baiH* under the control of a putative promoter, with *barB* and *barA* proposed as regulatory transcriptional elements. Additionally, *baiJ* (not shown) is present in a separate chromosomal region.

3.1.2 P. hiranonis *and* C. difficile

There is limited evidence for the direct involvement of *P. hiranonis* in colonisation resistance against *C. difficile* in the human gut; its individual species OTU was not identified as being highly correlated with susceptibility to *C. difficile* by Buffie *et al.* (2015), for example, but it has been implicated in resistance in a canine CDI model. Canines have been shown to be asymptomatic carriers of toxigenic *C. difficile* and a microbial protective mechanism has been suggested (Stone et al., 2016). *P. hiranonis* is the main 7α-dehydroxylating species in dogs (Giaretta et al., 2018), and Stone et al. (2019) have identified *P. hiranonis* as a relevant bacteria that may aid in CDI resistance in canines. Thanissery et al. (2020) have also demonstrated a negative association of commensal *P. hiranonis* and *C. difficile*.

P. hiranonis is a microbe of interest due to its similarity to *C. scindens*; *C. scindens* has been directly linked to CDI protection, and *P. hiranonis* has a similar 7α-dehydroxylation capacity to that of *C. scindens* (Doerner et al., 1997). However, there is limited evidence to directly implicate its production of bile acids to colonisation resistance. Moreover, there is also a growing body of evidence that questions the overall involvement of 7α-dehydroxylation in CDI resistance, and suggests other resistance mechanisms. Firstly, Kang et al. (2019) suggested that antibiotic production by *C. scindens* may be responsible. A secreted tryptophan-derived antibiotic, 1-acetyl-

β-carboline, was shown to inhibit the growth of *C. difficile*, and this was enhanced by secondary but not primary bile acids. Several other studies, discussed below, have also suggested that metabolic competition could play a role, particularly for metabolites involved in Stickland metabolism.

Stickland fermentation involves the use of amino acids as electron acceptors for energy generation; the coupling of proline, glycine and leucine fermentations, for example, allows regeneration of NAD^+ and formation of ATP (Girinathan et al., 2021; Ridlon et al., 2020). Many clostridia are capable of Stickland fermentation, particularly those in cluster XI, and it has been shown to play an important role in *C. difficile* colonisation. *C. difficile* is auxotrophic for proline, and two individual mutant strains lacking proline metabolism demonstrate decreased fitness or colonisation in a mouse model (Battaglioli et al., 2018; Reed et al., 2022). Stickland fermentation has also been shown in *C. scindens* (Girinathan et al., 2021), *P. hiranonis* and *C. hylemonae* (Ridlon et al., 2020), and their competition for Stickland metabolites has been proposed as a mechanism for colonisation resistance against *C. difficile*. Colonisation of mice with *C. scindens* or *P. hiranonis* demonstrated reduced proline concentrations due to its consumption, and protected against CDI in a bile acid independent manner (Aguirre et al., 2021). *In vitro* supernatant and coculture studies by Hromada et al. (2021) have also demonstrated bile acid independent inhibition of *C. difficile* by *P. hiranonis*, and the authors show that this is at least partially down to resource competition.

The largest hurdle to understanding the mechanisms of colonisation resistance against *C. difficile* by *P. hiranonis* and other 7α-dehydroxylating bacteria is the lack of genetic tools to develop loss-of-function mutations. To remove the ability to 7α-dehydroxylate BAs or utilise Stickland metabolites could enable the separation of these inhibitory effects, and subsequently work toward the improvement of tools to protect against CDI. Additionally, as discussed in 1.2.2, there are several *bai* genes whose function is yet to be determined. The generation of knockouts in the *bai* genes of *P. hiranonis* would help improve understanding of 7α-dehydroxylation in both this organism and others with similar *bai* operons.

3.1.3 Objectives

-Characterise and establish the conditions required for the genetic manipulation of *P. hiranonis,* including general characterisation and DNA transfer.

-Assess and optimise current clostridial tools for the generation of *bai* gene knockouts in *P. hiranonis*.

-Study 7α-dehydroxylation in knockout strains and assess their colonisation resistance mechanisms against *C. difficile in vitro*.

3.2 General characterisation and DNA transfer

P. hiranonis is a poorly studied bacterium and rarely appears in the literature. Because of this, characterisation of the species is limited and mainly identifies and focuses on its role in 7α-dehydroxylation. General characterisation was therefore carried out, both to expand knowledge of the species as a whole and in order to optimise DNA transfer in this study, particularly as there is no record of DNA transfer previously, nor in any other 7α-dehydroxylating species.

3.2.1 Growth Rate

The growth of *P. hiranonis* was measured over a period of 24 h (Figure 3.2), according to the protocol outlined in 2.3.5. Exponential growth occurred until approximately 9 h, reaching a maximum average OD_{600} of 2.75, with limited or no lag phase. The OD_{600} declined to 1.05 after 24 h.

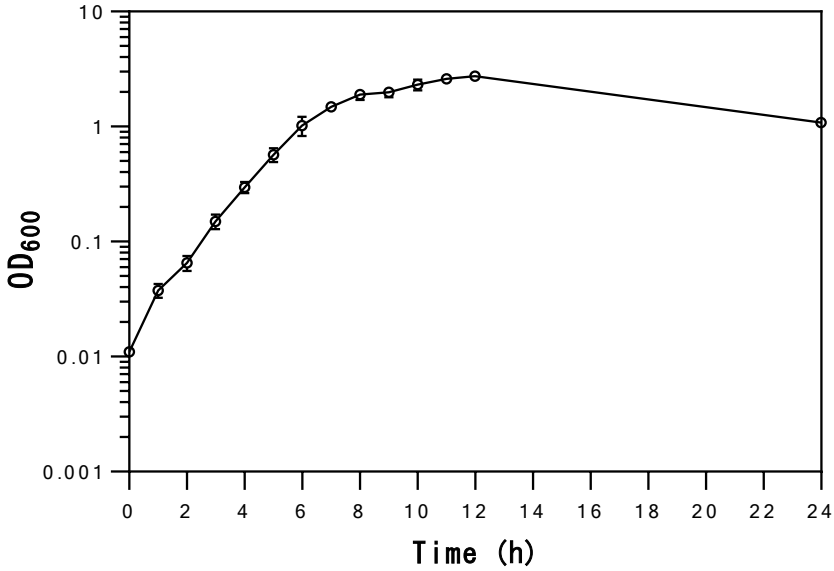


Figure 3.2 *P. hiranonis* growth curve. Growth curve of *P. hiranonis* WT measured by OD_{600}. Strain grown in liquid BHISS medium. Data represent mean values of three independent cultures ± SD.

3.2.2 Antibiotic Resistance

Antibiotic resistance assays were carried out for common antibiotics used in clostridia and those used as markers in the pMTL80000 vector series, to allow appropriate selection for the wildtype strain, and to assess the use of their resistance markers in future plasmids. Following preliminary determinations of susceptibility by plating onto selection, antibiotics were chosen for a broth dilution assay (2.3.6) to

determine the minimum inhibitory concentrations (MIC) in *P. hiranonis* (Figure 3.3).

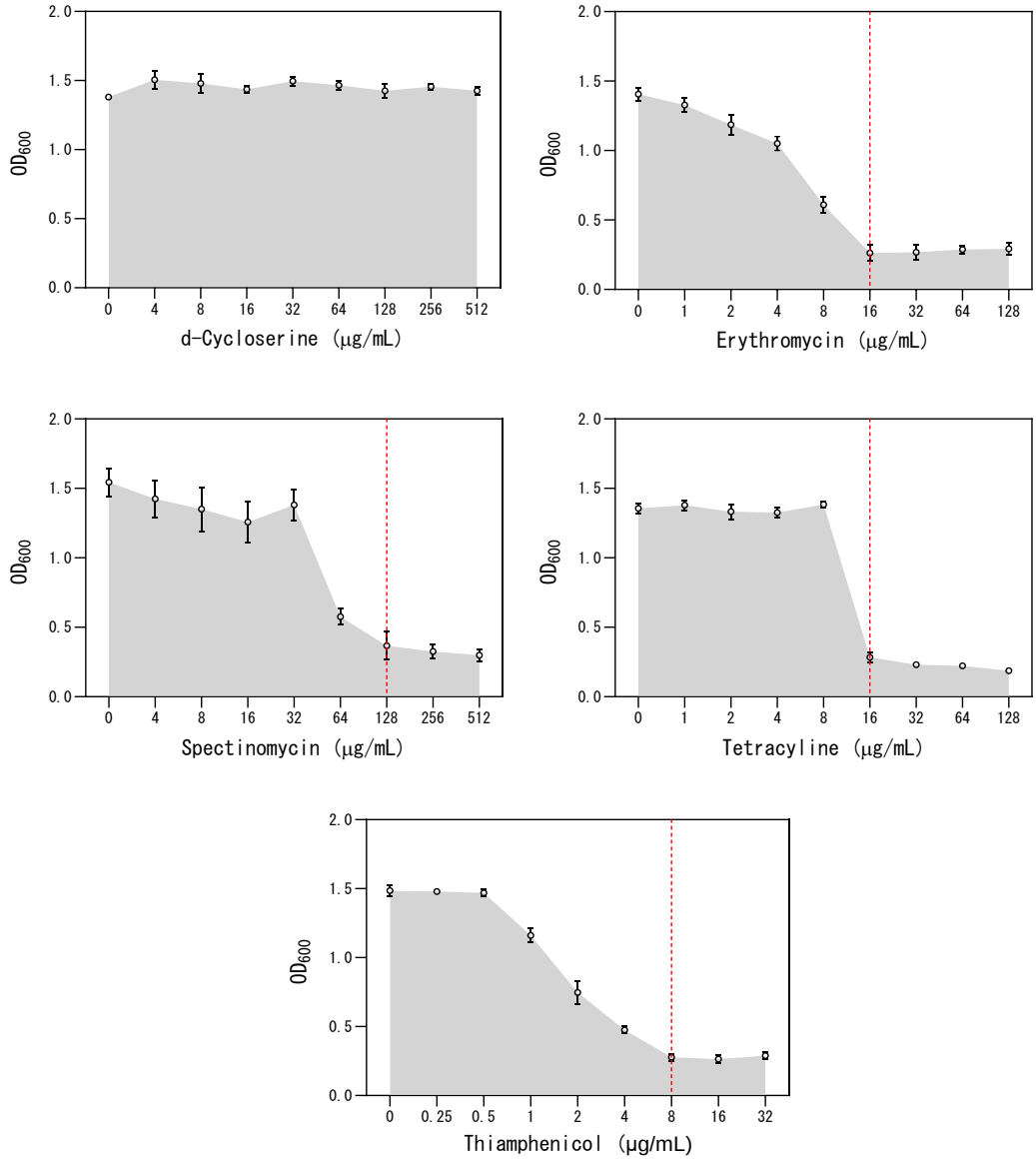


Figure 3.3 Minimum inhibitory concentrations of common antibiotics in *P. hiranonis*. *P. hiranonis* WT strain was cultured with a range of concentrations of chosen antibiotics. Starting from a 1 in 10 dilution of overnight culture, growth was measured after 24 h by OD_{600}. Data represent mean values of three independent cultures ± SD.

A dose response was not observed for d-cycloserine at the concentrations tested, illustrating resistance by *P. hiranonis*, in accordance with other clostridia, and will therefore allow d-cycloserine to be used for *E. coli* selection. A dose response was

observed for the other four antibiotics tested. The MIC, that is the lowest concentration at which no growth is observed, was estimated for: erythromycin (16 µg/mL), spectinomycin (128 µg/mL), tetracyline (16 µg/mL) and thiamphenicol (8 µg/mL). These MIC values will allow selection of *P. hiranonis* expressing respective resistance markers when antibiotics are used at standard clostridial working concentrations.

3.2.3 Restriction Modification Systems

Previous whole genome sequencing for *P. hiranonis* DSM 13275 was limited to Illumina Sequencing conducted by our group, and the genome sequence by Ridlon and Devendran., 2019 (reference sequence NZ_CP036523) compiled by Illumina MiSeq and Oxford Nanopore sequencing. Whilst these assemblies can provide sequence data for cloning work, Pacific Biosciences (PacBio) Single Molecule Real-Time (SMRT) sequencing can complement and improve the utility of Illumina data by read mapping and provision of base modification data. The base methylation data from PacBio sequencing can allow identification of restriction modification (RM) systems, and subsequently provide avenues to improve efficiency of DNA transfer.

A DNA sample of *P. hiranonis* WT was prepared for PacBio sequencing (2.5.3) using phenol-chloroform extraction (2.4.15), and quality assessed by measuring concentration, ratio of absorbance at 260/280 nm and running on an agarose gel to visualise DNA degradation or RNA contamination. Unfortunately, whilst the sample passed quality control checks by the PacBio establishments, sequencing data was unable to be produced.

Predicted RM systems

As PacBio data was unavailable to allow identification of RM systems by base modification data, *in silico* searches for restriction endonucleases and methylases were carried out. Relevant annotations in Illumina contigs identified the genes of interest listed in Table 3.1.

Table 3.1 Restriction endonucleases and methylases identified in the *P. hiranonis* genome. Enzymes identified by *in silico* searches of gene annotations in sequencing contigs. Genome sequence generated by Illumina sequencing (within research group).

BCI number	Gene	Description
00116	*bspRIM*	Modification methylase BspRI
00216	*mcrB*	5-methylcytosine-specific restriction enzyme B
00378	*dpnA_1*	Modification methylase DpnIIB
00379	*dpnM*	Modification methylase DpnIIA
00380	*dpnA_2*	Modification methylase DpnIIB
00381	*dpnB*	Type-2 restriction enzyme DpnII
00382	-	Type IIS restriction enzyme Eco57I
00424	-	Type IIS restriction enzyme Eco57I
00509	*hsdR*	Type I restriction enzyme EcoR124II R protein
00511	-	putative type I restriction enzymeP M protein
01178	-	endonuclease MutS2
01182	-	tRNA (cytidine(34)-2'-O)-methyltransferase
01183	*nth*	endonuclease III

BLASTP and domain searches (https://blast.ncbi.nlm.nih.gov/Blast.cgi) were carried out with the sequence of proposed restriction endonuclease or methylase genes to identify protein functions. This was also done for their up/downstream counterparts, to search for complete RM systems (Table 3.2).

Table 3.2 Conserved domains identified in proposed restriction modification system components of *P. hiranonis*. BLASTP results for proposed restriction endonuclease or methylase genes, identified by annotation, and their adjacent genes.

BCI	Name	Accession	Description	Interval	E-value
00115	HTH_XRE	Cd00093	Helix-turn-helix XRE-family like proteins	14-71	5.36e-05
00116	Dcm	COG0270	Site-specific DNA cytosine methylase	1-438	6.17e-69
00117	secG	PRK06870	Preprotein translocase subunit SecG	1-74	4.81e-10
00216	PRK11331	PRK11331	5-methylcytosine-specific restriction enzyme subunit McrB; Provisional	291-563	4.97e-76
00217	PRK09736	PRK09736	McrC, 5-methylcytosin-specific restriction enzyme subunit, provisional	1-343	5.04e-139
00378	Trm11	COG1041	tRNA G10 N-methylase	50-198	5.60e-16
00379	Dam	TIGR00571	DNA adenine methylase	7-277	1.55e-87
00380	N6_N4_Mtase	Pfam01555	DNA methylase	38-257	1.63e-60
00381	DpnII	Pfam04556	DpnII restriction endonuclease	6-282	8.09e-146
00382	N6_Mtase	Pfam02384	N-6 DNA methylase	7-184	5.18e-04
00423	Patatin	Pfam01734	Patatin-like phospholipase	4-192	5.41e-20
00424	BREX_1_MTaseX	NF033452	BREX-1 system adenine-specific DNA methyltransferase PgIX	27-386	4.62e-21

00424	Met_A_Alw26	TIGR02987	Type II restriction m6 adenine DNA methyltransferase	20-305	6.13e-12
00425	MurNAc-LAA	Cd02696	N-acetylmuramoyl-L-alanine amidase	180-351	4.17e-68
00509	hsdR	TIGR00348	Type I site-specific deoxyribonuclease, HsdR family	32-770	0
00510	RMtype1_S	Cd17291	Type I restriction modification system specificity (S) subunit	19-187	1.6e-67
00510	HsdS	COG0732	Restriction endonuclease S subunit	16-387	1.43e-26
00511	HsdM	TIGR00497	Type I restriction-modification system, DNA methylase subunit	8-520	0
01777	DUF523	Pfam04463	Protein of unknown function	2-146	9.49e-60
01178	MutS2	COG1193	dsDNA-specific endonuclease	3-766	0
01179	M20_pepD	Cd03890	M20 Peptidase D	10-483	0
01182	TrmL	COG0219	tRNA(Leu) C34 or U34 (ribose-2'-O)-methylase TrmL	1-155	1.72e-76
01183	Nth	COG0177	Endonuclease III	3-208	8.69e-113
01184	QueG	COG1600	Epoxyqueuosine reductase QueG	19-316	1.29e-60

Using the BlastP and conserved domain searches, the following putative RM systems (Figure 3.4) are suggested based on the presence of all required gene components adjacent in the genome:

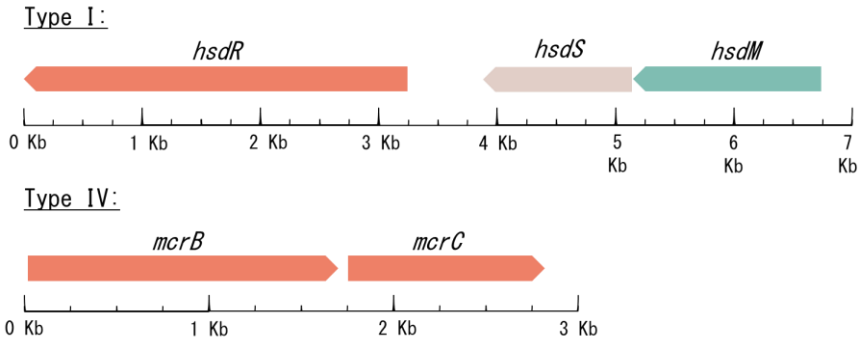


Figure 3.4 Putative restriction modification (RM) systems identified in *P hiranonis*. Two RM systems are proposed based on identification of all requisite genes by BLASTP searches: TypeI *hsdRSM* (BCI 00509, 00510, 00511) and Type IV *mcrBC* (BCI 00216 and 00217).

The two RM systems identified were: a type I system consisting of *hsdR*, *hsdS* and *hsdM*, and a type IV *mcrBC* system. All necessary components were identified adjacent on the operon, with confidence in identification given low E values. Other methylases were discounted due to the absence of adjacent restriction nucleases, exemplifying 'solitary' methylases, and/or higher E values.

The type IV *mcrBC* (methylcytosine restricting) system is a cytosine specific restriction system. The *mcrBC* recognition site consists of two RmC sites (R=A or G, mC= m4C or m5C) spaced between 40 and 3,000 base pairs apart, and DNA is cleaved between them (Bickle & Kruger, 1993; Stewart, Panne, Bickle, & Raleigh, 2000; Vasu & Nagaraja, 2013). This poses a barrier to DNA transformation, particularly from organisms that contain a *dcm* (DNA cytosine methyltransferase) system which methylates the second cytosine in the sequence CCWGG. To circumvent this, the sExpress *E. coli* conjugal donor (Woods et al., 2019) could be used for conjugations into *P. hiranonis*, as it lacks *dcm* methylation. With regards to the Type I RM system identified, more information is required to identify its subtype and recognition site; PacBio sequencing would allow this.

The gene in position 00381 was identified as encoding a DpnII restriction endonuclease, adjacent to 00379 and 00380 which were both identified as methylase domains, contrary to the expected one methylase characteristic of a Type II system. As 00379 is suggested to be a dam methylase, it may be only 00380 and 00381 that constitute the DpnII restriction system. However, other DpnII configurations have been observed, with the DpnII gene cassette of *Streptococcus pneumoniae* encoding two methyltransferases, DpnM and DpnA, alongside the endonuclease (Cerritelli, Springhorn, & Lacks, 1989). It is therefore possible that 00381 is associated with two methylases. Regardless of its exact structure, the presence of a DpnII system could present a barrier to cloning.

Whilst two RM systems at a minimum have been identified by *in silico* searches and provide insights for circumvention of barriers to genetic transfer, this manual method is unlikely to be exhaustive, with low confidence in the comprehensiveness of the results. For example, the lone entry for *P. hiranonis* on ReBase, the DGF055142 strain isolated from dog faeces, identifies a total of 8 RM systems, with 1 Type I and 5 Type II. The average nucleotide identity (ANI) between the two strains of *P. hiranonis* is calculated at 95.62% by the tool OrthoANI (https://www.ezbiocloud.net/tools/ani; Yoon, Ha, Lim, Kwon, & Chun, 2017).

The presence of at least two RM systems could limit conjugation efficiency into *P. hiranonis*, and this has therefore been tested in 3.2.4.

3.2.4 Conjugation Efficiency and Plasmid Stability

pMTL vector series

Transfer of DNA into clostridial hosts is achieved by either: electroporation or through conjugative mobilisation from an *E. coli* donor strain. For conjugation, the *E. coli* donor strain CA434 is one of a variety of strains that allows the *oriT*-mediated mobilisation of autonomous vectors, with transfer functions encoded by genes located on the resident R702 plasmid (N. P. Minton et al., 2016). To enable and maximise DNA transfer into *P. hiranonis*, a suitable Gram-positive replicon is required. The Gram-positive replicon allows replication of a plasmid in the Gram-positive acceptor strain, and the cloning vectors of the modular pMTL80000 series allow a series of replicons to be assessed (Heap, Pennington, Cartman, & Minton, 2009). Plasmids pMTL8X151 were studied in a conjugation efficiency assay; the Gram-positive replicon was varied (X) whilst the other modules of the plasmid remained constant (Table 3.3 and Table 3.4).

A 24-hour conjugation efficiency assay was carried out according to 2.3.3. The pMTL plasmids were conjugated into *P. hiranonis* using *E. coli* CA434 (Figure 3.5).

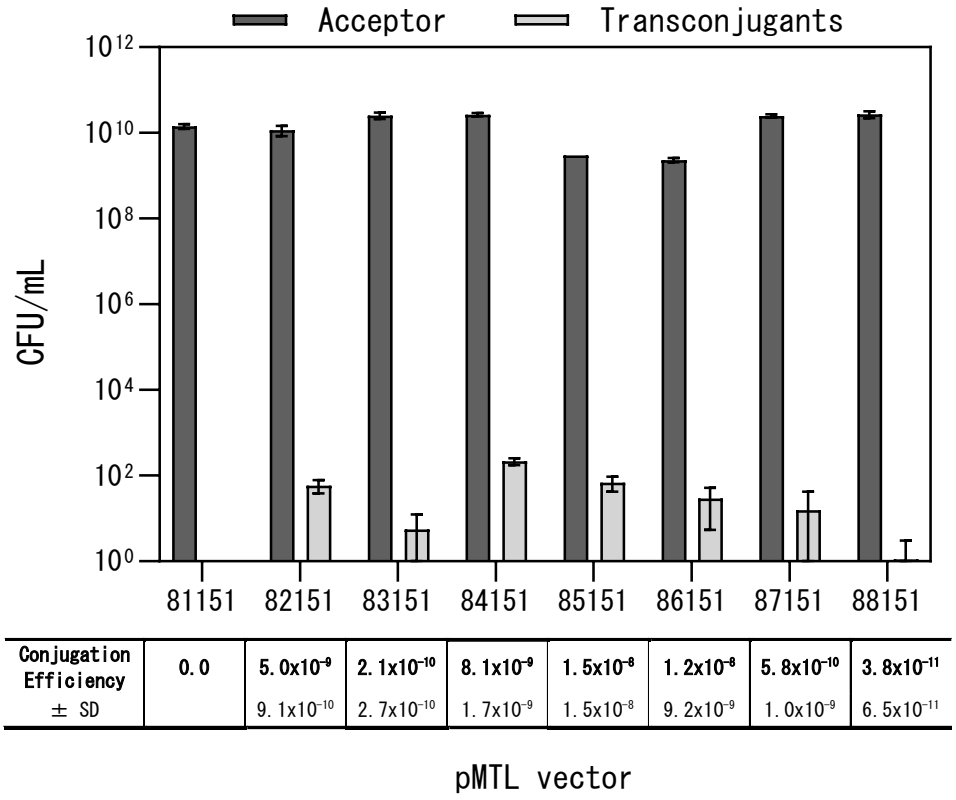
Table 3.3 Module elements of pMTL8X151 to enable a comparison of gram-positive replicons. The chosen fixed components for each module are outlined below.

Module	Chosen component	pMTL identifier
Gram-positive replicon	variable	X
Antibiotic resistance marker	catP	1
Gram-negative replicon	ColE1 + tra	5
Application specific module	Multiple cloning site	1

Table 3.4 Gram-positive replicons of the pMTL8000 series

Plasmid	Gram-positive replicon
pMTL81151	None (spacer)
pMTL82151	pBP1 (*Clostridium botulinum*)
pMTL83151	pCB102 (*Clostridium butyricum*)
pMTL84151	pCD6 (*Clostridioides difficile*)

pMTL85151	pIM13 (*Bacillus subtilis*)
pMTL86151	pIP404 (*Clostridium perfringens*)
pMTL87151	p19 (*Clostridium carboxidivorans*)
pMTL88151	pUB110 (*Staphylococcus aureus*)



Conjugation Efficiency ± SD	81151	82151	83151	84151	85151	86151	87151	88151
	0.0	5.0×10^{-9} ± 9.1×10^{-10}	2.1×10^{-10} ± 2.7×10^{-10}	8.1×10^{-9} ± 1.7×10^{-9}	1.5×10^{-8} ± 1.5×10^{-8}	1.2×10^{-8} ± 9.2×10^{-9}	5.8×10^{-10} ± 1.0×10^{-9}	3.8×10^{-11} ± 6.5×10^{-11}

Figure 3.5 Conjugation Efficiency of the pMTL vector series into *P. hiranonis*. A comparison of the different Gram-positive replicons (X) in the pMTL vector series, denoted by 8X151. Efficiency calculated as transconjugant colonies (CFU/mL)/ *P. hiranonis* conjugal acceptor colonies (CFU/mL). Data represent mean values of the conjugation in technical triplicate ± SD.

Conjugation into *P. hiranonis* was limited and efficiency was low throughout the different Gram-positive replicons. The control plasmid pMTL81151 containing a non-functioning replicon produced no transconjugants as expected. The replicons of pMTL82151, pMTL84151 and pMTL85151 gave the highest average efficiency of those tested and, whilst still low, demonstrate that conjugation into *P. hiranonis* can be achieved, enabling the possibilities of future genetic manipulation. The trend of low transfer efficiency is likely due to the presence of RM systems, putatively

identified in 3.2.3. Unfortunately, with no PacBio data and therefore no comprehensive list of RM systems and their recognition sites, a correlation between efficiency of transfer and number of recognition sites for each plasmid cannot be drawn.

The Gram-positive replicon of pMTL84151, taken from pCD6, produced the most consistent results. Moreover, the average number of transconjugant colonies produced was most conducive to subsequent mutagenesis steps (36 colonies/100µL of neat conjugal mixture), compared to that of 85151 (14 colonies/100µL of neat conjugal mixture) and 82151 (6 colonies/100µL of neat conjugal mixture). The pCD6 replicon was therefore deemed most suitable for use in vector design. Whilst the *E. coli* sExpress conjugal donor may allow increased conjugation efficiency due to avoidance of the putative type IV RM system identified in 3.2.3, attempts at conjugation for pMTL84151 were unsuccessful due to excess *E. coli* background growth, so this could not be verified.

These results suggest that RM systems are not an absolute barrier to DNA transfer in *P. hiranonis*; it is not genetic intractable like *C. scindens* (J. D. Kang et al., 2019a; Reed, Nethery, Stewart, Barrangou, & Theriot, 2020). This has been exemplified in other bacteria, for example Roer et al. (2015) demonstrated that the Type I EcoKI system in *E. coli* reduces uptake of DNA by conjugation but does not prevent it entirely.

A sufficient number of transconjugant colonies for pCD6 with *E. coli* CA434 suggests that this combination should be suitable for use in genetic manipulation strategies. However, if further optimisation is required the presence of a DpnII system could first be confirmed by incubation of *P. hiranonis* DNA with DpnII to visualise protection on a gel, and a *dam*+ (DNA adenine methylase) *E. coli* strain could circumvent it if present. Additionally, if genetic tools are able to, knockouts of key RM system genes could be generated; for example conjugation efficiency was increased by 8-and 10-fold by Lesiak et al. (2014) through deletions of *hsdR* components of two Type I systems in *Clostridium saccharobutylicum* using ClosTron. The same authors also demonstrated the use of *in vivo* methylation of donor DNA by cloning the methylase and specificity subunits into *E. coli*, to afford protection of the cloning vector for successful conjugation. Finally, another strategy is to synthesise plasmids without RM

recognition sites, as demonstrated by Johnston et al. (2019) for DNA transfer in *Staphylococcus aureus*, however PacBio base modification data would be required for this.

Plasmid Stability and Loss

In the natural environment, horizontal transfer of plasmids by conjugation allows the transfer of extra-chromosomal genetic elements, allowing adaptation and evolution of microbial populations under varying selection pressures (Millan et al., 2014; Wein, Hülter, Mizrahi, & Dagan, 2019). However, although some plasmids may provide beneficial advantages to the host, plasmid maintenance is an energy rich process and places a metabolic burden on the cell by exploiting host replication and gene expression machinery. Segregational instability also plays a role in plasmid maintenance, whereby plasmids are lost during cell division due to unequal segregation between daughter cells. This process is dependent on the copy-number of the plasmid, and therefore can be influenced by the replicon (Carroll & Wong, 2018; Millan et al., 2014; Wein et al., 2019). Taken together these two factors hinder the maintenance of the plasmid and, once selection is removed, plasmid free cells outcompete plasmid carrying cells (San Millan & Maclean, 2019; Wein et al., 2019). In the absence of selection, plasmid stability is dependent on the balance between its conjugation rate and these two factors (Lopatkin et al., 2017).

The natural processes of conjugation and plasmid loss are exploited for genetic engineering; once plasmid-based gene tools have achieved their desired function, usually encompassing insertion of genetic elements into the genome, the plasmid is eliminated and cured from the strain. This can be selected for using antibiotic resistance markers on the plasmid backbone. Therefore, whilst a high conjugation efficiency is desirable in a Gram-positive replicon, so too is the ability to lose the plasmid efficiently, i.e. within a minimal number of generations.

To ensure suitability of the pCD6 replicon for genetic engineering in *P. hiranonis*, a plasmid stability assay was carried out, as described in 2.4.14. The strain containing pMTL84151 underwent subcultures into media without selection and the proportion of plasmid retained calculated (Figure 3.6).

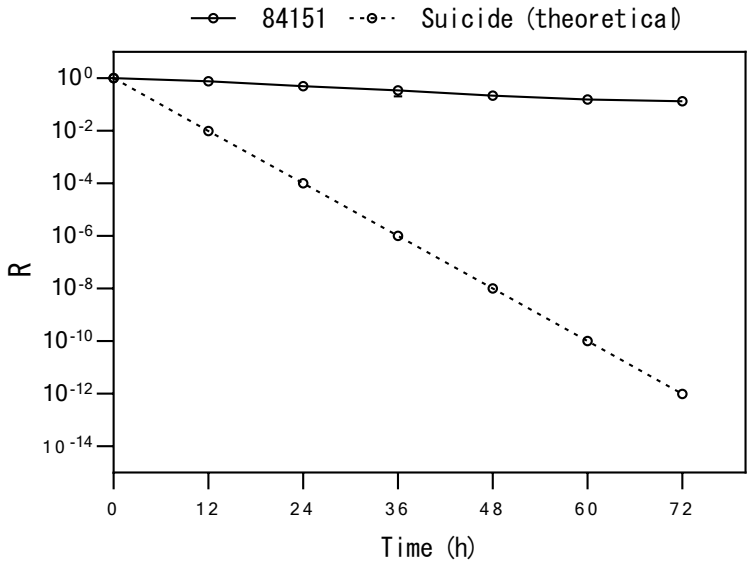


Figure 3.6 Plasmid stability of pMTL84151 in *P. hiranonis* without antibiotic selection. Loss of pMTL84151 over time given repeated sub-cultures in liquid BHISS media without selection every 12 hours. Samples were taken and plated to calculate the CFU/mL. R = plasmid loss as a function of time, measured by CFU/mL with and without thiamphenicol selection. Data represent mean values of three independent cultures ± SD.

Compared to the theoretical suicide plasmid, where the plasmid stability is 50% per generation and only 1% of cells retain the plasmid after 12 hours, pMTL84151 has a much slower rate of plasmid loss, at approximately 94%. However, at this rate 20% of cells lose the plasmid after a 12 hour culture without selection, which is adequate for the standard patch plating protocol for deliberate loss of an engineering plasmid (2.4.13). Without other Gram-positive replicons available for comparison it is difficult to hypothesise the influence of different variables on the stability of pMTL84151. This data does, however, demonstrate that the plasmid can be lost without additional elements on the plasmid backbone to aid with this.

Conjugation after plasmid loss

Whilst carriage of a plasmid can be costly to the host, in some cases this has been shown to alleviate over time. For example, in several Gram-negative strains lineages of transconjugants grew faster and/or had shorter lag times than their original counterparts, indicating fitness adaptation (Prensky, Gomez-Simmonds, Uhlemann, &

Lopatkin, 2021). Modelling by Prensky et al. (2021) also revealed that plasmid acquisition costs have an impact on conjugation dynamics, with the initial burden of the plasmid identified as one of the main constraints. These principles could explain observations by our research group, that conjugation efficiency can increase in a host that has previously maintained a plasmid (personal communications, Nigel Minton). This theory was investigated with regards to the improvement of conjugation efficiency in *P. hiranonis*, using a strain that had been cured of pMTL84151; plasmid curing was confirmed by plating onto thiamphenicol, and with a colony PCR for the plasmid (pCD6_F1 and catP_R1). Conjugation of pMTL84151 by *E. coli* CA434 into this strain was compared to the wildtype strain (Figure 3.7).

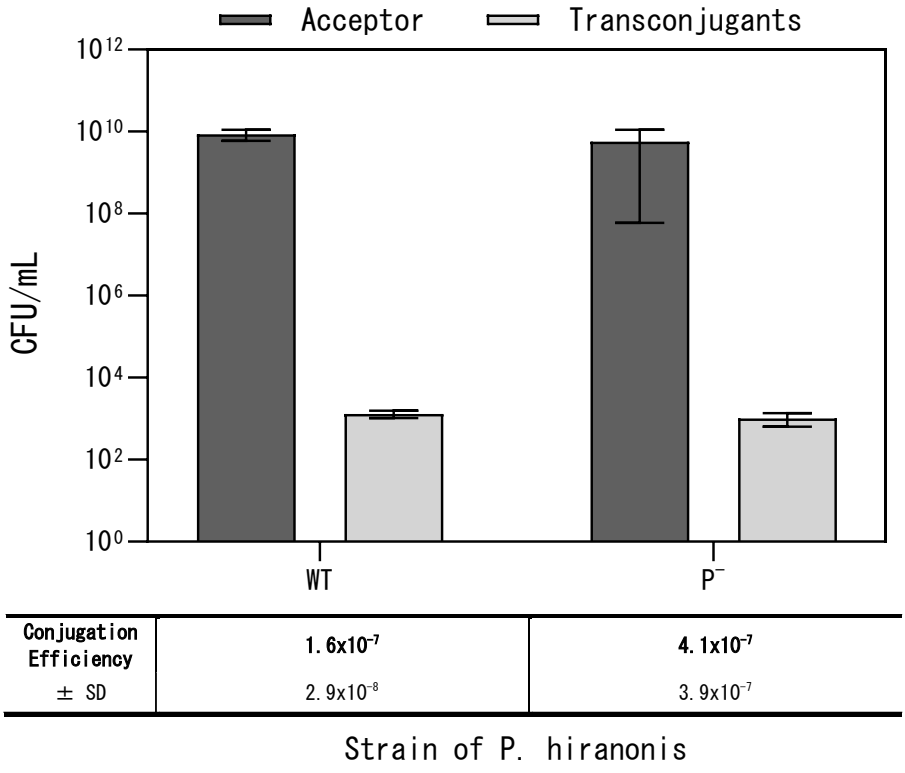


Figure 3.7 Conjugation efficiency of pMTL84151 into *P. hiranonis* previously cured of pMTL84151. WT = wildtype strain, P^- = WT previously conjugated and cured of pMTL84151. No significant difference in conjugation efficiency (unpaired two-tailed Student's t-test; p=0.3231). Data represent mean values of the conjugation in technical triplicate ± SD for the WT, and of three biological replicates ± SD for the P^-.

Unfortunately, a comparison of the conjugation efficiency for the two strains revealed no significant difference (unpaired two-tailed Student's t-test; p=0.3231). Prensky et al. (2021) have suggested that phenotypic adaptation and selection for plasmid acquisition are based on pre-existing heterogeneity, but occurrence of mutations that compensate for the fitness cost typically occur over hundreds of generations. Therefore, it is possible that one round of curing and subsequent growth for strain collection and usage there have not been enough generations to allow sufficient host adaptation, for either acquisition or fitness. With no improvement of efficiency in the cured strain, the *P. hiranonis* WT strain was therefore maintained as the conjugal acceptor.

3.3 Using existing Clostridial tools to generate gene knockouts

Currently there are no mutant strains of *C. scindens* because of its genetic intractability. It is not clear whether this has been a barrier for mutagenesis in the other 7α-dehydroxylating bacteria, or whether mutagenesis has not been attempted. If possible in *P. hiranonis*, knocking out genes of the *bai* operon and those involved in Stickland metabolism could allow the mechanism of colonisation resistance against *C. difficile* to be elucidated. It was first decided to attempt this using the genetic tools already established for clostridia.

3.3.1 ClosTron

The ClosTron backbone with pCD6, pMTL007c-E5, was chosen for conjugation into *P. hiranonis*. This backbone was first tested to ensure sufficient conjugation, and as there was no significant difference in efficiency to the pMTL84151 positive control (unpaired two-tailed Student's t-test; p≤0.1456), it was deemed suitable (Table 3.5).

Table 3.5 Conjugation Efficiency of pMTL007c-E5 in *P. hiranonis*. In comparison with the pMTL84151 positive control: p≤ 0.1456, unpaired, two-tailed Student's t-test.

Plasmid	Conjugation Efficiency	SD
pMTL84151	1.190×10^{-7}	5.574×10^{-8}
pMTL007c-E5	5.966×10^{-8}	1.520×10^{-8}

Design of ClosTron vectors

The following genes were chosen for ClosTron mutagenesis in *P. hiranonis*: the seven major *bai* genes; two genes encoding enzymes purportedly involved in Stickland metabolism (*grdA* and *prdC*) (Aguirre et al., 2021); and *hsdR* of the putative type I RM system. Clostron vectors based on the pMTL007c-E5 backbone were designed and generated according to 2.4.11, with chosen targets outlined in Table 3.6.

Table 3.6 Genes of interest for ClosTron mutagenesis in *P. hiranonis*. Chosen ClosTron targets for each gene outlined, including their score and sequence. Targets evaluated by the Perutka algorithm at www.ClosTron.com

Gene	Target	Target Score	Target Site
baiB	148S	8.84	ATATATATAGATAAAGAAGACAATGTGAGAGATATCACTTGGAAG
baiCD	670S	8.22	GAGTTAATGGATGAAGTAATGACAGAAGAAGAAATAGTAGAATTC
baiE	279A	4.25	TACCATCTTCCTGTAGCAGTATTTTCGCTGTCTATAGTTATTTCT
baiA2	357A	6.78	TGGTAAGCAGCCCAAGCACCATTGAATACACCTGTAACGTTTATA
baiF	317S	6.40	CAAAAGGTGGACAGTACGACAGACTAGGTCTTTCTGATGAAGTTA
baiG	616S	8.44	GATAACTTCGATAAAAAAGGTGCAGTATGCGTACTAATATTCTTC
baiH	112S	8.87	ATGAGTATGGACTACGAAGCTGCTGACGGAACAGTTCCAAAAAGA
grdA	388S	7.79	GAATTATTCGACGAAGAAGTTTACGAAGAACAGGTTGGAATGATG
hsdR	176A	7.64	TTTCAAGTTGAGCACGAAGGTTAGCTATTAAATCTTTTTCGTTAT
prdC	463S	7.70	ACTGTAATGGATATGGTTAATGCAGAAAAAGCTATAATAGGTATA

Utilisation of ClosTron vectors

The ClosTron vectors were transformed into *E. coli* CA434 and conjugated into *P. hiranonis*, selecting with thiamphenicol. Conjugation was successful for all 10 ClosTron vectors.

Following successful conjugation, mutagenesis was screened for by creating a 10-fold dilution series of transconjugant growth, and spreading the 10^{0}, 10^{-1} and 10^{-2} dilutions onto plates containing erythromycin. Unfortunately, despite multiple attempts, no transconjugants for any of the ClosTron plasmids demonstrated erythromycin resistance, and therefore it was not possible to screen for mutants by colony PCR.

The inability for transconjugants to grow on erythromycin is unlikely to be due to a non-functioning RAM, as this has been shown to work universally across Clostridia (Heap et al., 2007). It is most likely due to a lack of expression of the *ermB* gene, which is driven by P_{thl} from *C. acetobutylicum*. Ideally this would be assayed separately from the RAM but a plasmid containing the P_{thl} and *ermB* combination used in the RAM was not available, and nor was the sequence. If it was the promoter, this could have been assessed by conjugation of the appropriate plasmid into *P. hiranonis* and selection with erythromycin.

3.3.2 Endogenous CRISPR in P. hiranonis

As these results suggest ClosTron does not function in *P. hiranonis,* it was decided to explore other tools available for genetic manipulation. Since its discovery, development of CRISPR-Cas based genetic tools in bacteria has largely relied on the Type II CRISPR/Cas9 system from *Streptococcus pyogenes.* However, the harnessing of endogenous CRISPR systems offers an alternative method that can overcome some of the limitations of traditional *cas9* systems such as poor transfer rates and toxicity (Luo, Leenay, & Beisel, 2016).

Endogenous CRISPR systems are commonly found in *Clostridia*, present in 74% of strains (Pyne, Bruder, Moo-Young, Chung, & Chou, 2016). Type I-B systems have previously been identified and harnessed to make genetic modifications in several *Clostridium*, including *Clostridioides difficile* (Maikova et al., 2019), *Clostridium butyricum* (Zhou et al., 2021) and *Clostridium tyrobutyricum* (J. Zhang, Zong, Hong, Zhang, & Wang, 2018). In a similar manner to the Type II system, Type I-B systems allow selection of mutants by targeting of the Cas machinery to wild-type cells, causing their destruction. However, use of an endogenous Type I-B system avoids the need to encode the exogenous system on a resultingly large plasmid; the plasmid needs only to bear the editing template for homologous recombination and the artificial CRISPR miniarray (Maikova et al., 2019; Poulalier-Delavelle et al., 2023). Considering this, it was decided to explore the possibility of its use in *P. hiranonis* to generate gene deletions.

Identification of endogenous CRISPR systems

Characterisation of possible native CRISPR systems in *P. hiranonis* was carried out. The online CRISPR-Cas Finder tool (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index) was used to identify native CRISPR systems, and identified a type I-B CRISPR system, shown in Figure 3.8. Two CRISPR arrays were also identified, with Array 1 containing 10 spacer sequences with the Direct Repeat consensus ATTTACATTCCAATATGGTTCTAATCAAAT, and Array 2 containing 31 spacer sequences, with the Direct Repeat consensus ATTTAAATACATCAAATGTTATTAATAAGC .

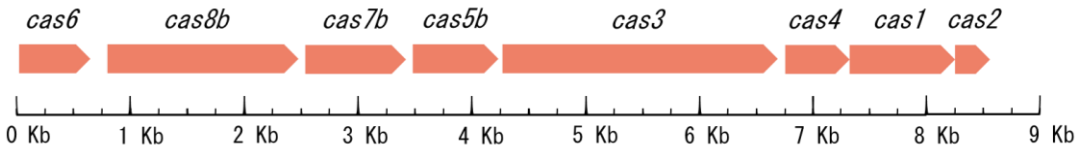


Figure 3.8 Type I-B CRISPR system in *P. hiranonis*. Genes identified by CRISPR-Cas Finder on analysis of *P. hiranonis* DSM 13275 genome published by Ridlon and Devendran., 2019 (reference sequence NZ_CP036523).

To enable the use of the native system the protospacer adjacent motif (PAM) sequence needs to be identified. CRISPR Target online software and its associated databases (http://crispr.otago.ac.nz/CRISPRTarget/crispr_analysis.html) was used to search for the possible origins of the spacer sequences identified. Unfortunately, a match was only identified for one of the sequences and therefore the PAM was unable to be identified by sequence alignments.

Identification of putative PAMs

For *in silico* prediction of the PAM sequences the automated Python script developed by Poulalier-Delavelle, Baker, Millard, Winzer, & Minton (2023) was used. Parameters of the script are provided in more detail in 2.6.7, but briefly the script takes each spacer identified by CRISPRFinder, searches BLAST databases for similarities in known bacteriophages and plasmids, and then applies a set of parameters to reduce false positives (Poulalier-Delavelle et al., 2023). Mismatches are allowed to reflect the

imperfect pairing afforded by PAMs to allow flexibility in recognition by the host, and the possible evolution of invading species (Semenova et al., 2011). Whilst the script was able to identify both the 5' and 3' adjacent sequences (3' not shown), the 5' sequence was chosen to assay, as most clostridial Type I-B CRISPR-Cas systems have been shown to recognise three nucleotides of the 5′ PAM (Pyne et al., 2016). Moreover, five nucleotides of the 5' sequence were chosen as it has been shown most PAMs are within this region (Shah, Erdmann, Mojica, & Garrett, 2013). The script returned three hits, shown in Table 3.7, but was not able to identify a consensus sequence for the 5' PAM.

Table 3.7 Identification of putative PAMs in *P. hiranonis*. PAM prediction by the automated python script by Poulalier-Delavelle et al. (2023), analysing spacer sequences identified by CRISPRFinder. Accession = GenBank accession number of the hit organism found with BLAST. Mismatch % = Mismatch percentage between query and hit. 5' = 5' sequence retrieved from the target organism.

Spacer	Array	Sequence	Organism	Accession	Hit sequence	Mismatch %	5'
11	1	CAACGTTAAACACTTCTTTGAGTTTTTCAATTACTT	Calothrix sp. NIES-4105 plasmid plasmid1 DNA	AP018291	CATCGTTAAAC-CTGCTTTTGATTTTTTCAATTACTT	11	**AATTT**
23	1	TAAACCAGAAACTGTAACTCTTAAAAATAATATAAT	Candidatus Arthromitus sp. SFB-rat-Yit DNA	AP012210	TAATCCAGAAATTTTAACCCTAAAAAATAATATAA	17	**ACTAC**
26	2	GAAAATAAATCAAGAAATAGAAGCATTAAAAGCTC	Enterococcus phage vB_OCPT_Ben, complete genome	MN027503	GAAACTAAA-CAAGAAACTAGAAGCATTAAAA	17	**GATAT**

PAM vector construction

To assay for a correct PAM sequence a plasmid interference assay can be used; the putative PAM can be placed upstream of the respective protospacer, resulting in significantly reduced or unachievable transfer into the host due to recognition by Cas machinery (Zhou et al., 2021). As no consensus PAM was identified, all three candidate PAMs and their relevant protospacer were inserted into pMTL84151, as well as respective control sequences (the upstream five nucleotide sequence of the

direct repeat: CAAAT for array 1; TAAAT for array 2). To generate these vectors, shown in Table 3.8, pMTL84151 was digested with XhoI and hifi assembly was used for insertion by single-stranded (SS) oligonucleotide (2.4.9).

Table 3.8 Vectors for a plasmid interference assay in *P. hiranonis* for PAM identification. Plasmids for protospacers 11, 23 and 26, containing either the putative PAM (protospacer_5) or control (protospacer_DR). 5 nucleotides from the 5' of the direct repeat was also included at the 3' of the protospacer. The three sequences were then included on the single stranded (SS) primer for hifi assembly into pMTL84151.

Plasmid	5' sequence	Protospacer	3' sequence	SS Primer
p84151_11-5	AATTT	CAACGTTAAACACTTCTTTGAG TTTTTCAATTACTT	ATTTA	84_proto11_5
p84151_11-DR	CAAAT	CAACGTTAAACACTTCTTTGAG TTTTTCAATTACTT	ATTTA	84_proto11_DR
p84151_23-5	ACTAC	TAAACCAGAAACTGTAACTCTT AAAAATAATATAAT	ATTTA	84_proto23_5
p84151_23-DR	CAAAT	TAAACCAGAAACTGTAACTCTT AAAAATAATATAAT	ATTTA	84_proto23_DR
p84151_26-5	GATAT	GAAAATAAATCAAGAAATAGA AGCATTAAAAGCTC	GCTTA	84_proto26_5
p84151_26-DR	TAAAT	GAAAATAAATCAAGAAATAGA AGCATTAAAAGCTC	GCTTA	84_proto26_DR

Plasmid Interference Assay

A plasmid interference assay (2.3.7) was carried out to test for a reduction in conjugation efficiency of the PAM plasmids compared to their controls (Figure 3.9).

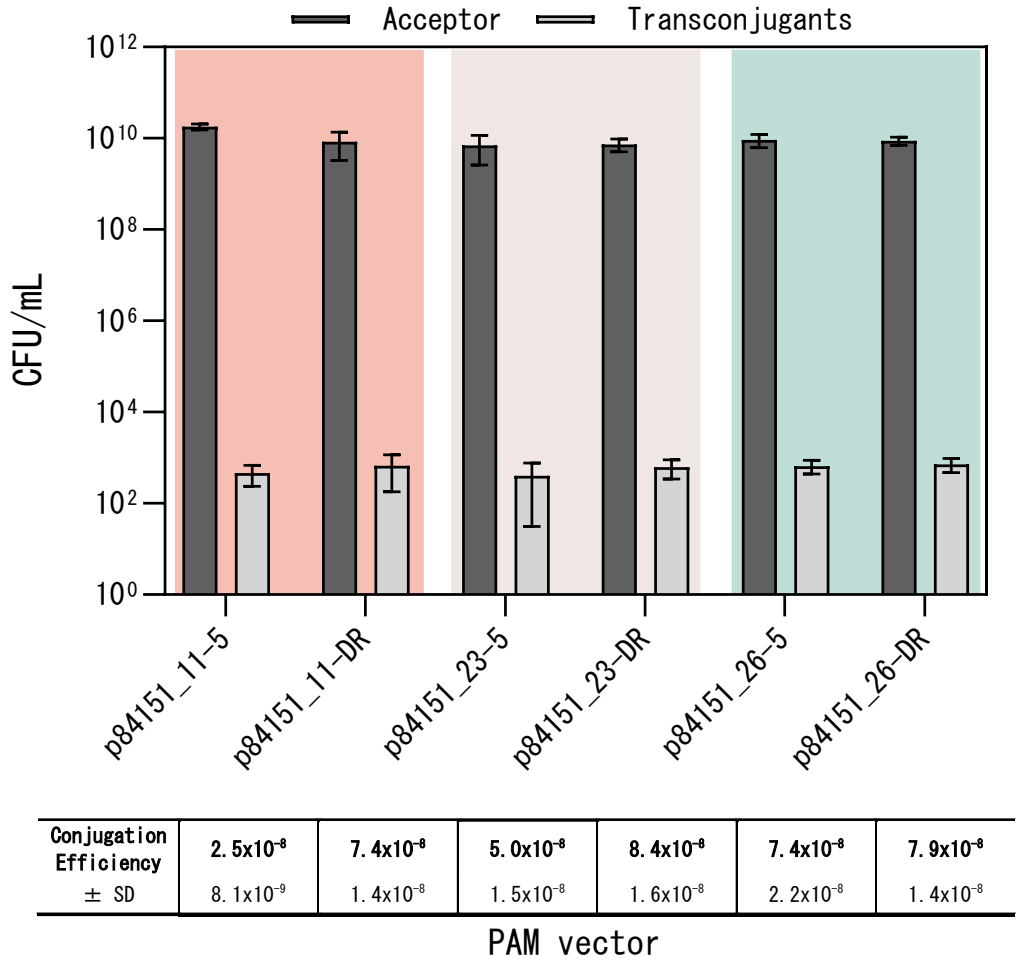


Figure 3.9 Plasmid interference assay for putative PAMs in *P. hiranonis*. Conjugation efficiency of plasmids containing putative PAMs and their protospacer to compare to their respective control plasmid**.** P84151_'Number' denotes the spacer number, '-5' denotes a five nucleotide PAM, '-DR' denotes a five nucleotide control sequence derived from the direct repeat. Statistical analysis was carried out to compare conjugation efficiencies for the PAM and its control (unpaired two-tailed Student's t-tests): 11 (p=0.0055), 23 (p=0.0558), 26 (p=0.7476), Data represent mean values of the conjugation in technical triplicate ± SD.

When comparing the conjugation efficiency for the PAM and direct repeat control for each protospacer, only protospacer 11 had a significant difference (unpaired two-tailed Student's t-test: protospacer 11 p=0.0055; protospacer 23 p=0.0558; protospacer 26 p=0.7476). There was an approximate 3-fold decrease in the conjugation efficiency of p84151_11-5 compared to p84151_11-DR. Whilst the reduction is statistically

significant, the conjugation efficiency for all plasmids is relatively low, so could be easily influenced by small changes in conjugation variables.

Published results for correctly identified PAMs from an interference assay demonstrate a much larger, or complete, reduction in transformation; three tested PAMs for a type 1-B system were completely unable to transform in comparison to their transformable controls in *Clostridium pasteurianum* (Pyne et al., 2016), and a correct PAM with three different protospacers had reductions of 10-fold, 100-fold and 100-fold for a type I-E system in *Lactobacillus crispatus* (Hidalgo-Cantabrana, Goh, Pan, Sanozky-Dawes, & Barrangou, 2019). Considering this, the transconjugant colonies for protospacer 11 should have been screened for escape mutants, for example mutated PAMs or spacer sequences, as these would have enabled avoidance of the CRISPR/Cas system and inaccurately increased conjugation efficiency, preventing elucidation of the true reduction in efficiency.

Whilst PAMs with a relatively weak interference score (reduction in conjugation efficiency of 51% and 62%) have been successfully used in knockout generation in *Clostridium butyricum*, extensive optimisation was required. This included the choice of promoters in the CRISPR expression cassette, and of protospacer length, which required balancing with conjugation and genome editing efficiency (Zhou et al., 2021).

The relatively small reduction in conjugation efficiency for protospacer 11 and its PAM compared to published results in other organisms suggests that it is either not a valid PAM, or that extensive optimisation is required for its use in genetic engineering. As conjugation efficiency is inherently low with pCD6, there is risk that any reduction in transfer on optimisation of the CRISPR system would result in loss of conjugation ability. The chances of success are therefore low considering both factors. Moreover, there is already an optimised, universal CRISPR-*cas9* tool available for testing (Cañadas et al., 2019), and it was decided to proceed with this system, and not endogenous CRISPR.

3.3.3 Use of RiboCas technology to generate bai gene knockouts

As discussed in 1.3.3, the RiboCas system generated by Cañadas et al. (2019) is a CRISPR/Cas9 based genome editing tool that employs a synthetic riboswitch for control of *cas9* expression to overcome the common obstacles experienced with traditional Type II editing systems. The RiboCas plasmid, pRECas1_MCS, allows insertion of elements into the application specific module, illustrated in Figure 3.10, for targeted chromosomal editing. This includes the sgRNA, containing the 20 nucleotide protospacer sequence for targeting of *cas9* to the DNA region of interest, and the homology arms which allow mutant generation by homologous recombination.

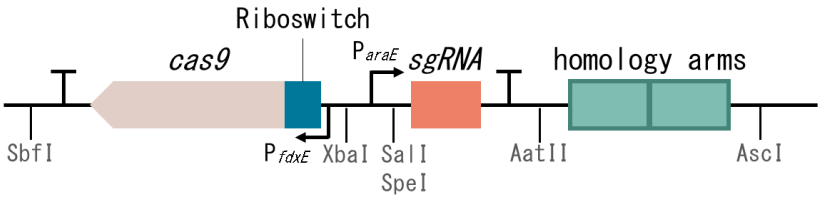


Figure 3.10 RiboCas application-specific module of pRECas1_MCS. The RiboCas plasmid, pRECas1_MCS, developed by Cañadas et al. (2019), consists of the pMTL83151 backbone with insertion of the CRISPR components into the application specific module between the restriction sites SbfI and AscI. The components specific for the desired mutation can be inserted between set restriction sites, SalI and AatII for the sgRNA, and AatII and AscI for the homology arms.

Design and Generation of RiboCas vectors

RiboCas vectors were designed to generate individual gene deletions of the seven major *bai* genes in *P. hiranonis*, according to 2.4.11. Firstly, to allow subsequent conjugation into *P. hiranonis*, the Gram-positive replicon of pRECas1_MCS was substituted from pCB102 to pCD6, by digestion with FseI and AscI and ligation of the replicon digested from pMTL84151. This generated pRECas_CD6_MCS to use as the backbone for insertion of the application specific module elements.

For each gene the editing template was generated by SOEing PCR (2.4.2), and included the BM8 bookmark (Table 3.9) to allow subsequent CRISPR-mediated complementation. For the sgRNA SEED the top-scoring sequences were chosen, according to Benchling's CRISPR tool (Benchling.com) (Table 3.10), and the sgRNAs generated using a primer-dimer PCR reaction (2.4.2). Following digestions of the pRECas_CD6_MCS backbone (SalI and AscI), sgRNA (SalI and AatII) and editing template (AatII and AscI), a three-way ligation was carried out for fragment insertion. This generated plasmids pRECas_CD6_X, where X denotes the letter of the targeted *bai* gene. It was attempted to generate three sgRNA variants per gene target to allow for non-functional sgRNAs. However, as difficulties were faced with cloning only one final plasmid was generated per gene, with primers used for the final plasmids outlined in Table 3.11. Final plasmids were confirmed by colony PCR with sgRNA_F4 and pCD6_R1, followed by Sanger sequencing.

Table 3.9 BM8 bookmark sequence. Elements of the BM8 bookmark (a 20 nucleotide protospacer adjacent to a 3 nucleotide PAM with additional nucleotide to avoid frameshift). Orientation is in relation to the coding sequence it replaces. (Seys, Rowe, Bolt, Humphreys, & Minton, 2020).

	Bookmark Sequence (24 nt)			
Extra nt	**Protospacer (20 nt)**	**PAM**	**Orientation**	**Origin**
G	TCCGGAGCTCCGATAAAAAA	TGG	+/-	*Bacillus subtilis*

Table 3.10 Guide sequence selection for *P. hiranonis bai* gene RiboCas vectors. SEED sequences identified for each target gene by the CRISPR tool at Benchling.com. SEED sequences outlined below for the final RiboCas plasmid successfully cloned for each target.

***bai* gene target**	**Guide SEED sequence**	**Position**	**Strand**	**PAM**	**On-target score**	**Off-target score**
B	CTGCTGTATGAGGAGCACCA	2439	-	TGG	73	100
CD	CATTGTATAGGATGTGACCA	4441	+	GGG	72.5	100
E	AAAAACGTTGGAATAAACGG	5778	+	TGG	71.4	99.9
A2	ATTGGAGTGTTAACAACACC	6607	-	TGG	56.3	100
F	AAAAAACCTGACAACCAGAG	7040	+	AGG	75.7	100
G	TTACTAGTAGGTCTTATAGG	8370	+	TGG	63.3	98.8
H	TTTATATGGCAACGACACTG	11164	-	AGG	72.4	100

Table 3.11 Primers used in the construction of RiboCas plasmids for *bai* gene knockouts.

bai gene target	LHA primers	RHA primers	sgRNA* forward primer	Final plasmid
B	baiB_LF1 BaiB_BM8_LR1	BaiB_BM8_RF1 baiB_RR1	sgBaiB_F1	pRECas_CD6_B
CD	baiCD_LF2 BaiCD_BM8_LR1	BaiCD_BM8_RF1 baiCD_RR2	sg_BaiCD_F1	pRECas_CD6_CD
E	baiE_LF1 BaiE_BM8_LR1	BaiE_BM8_RF1 baiE_RR1	sg_BaiE_F1	pRECas_CD6_E
A2	baiA2_LF3 BaiA2_BM8_LR1	BaiA2_BM8_RF1 baiA2_RR1	sg_BaiA2_F1	pRECas_CD6_A2
F	baiF_LF1 baiF_BM8_LR1	baiF_BM8_RF1 baiF_RR1	sg_BaiF_F1	pRECas_CD6_F
G	baiG_LF1 baiG_BM8_LR1	baiG_BM8_RF1 baiG_RR1	sg_BaiG_F1	pRECas_CD6_G
H	baiH_LF1 baiH_BM8_LR1	baiH_BM8_RF1 baiH_RR1	sg_BaiH_F1	pRECas_CD6_H

* All sgRNAs were created using sgRNA_Rev as the reverse primer.

Conjugation and Induction of RiboCas vectors

Following construction of the *bai* KO RiboCas vectors they were transformed into chemically competent *E. coli* CA434 and conjugated into *P. hiranonis*, allowing 24 h for conjugation. Transconjugants were selected for with thiamphenicol and re-streaked to confirm presence of the plasmid. Colony PCR was also carried out and confirmed presence of the plasmid (not shown). Successful conjugation was achieved for all seven of the RiboCas vectors.

Transconjugants for vectors pRECas_CD6_B, pRECas_CD6_E and pRECas_CD6_F were streaked onto BHISS agar plates containing thiamphenicol, to maintain plasmid selection, and 5mM theophylline, to induce the RiboCas CRISPR system. All re-streaks had growth, and colonies were selected from the re-streaks for colony PCR, to

confirm gene deletion. As Figure 3.11 demonstrates, deletion of *bai* genes did not occur, and a wildtype genotype was present. Re-streaks of the colonies on the theophylline plates were attempted but produced the same results.

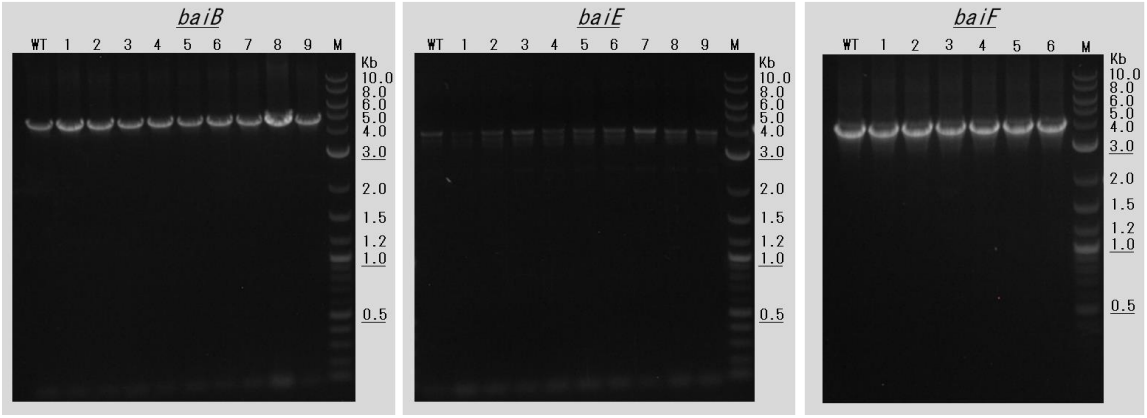


Figure 3.11 Colony PCR screening for *bai* gene deletions. Agarose gel electrophoresis visualisation of colony PCRs with flanking primers to assess for deletions by RiboCas in chosen *P. hiranonis* genes. WT = wildtype *P. hiranonis*, M = DNA Marker (2-log ladder; NEB). Numbered lanes = transconjugant colonies plated on thiamphenicol selection with theophylline inducer. Primers and expected fragment sizes (bp) *baiB*: CHIR_baiop_scr_F1 and Ch_scr_R2. WT = 4639, KO = 3099. *baiE*: Ch_scr_F4 and Ch_scr_R4. WT = 3793, KO = 3286. *baiF*: Ch_scr_F6 and Ch_scr_R5. WT = 3657, KO = 2386.

A functioning RiboCas system would not allow for wildtype growth on theophylline, as it would induce cleavage and be lethal. As wildtype growth on theophylline occurred for three different vectors it was unlikely due to non-functioning sgRNAs, and was likely due to insufficient expression of either *cas9* and/or the guide RNA. Whilst the RiboCas system has previously been tested and confirmed in different Clostridia (Cañadas et al., 2019), this was the first time it has been tried in *P. hiranonis*. These preliminary results demonstrate that further troubleshooting and optimisation is required for it to function.

3.3.4 Troubleshooting RiboCas

Further assays were carried out to confirm that the RiboCas system is non-functioning in *P. hiranonis*, and to attempt to identify the causal elements.

SEED sequence

Each of the attempted gene KOs in 3.3.3 tested one SEED sequence. However, sgRNA prediction can be limited and different SEEDs can have varying degrees of efficiency (Konstantakos, Nentidis, Krithara, & Paliouras, 2022). Although it is unlikely for all seven targets, to confirm that the failure of RiboCas was not purely down to inefficient SEED sequences two additional SEED sequences were chosen for one target, *baiCD*, to test a total of three SEEDs (Table 3.12). The same HAs were used as in pRECas_CD6_CD.

Table 3.12 Three guide SEED sequences present in *baiCD*. Three highly scored SEED sequences according to Benchling's CRISPR tool to provide sgRNA fragments to assess for targeting of the RiboCas system to generate a *baiCD* knockout. Inserted into RiboCas along with relevant homology arms to produce the final vectors of pRECas_CD6_CD/CD2/CD3

***bai* gene target**	**Guide SEED sequence**	**Position**	**Strand**	**PAM**	**On-target score**	**Off-target score**
CD	ATCCTATAGCTATAACAACG	4813	-	TGG	76.3	99.9
CD2	GGTGTTGGACGTATAAACAC	4297	+	AGG	69.2	100.0
CD3	GACGTAGCTGACTTATCAAG	4162	+	AGG	68.0	100.0

All three plasmids (pRECas_CD6_CD, pRECas_CD6_CD2 and pRECas_CD6_CD3) were able to conjugate into *P. hiranonis* and had the same 100% rate of re-streaking of transconjugants. Once again colonies from the theophylline selection plates were screened for a *baiCD* deletion by colony PCR (Ch_scr_F3 and Ch_scr_R2), and all but one colony were WT. One colony from pRECas_CD6_CD demonstrated a band of the size for the designed KO (lane 9, Figure 3.12), and this was confirmed by Sanger Sequencing. The plasmid was then cured by one round of patch plating (2.4.13) and KO genotype reconfirmed by colony PCR, to make the final strain.

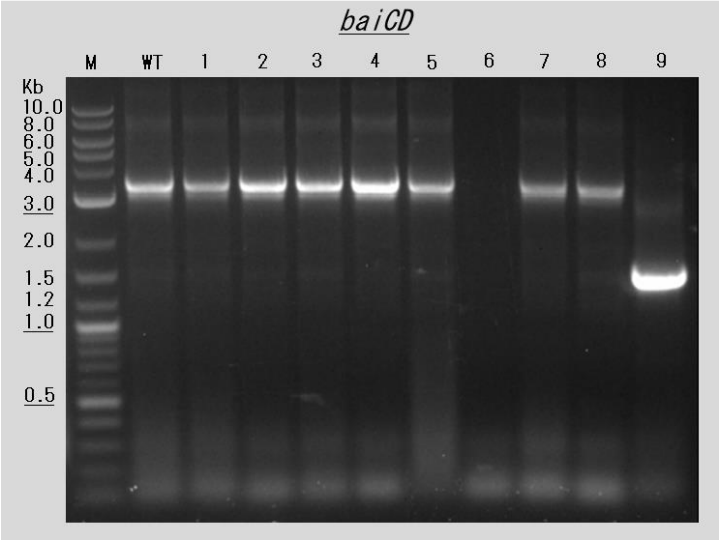


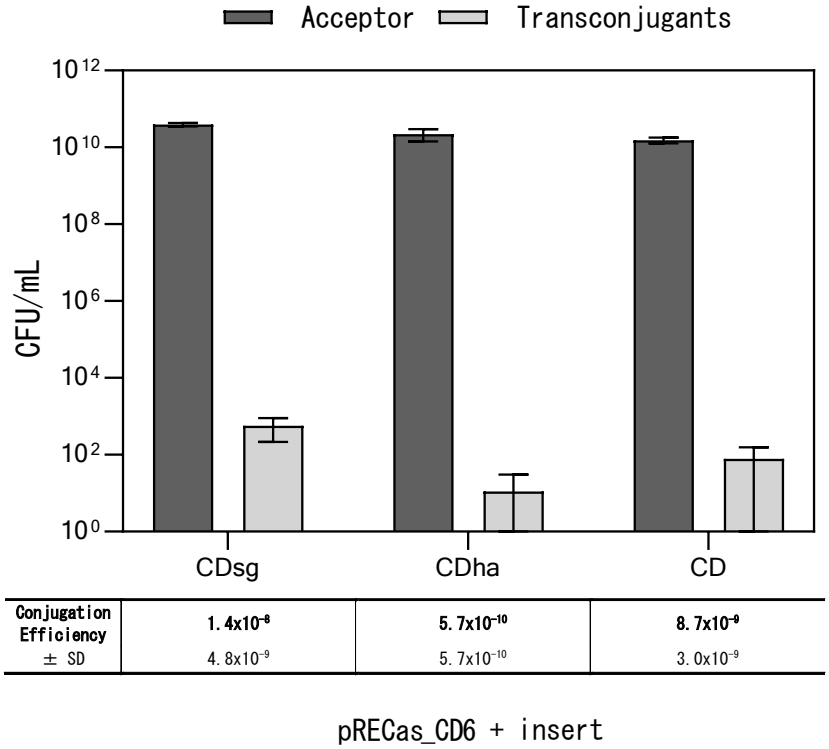
Figure 3.12 Colony PCR screening for *baiCD* deletion. Agarose gel electrophoresis visualisation of colony PCRs with flanking primers (Ch_scr_F3 and Ch_scr_R2) to assess for deletions by RiboCas in *P. hiranonis baiCD*. WT = wildtype *P. hiranonis*, M = DNA Marker (2-log ladder; NEB). Numbered lanes = transconjugant colonies plated on thiamphenicol selection with theophylline inducer. Expected fragment sizes (bp) WT = 3167, KO = 1257.

This is the first recorded mutant in *P. hiranonis*, the first recorded *bai* gene mutant in the classical *Clostridial* 7α-dehydroxylators, and only the second recorded *bai* gene mutant in the literature; Jin et al. (2022) have generated an insertional mutant of *baiH* in the commensal *Faecalicatena contorta*, newly suggested as a 7α-dehydroxylator in their paper. Further characterisation of this KO strain is carried out in 3.5.

Further screens for *baiCD* KOs were carried out for the colonies plated onto theophylline, but in a total of 40 colonies this was the only deletion; the rest were WT. This is concordant with the first round using pRECas_CD6_CD in which no deletions were obtained (3.3.3). This suggests that the RiboCas system either is non-functional, and this KO was obtained by chance, or that it is functioning very weakly and allowing extremely inefficient selection. Therefore, the system still requires optimisation for future use and further *bai* gene KOs.

Testing functionality of RiboCas components

Following the successful *baiCD* KO it was chosen as the target for further testing of the RiboCas system as it removes the risk lethality is preventing a gene KO. To confirm that the RiboCas system is non-functioning in *P. hiranonis*, plasmids were constructed containing an sgRNA only (pRECas_CD6_CDsg) and HAs only (pRECas_CD6_CDha), to compare their conjugation efficiency alongside the plasmid containing both sgRNA and HAs (pRECas_CD6_CD)(Figure 3.13). Without the theophylline inducer of *cas9*, one would expect all three plasmids to have a similar transformation efficiency, as *cas9* should not be able to induce lethal cleavages. On induction with theophylline, if the RiboCas system is functioning, it would be expected that transconjugants of pRECas_CD6_CDsg would have no growth as without an editing template for homologous recombination the target sequence is present in the genome and will be lethally cleaved.



Conjugation Efficiency ± SD	CDsg	CDha	CD
	1.4×10^{-8}	5.7×10^{-10}	8.7×10^{-9}
	4.8×10^{-9}	5.7×10^{-10}	3.0×10^{-9}

Figure 3.13 Testing conjugation efficiency of RiboCas components. Comparison of the conjugation efficiency in *P. hiranonis* of the pRECas_CD6_MCS backbone with separate elements of the RiboCas application specific module, to determine function. Three plasmids of pRECas_CD6_MCS with the following inserts, used for the *baiCD* knockout: sgRNA only (pRECas_CD6_CDsg; CDsg), HAs only (pRECas_CD6_CDha; CDha) and both sgRNA and HAs (pRECas_CD6_CD). No significant difference between any conjugation efficiency (one-way ANOVA with Tukey's test for multiple comparisons, p-values: CDsg vs CDha, p=0.0574; CDsg vs CD, p=0.4943; CDha vs CD, p=0.2622). Data represent mean values of the conjugation in technical triplicate ± SD.

On selection without inducer there was no significant difference in conjugation efficiencies of pRECas_CD6_CD, pRECas_CD6_CDsg or pRECas_CD6_CDha (one-way ANOVA with Tukey's test for multiple comparisons, p-values: CDsg vs CDha, p=0.0574; CDsg vs CD, p=0.4943; CDha vs CD, p=0.2622). There was, however, large variation observed among the triplicates for pRECas_CD6_CDsg and pRECas_CD6_CDha. Without inducer these efficiencies are as expected. The successful conjugation and growth of pRECas_CD6_CDsg transconjugants could

demonstrate a limited background leaky expression of *cas9*, but it is not possible to distinguish between this and a total lack of expression by P_{fdxE}.

Following conjugation, transconjugants were re-streaked onto BHISS supplemented with thiamphenicol to confirm presence of the plasmid, and then re-streaked onto BHISS plus theophylline. Strains harbouring each of the three plasmids grew on plates containing theophylline (8/8 colonies), and colony PCR of *baiCD* showed a WT genotype across all samples for all three plasmids. The WT genotype for pRECas_CD6_CDha demonstrates an undetectable spontaneous homologous recombination rate without targeting of *cas9* by the sgRNA. The growth of strains containing pRECas_CD6_CDsg demonstrates a non-functioning RiboCas system irrespective of possible homologous recombination issues, as this should be lethal, as does the WT genotype for pRECas_CD6_CD, as one would expect some selection of gene deletions. This could be due to a non-targetting SEED sequence in the sgRNA, although is unlikely considering the system was already shown as essentially non-functional for three different SEED sequences (3.3.4). Alternatively, it could demonstrate a lack of expression of the sgRNA by its P_{araE} promoter, lack of expression of *cas9* by P_{fdxE}, or a lack of P_{fdxE} inducibility with theophylline. Taken together these results further demonstrate that the RiboCas system does not currently function in *P. hiranonis*.

Promoters

The first attempts to identify the causal error of RiboCas in *P. hiranonis* highlighted a high probability of inactive promoters. Whether they are the only cause remains to be determined but clarifying promoter functionality in *P. hiranonis* is important regardless, for both RiboCas and possible future tools. To achieve this, a reporter assay was designed for promoters used in RiboCas and other promoters commonly used in clostridial tools. Reporter assays in *P. hiranonis* are not present in the literature.

Previous reporter systems in clostridia have been limited to enzyme reporters, for example using the *catP* gene and measuring CAT (chloramphenicol acetyltransferase) activity (Shaw, 1975). Unfortunately, unlike fluorescent reporters these do not allow real time analysis for flow cytometry or protein localisation, but fluorescent reporters generally require oxygen for chromophore maturation and therefore cannot be used in

anaerobic clostridia. To overcome these issues, commercial ligands such as TF-Amber (www.the-twinkle-factory.com) have now become available for use with the fluorescence- activating and absorption-shifting tag protein (FAST), and have been shown to perform under anaerobic conditions, for example in *Clostridium acetobutylicum* (Streett, Kalis, & Papoutsakis, 2019) and *Clostridium autoethanogenum* (Poulalier-Delavelle et al., 2023). Additionally, the FAST assay requires fewer steps and is less labour intensive than assays based on CAT activity, as it only requires labelling with fluorogen and is not reliant on protein quantification. It also has a more versatile repertoire, for example protein-protein interactions and multi-labelling. Because of this it was decided to optimise the FAST assay in *P. hiranonis*.

A suite of promoters commonly used in clostridia was chosen to assay, and these were cloned into pMTL84151 upstream of *FAST*. Promoters were cloned with NotI and NdeI either using a primer dimer reaction, or with PCR using a plasmid template, with primers outlined in Table 3.13. The FAST fragment (*FAST* upstream of the *pyrEhydA* terminator) was generated with NdeI and AscI sites by PCR from pMTL8315_ptcdB_FAST (ClProm_FAST_NdeI_F1 and ClProm_FAST_AscI_R1). Promoter and FAST fragments were digested with respective enzymes and ligated into the digested pMTL84151 backbone (NotI and NdeI). Insertion was confirmed by colony PCR (primers ColE1_tra_F2 and pCD6_R1) and Sanger Sequencing. Final plasmids were then transformed into *E. coli* CA434.

Table 3.13 Promoters to assay for use in *P. hiranonis* genetic tools. Promoters commonly used in clostridial tools chosen to assess in *P. hiranonis* by a FAST reporter assay. Primers used in construction of reporter plasmid for each promoter. Promoter and FAST sequence inserted into pMTL84151. PD = primer dimer.

Promoter	Description	Construction method	Promoter Primers	Promoter Template	Final plasmid
Promoter less	No promoter for negative control.	-	-	-	pMTL84151_FAST_neg
P_{fdx}	Promoter of the *Clostridium sporogenes* ferredoxin gene. Constitutive.	PCR	ClProm_Pfdx_NotI_F1 ClProm_Pfdx_NdeI_R1	pMTL_HZ13_Pfdx	pMTL84151_FAST_fdx
P_{fdxE}	P_{fdx} under the control of the theophylline responsive riboswitch E.	PCR	ClProm_PfdxE_NdeI_F1	pRECas1_MCS	pMTL84151_FAST_fdxE

			ClProm_PfdxE_NotI_R1		
P_{araE}	Promoter of the *Clostridium acetobutylicum araE* gene. Putatively constitutive.	PCR	ClProm_ParaE_NotI_F1 ClProm_ParaE_NdeI_R1	pRECas1_MCS	pMTL84151_FAST_araE
P_{thl}	Promoter of the *Clostridium acetobutylicum thl* gene. Putatively constitutive.	PD	ClProm_Pthiol_NdeI_R2 ClProm_Pthiol_NotI_F2	psbrcas9_630pyrE_ptcdb_PI	pMTL84151_FAST_thl
P_{j23119}	Synthetic promoter.	PD	ClProm_Pj23119_NotI_F1 ClProm_Pj23119_NdeI_R1	pRECas2_MCS	pMTL84151_FAST_j23

Conjugation into *P. hiranonis* was attempted for all six reporter plasmids, selecting with thiamphenicol. All but pMTL84151_FAST_fdx had successful transfer, despite re-attempting the conjugation.

Assessment of RiboCas promoters

In preparation for the FAST reporter assay, growth of *P. hiranonis* harbouring pMTL84151_FAST_neg was assessed. Overnight cultures were prepared by inoculation of BHISS broth with *P. hiranonis* and used to inoculate 50 mL of BHISS broth to produce a starting OD_{600} of 0.05. Thiamphenicol selection was included in the overnight and final culture of pMTL84151_FAST_neg.

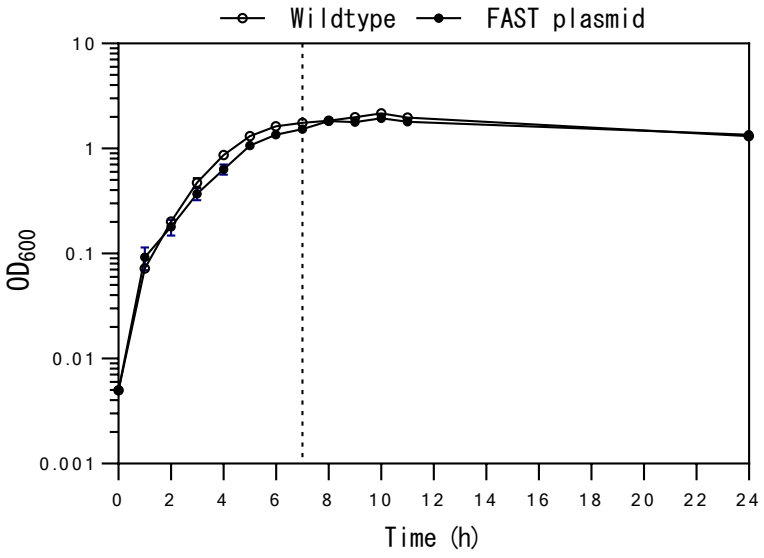


Figure 3.14 Growth of *P. hiranonis* with FAST control plasmid. Growth curves of *P. hiranonis* with and without pMTL84151_FAST_neg, measured by OD_{600}. WT strain grown in liquid BHISS medium, plasmid strain grown in BHISS + thiamphenicol. Data represent mean values of three independent cultures ± SD.

The assay demonstrates that growth is consistent enough to take one time point for both WT and plasmid strains, and 7 h was chosen as a mid-exponential point for sampling (Figure 3.14).

The FAST reporter assay was then carried out for the clostridia promoters, shown in Figure 3.15, according to the protocol 2.3.8. For the inducible promoter $P_{fdxE,}$ theophylline at 5mM was added after 4 h.

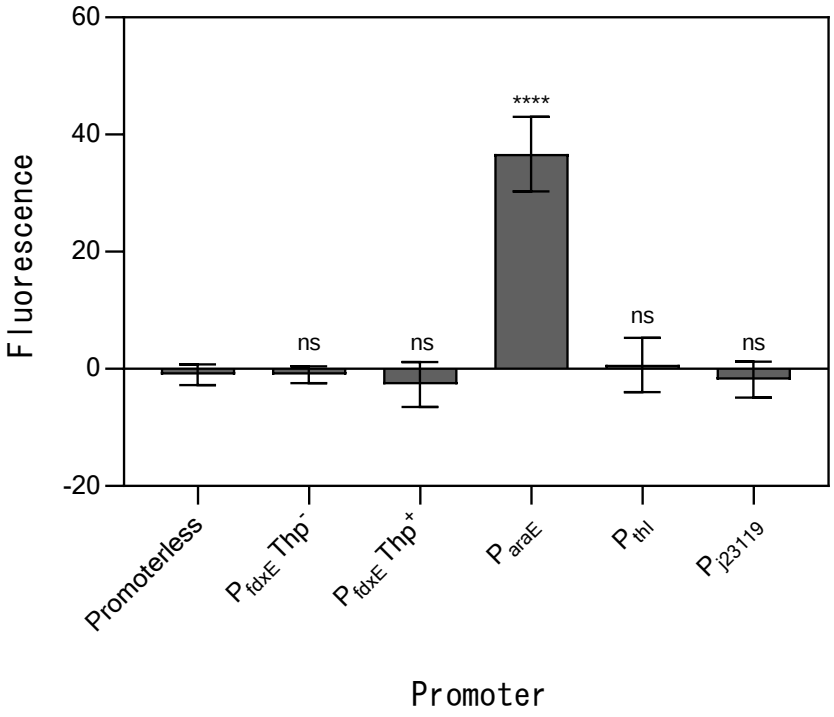


Figure 3.15 FAST reporter assay for *Clostridial* promoters in *P. hiranonis*. *P. hiranonis* harbouring the following reporter plasmids was cultured for 7 h in BHISS liquid medium + thiamphenicol: pMTL84151_FAST_neg, pMTL84151_FAST_fdxE, pMTL84151_FAST_araE, pMTL84151_FAST_thl and pMTL84151_FAST_j23. At 4 h theophylline (5 mM) was added to $P_{fdxE}Thp^{+}$. Samples were washed and the TF-Amber ligand added. Fluorescence was measured immediately at an excitation wavelength of 499 nm and emission at 558 nm using a microplate reader. Readings were blanked to the fluorescence of the plasmid-less WT strain. Statistical analysis was carried out using one-way ANOVA with Dunnett's test for multiple comparisons against the promoter-less control; p-values are indicated as non-significant (ns), or $p<0.0001$ (****). Data represent mean values of three independent cultures in technical triplicate ± SD.

The results indicate that of all the promoters tested, only P_{araE} has a higher expression than the promoter-less control in *P. hiranonis* (Dunnett's test, $p<0.0001$); the other promoters had no signal and were no different to the promoter-less control. This is surprising given that the other promoters have demonstrated activity in several other clostridia, with P_{thl} from *C. acetobutylicum* in particular demonstrating high levels of expression (Tummala, Welker, & Papoutsakis, 1999). The lack of activity of P_{fdxE} with and without theophylline either demonstrates low leaky expression and a non-

inducibility with theophylline, or that the promoter is totally non-functional in *P. hiranonis*. Unfortunately, this cannot be ascertained without the P_{fdx} reporter plasmid. Regardless, it does suggest that *cas9* is not being expressed in RiboCas.

The P_{araE} promoter was shown to have some expression. However, without a positive control it is not possible to qualify the level of said expression and it therefore may not be sufficient for the sgRNA in RiboCas. Promoter studies have not previously been carried out in *P. hiranonis*, and therefore a positive control is not available. Whilst this would have improved the assay reliability, the signal from P_{araE} does validate the assay, as does the lack of expression from the promoter-less control, suggesting that FAST is a valid reporter gene.

Theophylline dose-response

To confirm a lack of expression of P_{fdxE}, and to ensure that the 5 mM concentration of theophylline used was not responsible for a lack of P_{fdxE} expression in the FAST assay, a theophylline dose-response was assessed (Table 3.14).

Table 3.14 Theophylline dose response of P_{fdxE} expression in *P. hiranonis*. *P. hiranonis* harbouring pMTL84151_FAST_fdxE was grown in BHISS medium with theophylline selection for 7 hours. Theophylline was added at 4 hours. Samples were taken and fluorescence measured (excitation wavelength of 499 nm and emission at 558 nm) on addition of TF-Amber.

Theophylline conc. (mM)	Relative fluorescence (± S.D)
0.0	10 (± 7)
2.5	2 (± 6)
5.0	10 (± 4)
7.5	7 (± 5)
10.0	4 (± 2)
pMTL84151_FAST_neg	8 (± 10)

With no fluorescence measured at any concentration, and there being no difference to the promoter-less control, this confirms that there is no expression of P_{fdxE}.

3.4 Modification of genetic tools for *P. hiranonis*

As diagnostic assays for the RiboCas system suggest that it is a lack of sgRNA and/or *cas9* expression that is preventing gene KOs, due to inactive promoters, it was decided to identify and characterise native promoters for use in *P. hiranonis* genetic tools, and to use them in a modified RiboCas system.

3.4.1 Identification and assessment of native promoters

Choice of candidate native promoters

It was decided that to allow the highest chance of identifying expression from native promoters in *P. hiranonis*, promoters for genes that are likely to be constitutively expressed and at high levels should be studied. This should limit the impact of sample timing on fluorescence levels and ensure that any sensitivity issues of the FAST assay, which were not eliminated as a possible issue previously, should not prevent identification of a suitable promoter.

To provide a starting point for promoter choice, papers by Karlin, Mrázek, Campbell, & Kaiser (2001) and Song et al. (2016) were studied. Karlin et al. (2001) have predicted highly expressed genes from the genomes of four bacteria; *E. coli*, *Haemophilus influenza*, *Vibrio cholerae* and *Bacillus subtillis.* Predictions were made based on codon frequencies, translation and transcription processing, chaperone/degradation standards and average gene codon frequencies. Whilst these bacteria may not be closely related to *P. hiranonis*, it was hoped that the promoter responsible for a gene highly expressed across all four diverse species would also afford high expression in *P. hiranonis*. In addition to predicted values, Song et al. (2016) were also searching for a native promoter to use in *B. subtillis* genetic tools, so measured *gfp* fluorescence downstream of 84 predicted promoter sequences.

Genes with high predicted expression across all four species in Karlin et al. (2001), or shown to be highly expressed in *B. subtillis* by Song et al. (2016), were assigned a low/medium/high rating of high expression based on their scores in the paper. They were then searched for by name in the *P. hiranonis* genome. Searches were focussed on genes for a range of major functions, for example transcription/translation and metabolism. If identified in *P. hiranonis*, the location in the genome was noted to

enable a range of distances from the origin to be covered, as this can affect gene expression (Lato & Golding, 2020). Only genes that were not directly adjacent to upstream genes were included and the distance from the upstream gene was noted, to aid in promoter identification. The upstream intergenic region was then inputted into the Softberry Inc promoter prediction software (BPROM) (http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb) to determine if the promoter of the highly expressed gene could be putatively identified. These results are outlined in Table 3.15.

Table 3.15 Promoter candidates in *P. hiranonis* from genes with high predicted expression in other organisms. Genes predicted as highly expressed in bacteria by Karlin et al. (2001) (paper 1) and Song et al. (2016) (paper 2). Predicted expression assigned based on low/medium/high score of high expression. Location identified in *P. hiranonis*, and upstream intergenic sequences assessed for promoter presence by BPROM software.

Gene	Name	Function	Location	Predicted level	Upstream intergenic	Paper	Promoter predicted
rpoB	DNA-directed RNA polymerase subunit beta	translation transcription	94,630	Medium	295 bp	1	yes
fusA	elongation factor G	translation transcription	103,000	high	62 bp	1	no
rplC	50S ribosomal protein L3	ribosomal	107,000	high	94 bp	1	no
fba	class II fructose-1,6-bisphosphate aldolase	glycolysis	343,000	high	313 bp	1	yes
thrS	threonine--tRNA ligase	Translation transcription	532,000	medium	95 bp	1	yes
dnaJ	molecular chaperone DnaJ	Misfolding chaperone	750,000	high	200 bp	2	yes
gap	type I glyceraldehyde-3-phosphate dehydrogenase	glycolysis	935,000	high	381 bp	1	yes
pgk	phosphoglycerate kinase	glycolysis	936,000	high	280 bp	1	yes
trxA	thioredoxin	redox	1,111,000	medium	167 bp	2	yes
frr	ribosome recycling factor	translation transcription	1,577,000	low	172 bp	1	yes
secG	preprotein translocase subunit SecG	chaperone/degradation	2,060,000	low	363 bp	1	yes
groL	chaperonin GroEL	Misfolding chaperone	2,370,000	high	68 bp	2	no

groS	co-chaperone GroES	heat shock protein	2,372,000	high	175 bp	2	yes

As shown in Table 3.15 a promoter could be predicted in the upstream intergenic region for most of the highly expressed genes that are present in the *P. hiranonis* genome. Promoters were chosen from genes to cover variable locations, functions and expression levels (Table 3.16).

Table 3.16 *P. hiranonis* genes chosen for promoter analysis. Hypothesised expression based on predicted expression in other bacteria by Karlin et al. (2001) and Song et al. (2016).

Gene	Function	Locus tag	Hypothesised expression
secG	Chaperone/degradation	KGNDJEFE_00625	Low
rpoB	Translation/transcription	KGNDJEFE_00058	Medium
thrS	Translation/transcription	KGNDJEFE_02019	Medium
trxA	Redox	KGNDJEFE_01781	Medium
fba	Glycolysis	KGNDJEFE_00256	High
groS	Chaperone/degradation	KGNDJEFE_00378	High
pgk	Glycolysis	KGNDJEFE_00620	High

Comparison of native promoters

As there is no analysis of *P. hiranonis* promoters in the literature, the predicted promoters by Softberry Inc were compared for their promoter elements to assess if consensus sequences could be determined. This was done using Weblogo (http://weblogo.berkeley.edu/) (Crooks, Hon, Chandonia, & Brenner, 2004) (Figure 3.16).

A)

Gene	Predicted -35 and -10 box	Spacer	BP to ATG
secG	AATTTTTGTCGCAGTAAATTATAACCAAATGTAAAATTTGTG	17	150
rpoB	TTTTTTTTACAAAAAATAGAGTTTTTTCTGTTATATTGCATT	17	199
thrS	ATAAATTGCAAAATATAAAATTAATATGTAAATTATTATCAATT	19	31
trxA	TACTCTGGAAGCCTTTTGGCTTCCAGGTAAAATAAATA	13	26
fba	TAATTTTTATTTTTCACGGTATTTTTTATTGTTATGA	12	164
groS	AAGTATTGAAAAATGAAAATAAAAGGTATATTATAATAATTG	17	89
pgk	GTATATTTATAATAGGGAAGTGTGATTAAGTCATACTTTTAA	17	36

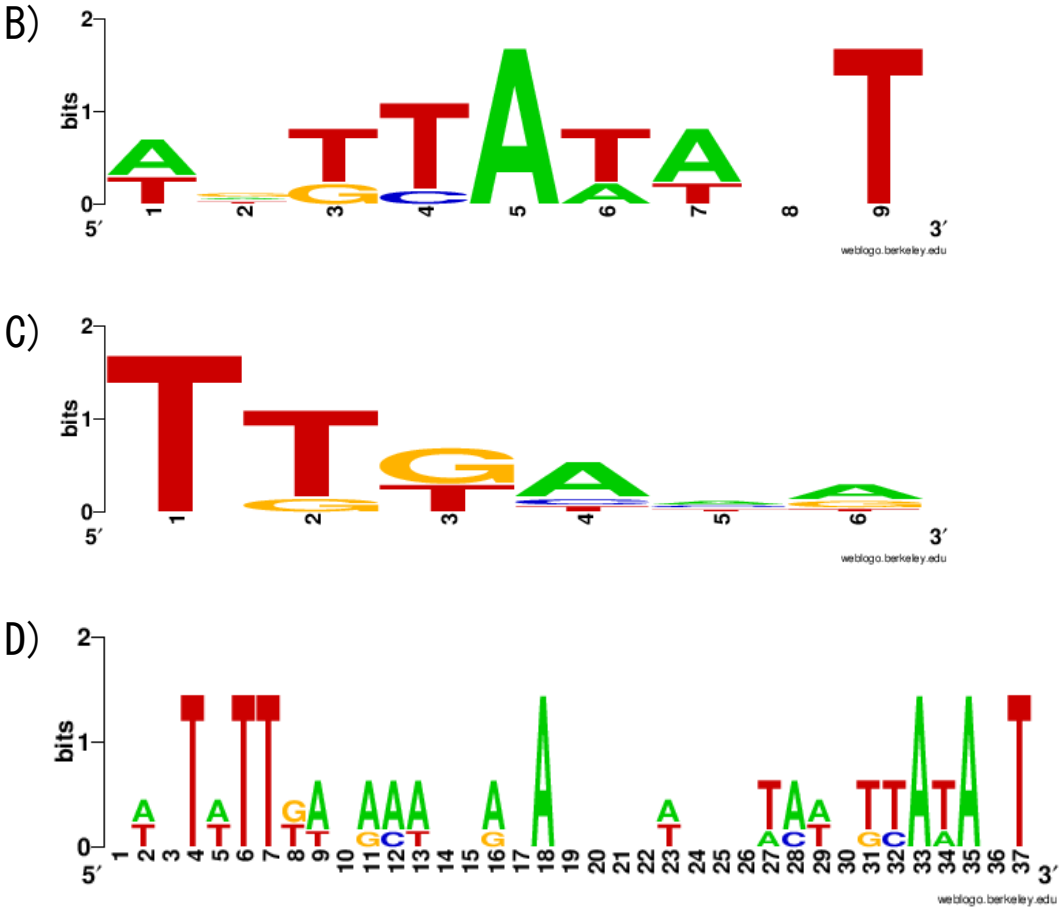


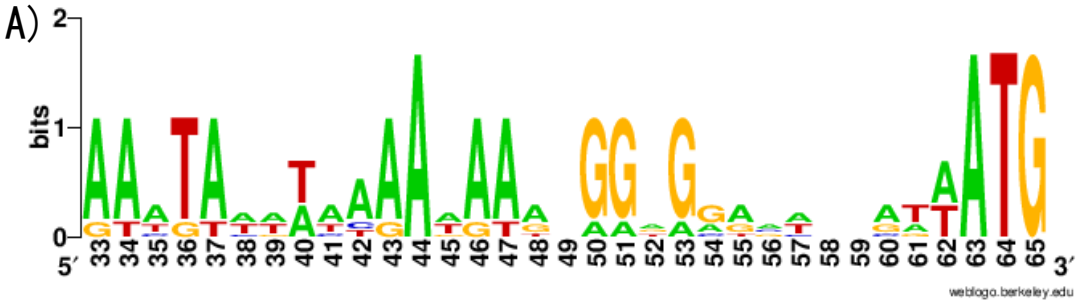
Figure 3.16 Comparison of putative promoter elements in *P. hiranonis*. The promoters of the 7 chosen genes were predicted by BPROM software. A) Predicted -35 and -10 elements (blue), the number of bases (BP) between the elements (spacer), and number of bases between the -10 element and ATG of chosen gene. B) Comparison of the 9 nucleotides of the -10 elements. C) Comparison of the 6 nucleotides of the -35 elements. D) Comparison of the four spacer sequences of equal 17 base length (promoters for *secG*, *rpoB*, *groS*, *pgk*). All sequence logos generated by Weblogo (http://weblogo.berkeley.edu/).

When comparing the -10 elements (Figure 3.16B) as predicted by the BPROM software (Figure 3.16A), there is some conservation, and some similarity to the traditional TATAAT of the Pribnow box. The -35 elements predicted by the software (Figure 3.16A) also showed some similarity to the classical TTGACA (Figure 3.16C), although this was stronger in the first three nucleotides and weaker in the last three nucleotides. There was some variation in the number of nucleotides between the -35 and -10 elements (spacer), ranging from 12-19, but the most common spacer was 17 nucleotides, predicted for four of the promoters. When aligning these four sequences (Figure 3.16D), again some conservation is shown for the -35 element, estimated at position 6 in the web logo, and for the -10 element, estimated at position 32 in the web logo.

There was a wide range of distances between the -10 element and the ATG of the chosen gene, from 26 bp to 199 bp (Figure 3.16A). This is an estimate of the distance between the transcriptional start site (TSS), found approximately downstream of the -10 element, and the ribosome binding site (RBS). This region between the 5' mRNA end and the RBS is known as the 5'-untranslated region (5'-UTR), and can influence translation efficiency, mRNA stability and concentration (F. Chen, Cocaign-Bousquet, Girbal, & Nouaille, 2022; Kosuri et al., 2013). A variability in this distance has also been demonstrated in *E. coli*; the most frequent distance is 12-40 nucleotides, but it has a range of -2 to 553 nucleotides (Rangel-Chavez, Galan-Vasquez, & Martinez-Antonio, 2017). Whilst 5'-UTR distance is one of the various factors that can influence translation, the rules to predict it are still unclear; it has been shown to be affected by a number of factors such as length and location of Shine-Dalgarno sequences, formation of secondary structures and presence of AU-enhancers (Evfratov et al., 2017; Rangel-Chavez et al., 2017). Moreover, the theoretical translation initiation rate has also been shown as poor predictor of the resulting protein level (F. Chen et al., 2022). Therefore, it is difficult to gain an insight into predicted expression levels by the chosen native promoters from their sequence and 5'-UTR alone.

The chosen promoters were also analysed with regards to their RBS (Figure 3.17), which has not been previously elucidated. Firstly the 20 bp upstream of the ATG was aligned, which revealed a likely consensus of GGXG (position 50-53 in Figure 3.17A),

which shows similarity to the core consensus AGGAGG RBS sequence. The promoters were then aligned according to this sequence, from 7 bp before and 5 bp after (Figure 3.17B), to take into account different distances to the ATG. This sequence was then used to estimate an RBS consensus of AANAANNGGNG (Figure 3.17C). Without knowing the length of the RBS it is difficult to predict the exact sequence, but this prediction is in line with other known RBS's, for example the 11 nucleotide AAAGGAGGTGT *B. subtillis* consensus RBS (http://parts.igem.org/Ribosome_Binding_Sites/Prokaryotic/Constitutive/Miscellaneous).



C)

Gene	Sequence																					
secG	T	C	G	A	T	A	A	A	C	A	G	G	A	G	G	G	A	A	C	A	G	T	**ATG**
rpoB	A	A	A	A	T	G	A	G	A	G	G	T	G	A	A	G	A	G	T	G	A	A	**ATG**
thrS	T	A	A	A	A	A	A	G	G	G	G	A	G	G	A	C	A	C	A	A	T	A	**ATG**
trxA	A	A	A	A	A	A	A	T	T	G	G	A	G	G	A	A	C	A	A	A	A	A	**ATG**
fba	A	A	A	A	G	A	A	A	A	G	G	T	G	G	T	A	T	T	T	A	T	T	**ATG**
groS	A	T	A	A	A	A	T	A	G	G	A	G	G	C	G	C	T	A	G	G	T	A	**ATG**
pgk	C	A	A	A	T	A	A	A	A	G	G	A	G	A	A	A	T	T	A	C	T	T	**ATG**

Figure 3.17 RBS comparisons and predictions. Comparing the predicted RBS sequences from the upstream sequences of 7 genes in *P. hiranonis*: *secG, rpoB, thrS, trxA, fba, groS, pgK.* A) Alignment of the 20 nucleotides upstream of the ATG for all 7 genes. B) Alignment of all 7 genes bases on their GGNG sequences; 7 nucleotides upstream, 5 nucleotides downstream. C) 22 nucleotides upstream of the ATG, with predicted AANAANNGGNG RBS highlighted in grey, and matching nucleotides in blue. All sequence logos generated by Weblogo (http://weblogo.berkeley.edu/).

Using the estimated RBS sequence of AANAANNGGNG illustrates a range of distance between the RBS and ATG, from 8-10 nucleotides, although the most common is 9. Again this is similar to that seen in *E. coli*, which also had a range (1 to 15 nucleotides), but an average of 6 (Rangel-Chavez et al., 2017). Experimental data

would be required to confirm the RBS sequence in *P. hiranonis*, but this was considered outside the scope of the project presently.

These promoter predictions are only as good as the software used. Cassiano & Silva-Rocha (2020) have reviewed several promoter prediction tools with regards to specificity, sensitivity, accuracy, and BPROM performed the worst compared to the other nine tools tested. However, many of these other tools are no longer available online, and BPROM does have the advantage of an easy to use interface with fast results, available as a web application. Moreover, the authors note that the majority of tools have been trained using data from *E. coli*, and therefore they are all limited in their application to other organisms (Cassiano & Silva-Rocha, 2020). Regardless, they do provide estimates of promoter elements, which serve as a starting point for experimental design.

Generation of FAST reporter plasmids to assess native promoters

For assessment of the native promoters chosen, FAST plasmids were constructed for a reporter assay. This would allow identification of promoter activity when positioned on a plasmid, and therefore determine suitability for use in genetic tools. It should also allow validation of the FAST assay in *P. hiranonis*; if none of the native promoter generate a fluorescent signal then it is unlikely that the reporter gene is suitable.

The sequence of the upstream intergenic region of the identified genes was taken up to a maximum of 250 bp. For six of the seven promoters a geneblock was synthesised (2.4.10) to include the promoter region directly upstream of the *FAST* gene. Primers were then designed to generate this fragment flanked by NotI and XhoI sites. This fragment and pMTL84151 were then digested and ligated. The promoter/FAST geneblock for *fba* was unable to be synthesised as the complexity score was too high. This reporter plasmid was instead generated using hifi assembly, utilising the same restriction sites as the geneblock plasmids. All primers used are shown in Table 3.17. Promoter/FAST insertion was then screened for by colony PCR using primers ColE1_tra_F2 and pCD6_R1. Unfortunately, no transformants could be generated for P_{rpoB}. Final plasmids were then transformed into *E. coli* CA434.

Table 3.17 *P. hiranonis* native promoter reporter plasmids. *P. hiranonis* promoters chosen to assess by a FAST reporter assay. Primers used in construction of reporter plasmid for each promoter. Promoter and FAST sequence inserted into pMTL84151. Gene blocks generated by Integrated DNA Technologies.

Promoter	Sequence	Construction method	Final plasmid	Primers
secG	250 bp	Gene block	pMTL84151_FAST_secG	PsecG_FAST_Not_F1 CHIRprom_FAST_Xho_R1
rpoB	250 bp	Gene block	N/A	PrpoB_FAST_Not_F1 CHIRprom_FAST_Xho_R1
thrS	250 bp	Gene block	pMTL84151_FAST_thrS	PthrS_FAST_Not_F1 CHIRprom_FAST_Xho_R1
trxA	98 bp	Gene block	pMTL84151_FAST_trxA	PtrxA_FAST_Not_F1 PtrxA_FAST_Xho_R1
fba	235 bp	Hifi	pMTL84151_FAST_fba	Pfba_hifi_F1 Pfba_hifi_R1 FAST_fba_F1 FAST_fba_R1
groS	178 bp	Gene block	pMTL84151_FAST_groS	PgroS_FAST_Not_F1 CHIRprom_FAST_Xho_R1
pgk	250 bp	Gene block	pMTL84151_FAST_pgk	Ppgk_FAST_Not_F1 CHIRprom_FAST_Xho_R1

Conjugation into *P. hiranonis* was attempted for all six reporter plasmids, selecting with thiamphenicol. All but pMTL84151_FAST_trxA had successful transfer, despite re-attempting the conjugation.

Promoter FAST assay

The five *P. hiranonis* reporter strains, plus the promoter-less control generated in 3.3.4, were grown in BHISS with thiamphenicol selection and samples taken at 7 h for a FAST reporter assay (Figure 3.18).

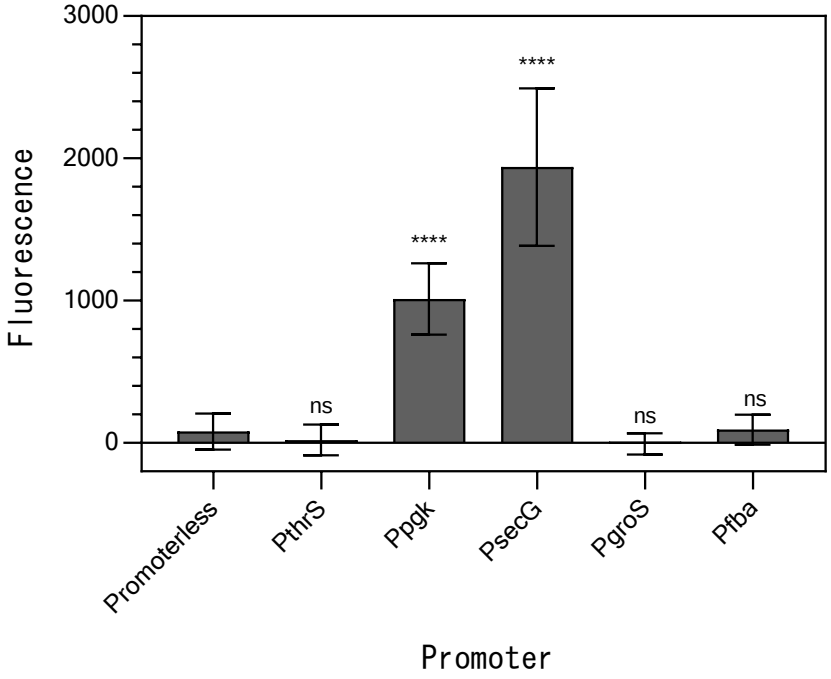


Figure 3.18 FAST reporter assay for native promoters of *P. hiranonis*. *P. hiranonis* harbouring the following reporter plasmids was cultured for 7 h in BHISS liquid medium + thiamphenicol: pMTL84151_FAST_neg, pMTL84151_FAST_thrS, pMTL84151_FAST_pgk, pMTL84151_FAST_secG, pMTL84151_FAST_groS, pMTL84151_FAST_fba. Samples were washed and the TF-Amber ligand added. Fluorescence was measured immediately at an excitation wavelength of 499 nm and emission at 558 nm using a microplate reader. Readings were blanked to the fluorescence of the plasmid-less WT strain. Statistical analysis was carried out using one-way ANOVA with Dunnett's test for multiple comparisons against the promoter-less control; p-values are indicated as non-significant (ns), or $p<0.0001$ (****). Data represent mean values of three independent cultures in technical triplicate ± SD.

When comparing the *P. hiranonis* native promoters to the promoter-less control, P_{thrS}, P_{groS} and P_{fba} had no significant difference, but P_{pgk} and P_{secG} did (one-way ANOVA with Dunnett's test; $p<0.0001$ for both promoters). Whilst these results are encouraging as they indicate that the FAST assay is functional in *P. hiranonis*, and identify two candidate promoters for further use, it was surprising that the other three did not demonstrate expression; these were chosen based on a probability of high expression. Because of this, the FAST assay was repeated to ensure that the results

were reproducible, but also to take samples at multiple phases of growth to ensure that all possible time points of expression were covered (Figure 3.19). As demonstrated in Figure 3.14, time points of 3 h, 5 h and 7 h should be sufficient to do this.

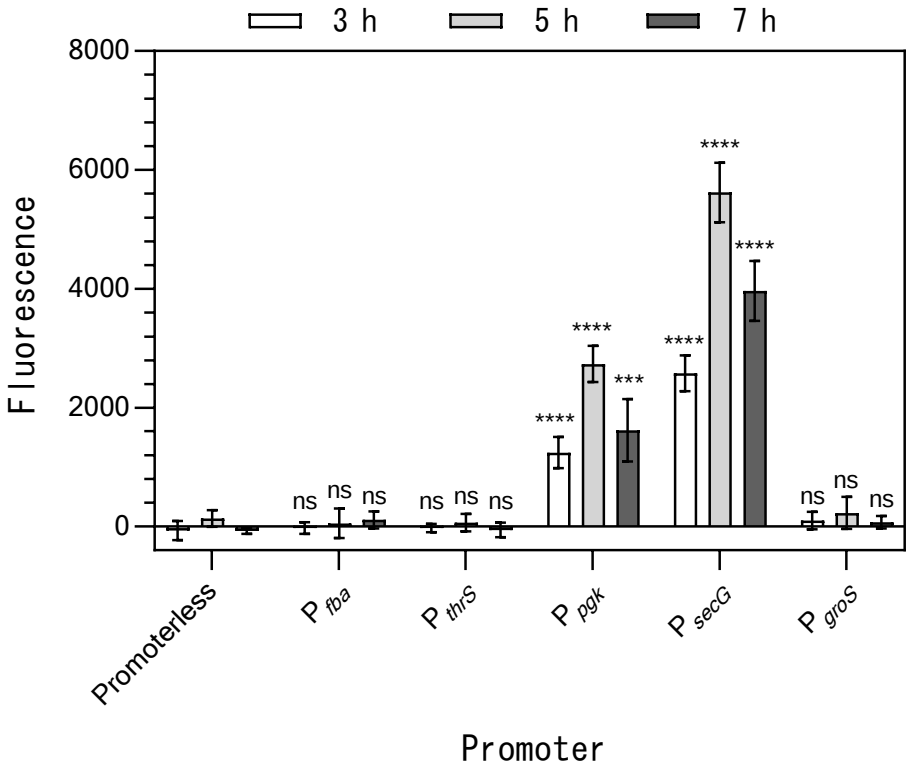


Figure 3.19 Native promoter FAST assay to compare timing of sampling. *P. hiranonis* harbouring the reporter plasmids was cultured for 3, 5 and 7 hours in BHISS liquid medium + thiamphenicol. Data represent mean values of three independent cultures in technical triplicate ± SD. Statistical analysis was carried out using one-way ANOVA with Dunnett's test for multiple comparisons against promoter-less control; p-values are indicated as non-significant (ns), p<0.005 (***), p<0.0001 (****).

A two-way ANOVA with Tukey's test for multiple comparisons, p-values. Pgk: 3h vs 5h <0.0001, 3h vs 7h 0.2182, 5h vs 7h <0.0001. SecG: 3h vs 5h, 3h vs 7h, 5h vs 7h all p-values are <0.0001. P-values not indicated on graph.

At all three time points the results mirrored observations of the initial assay; P_{pgk} and P_{secG} demonstrated expression, whereas the other three promoters did not, and P_{secG} had consistently higher expression than P_{pgk}. Moreover, they both followed a similar

pattern of expression over time, with highest expression at 5 h (two-way ANOVA with Tukey's test for multiple comparisons). At 5 h the expression of P_{secG} is approximately twice that of P_{pgk} (means of 5621 vs 2734), with a p-value of 0.001 (unpaired two-tailed Student's t-test). These results validate further the FAST assay, and propose with reasonable confidence P_{pgk} and P_{secG} as candidate promoters to use in *P. hiranonis* genetic tools, with two varying expression levels.

Several results from the assay are not as expected from predicted expression levels. Firstly, it was hypothesised that all chosen promoters would allow expression of FAST, but only P_{pgk} and P_{secG} did. Moreover, although one cannot comment on absolute expression levels as the results are all relative given a lack of confirmed strong promoter as a positive control, the levels of P_{pgk} and P_{secG} expression in the FAST assay were different to those predicted; P_{pgk} as high and P_{secG} as low, whereas in the assay they are ranked in reverse. Finally, as P_{fba} and P_{groS} are from the same functional categories as P_{pgk} and P_{secG}, in glycolysis and protein folding/degradation respectively, they should theoretically be expressed in similar conditions albeit to possibly different expression levels, so one might have also expected these to be expressed.

These unexpected results perhaps demonstrate the interplay of other factors on downstream gene expression, independent of promoter strength. As the promoter, 5'-UTR and RBS were all encompassed in the sequence included in the reporter plasmids as it was not possible to confidently separate the promoter alone, the assay does not allow direct comparisons of promoter strength. As indicated previously, the 5'-UTR for example can influence the efficiency of translation initiation, as can the strength of the RBS, which was predicted to differ between the chosen promoters (Figure 3.17). Moreover, the promoters were chosen based on predictions of high gene expression by Karlin et al. (2001), but these predictions were not solely based on promoters and were used as a starting point only to choose candidate genes.

As all promoters and RBS's were native to *P. hiranonis*, it is highly unlikely that any of the promoters studied are non-functional. The most likely explanation for a lack of expression for P_{thrS}, P_{groS} and P_{fba} is that they require inducers that were not present in the assay conditions, although literature searches were carried out originally to reduce the risk of this. If future work required more candidate promoters to be identified, or

to increase understanding of the promoters tested, a variety of conditions could be tested, for example different temperatures or pH. However, for example in *B. subtillis* this did not always produce results as expected; the authors expected that as it controls a heat shock protein P_{groS} would respond to stress treatments, but expression did not differ (Y. Song et al., 2016). It would therefore be unlikely that this would allow a targeted approach to promoter optimisation, would require a larger promoter library, and would be a complex undertaking.

There are also further possible reasons why P_{thrS}, P_{groS} and P_{fba} were not expressed in either assay. It is possible that for P_{fba}, where the sequence taken upstream of *fba* was limited to 250bp, that this was not sufficient for a functional promoter. However, for *thrS* and *groS* their complete upstream intergenic region was taken, so is an unlikely explanation. It could be that the time points taken did not cover their points of expression, however this is also unlikely as most phases of growth were covered, and they were sufficient for P_{pgk} and P_{secG} which share similar functions. It could also be that they have very low expression levels but that the FAST assay in *P. hiranonis* is not sensitive enough to detect these, and it is not possible to clarify this without a control promoter.

Whilst there are still outstanding questions in relation to the apparent expression of the promoters in this assay, the main outcome is that it gives confidence in the sequences chosen for P_{pgk} and P_{secG} to provide expression of exogenous genes in *P. hiranonis*, when grown under the same conditions used in the FAST assay.

3.4.2 Use of native promoters in RiboCas

Construction of RiboCas vectors with native promoters

The identification of P_{pgk} and P_{secG} as functioning promoters allows their consideration for use in genetic tools in *P. hiranonis.* As the existing promoters of the RiboCas system were identified as a probable cause of its inability to generate gene knockouts, it was decided to exchange them. The promoter with lower expression, P_{pgk}, was chosen to replace the P_{fdxE} promoter of *cas9*, to attempt to reduce any possible toxicity issues that may arise from constitutive expression. The other promoter P_{secG} was chosen to replace the P_{araE} promoter of the sgRNA.

Replacing the two promoters in pRECas1_CD6_MCS gave the new backbone pCHRE1 for insertion of relevant application module specific elements (Figure 3.20). Other elements in pCHRE1 remain the same as pRECas1_CD6_MCS. Hifi PCR was used to construct pCHRE1 from three fragments (*cas9* plus the two promoters), plus the backbone (pRECas1_CD6_MCS digested with NotI and SalI). Templates and primers are listed in Table 3.18. All restriction sites remain the same in pCHRE1 as pRECas1_CD6_MCS. The final plasmid was confirmed by two colony PCRs with primers Cas9scr1 and ColE1_tra_F2, and Cas9scr1R and pCD6_R1, followed by Sanger sequencing.

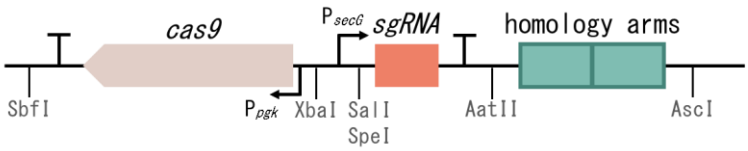


Figure 3.20 Application specific module of the modified RiboCas system, pCHRE1. The promoters for *cas9* and the sgRNA module in pRECas1_MCS (Cañadas et al., 2019) have been exchanged for native promoters from *P. hiranonis:* P_{pgk} and P_{secG} respectively. The generates the new RiboCas backbone pCHRE1. The other components of the application specific module and plasmid backbone remain the same as pRECas1_MCS.

Table 3.18 Fragments used to generate pCHRE1. Primers used to generate the RiboCas backbone with native *P. hiranonis* promoters for *cas9* and the sgRNA.

Fragment	Length (bp)	Primers	Template
cas9	4148	CHRE1_cas9_F1 CHRE1_cas9_R2	pRECas1_MCS
P_{pgk}	286	CHRE1_Ppgk_F1 CHRE1_Ppgk_R1	*P. hiranonis* DNA
P_{pgk}	255	CHRE1_PsecG_F1 CHRE1_PsecG_R1	*P. hiranonis* DNA

The pCHRE1 backbone was then used to generate a sample of RiboCas vectors for *bai* gene knockouts. These were generated for *baiB, baiCD* and *baiF* by digestion of pRECas1_MCS_B, pRECas1_MCS_CD and pRECas1_MCS_F with SalI and AscI to generate the sgRNA/homology arm editing cassette, for insertion by ligation into the MCS of pCHRE1 digested with SalI and AscI. This generated pCHRE1_B, pCHRE1_CD and pCHRE1_F. Final plasmids were confirmed by colony PCR with Cas9scr1R and pCD6_R1, followed by Sanger sequencing. The three plasmids were then transformed into *E. coli* CA434.

Conjugation of vectors and results

The plasmids pCHRE1_B, pCHRE1_CD and pCHRE1_F were conjugated into *P. hiranonis*, allowing 24 h for conjugation. Transconjugants were selected for with thiamphenicol and re-streaked to confirm presence of the plasmid. Colony PCR was then carried out to assess for gene knockouts using for *baiB* (CHIR_baiop_scr_F1 and Ch_scr_R2) for *baiCD* (Ch_scr_F3 and Ch_scr_R2) and for *baiF* (Ch_scr_F6 and Ch_scr_R5). Unfortunately, all transconjugant colonies had the WT genotype, and no KOs were achieved for any of the target genes.

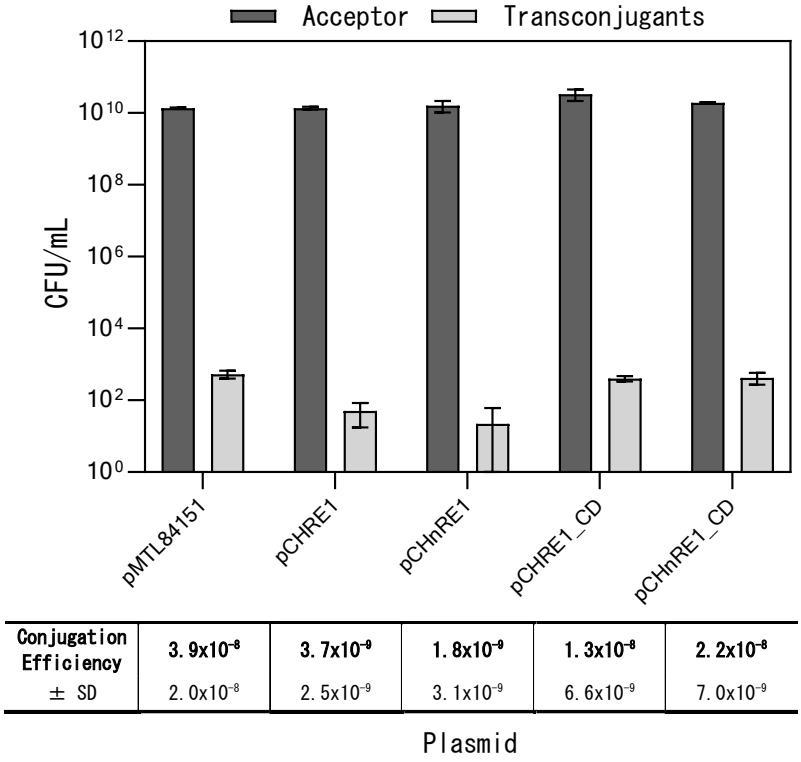
Troubleshooting RiboCas

The lack of knockouts using pCHRE1 was unexpected, as it was assumed that replacing the promoters with those confirmed to be active would overcome previous issues. One of the predicted new issues was that of toxicity, that the constitutive expression of *cas9* would be too toxic for *E. coli* or *P. hiranonis*, or both. As pCHRE1 was able to be cloned in *E. coli* this suggests that there is no toxicity; either P_{pgk} is non-functional in *E. coli*, and therefore *cas9* is not expressed, or it is expressed at low enough levels to not be toxic. The other issue of toxicity could have arisen in *P. hiranonis*, which could have prevented transfer of the plasmid by conjugation, however, this was not the case. The constitutive expression of *cas9* could have also resulted in no mutants due to limited time for homologous recombination before DSBs occurred. To ensure that single nucleotide polymorphisms (SNPs) had not arisen which could have overcome toxic effects in *E. coli* or *P. hiranonis*, both the *cas9* and promoter regions of pCHRE1 were sequenced, and no SNPs were observed.

To investigate the issues of toxicity further, it was decided to trial the use of a nickase instead of *cas9*. Whilst the exact mechanism is still to be elucidated, a nickase *cas9* has been shown to improve editing efficiency and be more effective than the *cas9* for editing in *Clostridium acetobutylicum* (Q. Li et al., 2016) and *Clostridium cellulolyticum* (Xu et al., 2015). Explanations for this include that the reduced lethality of nicks compared to DSBs could still allow for selection against WT but allow more time for homologous recombination, or that the nickase could improve the rate of homologous recombination in a similar manner to eukaryotic nickases (Luo et al., 2016). Regardless of mechanism, it was decided to replace *cas9* in pCHRE1 with a nickase.

It was decided to use the nickase *cas9* (*ncas9)* with a point mutation in the RuvC nuclease domain, D10A, resulting in a single functioning nuclease domain (HNH) that induces a single and not double stranded break (Luo et al., 2016). The pCHRE1 plasmid was generated with *ncas9* instead of *cas9* by hifi PCR. The fragments *ncas9* (pCHn_nCas9_F1 and pCHn_nCas9_R1, template vFS40), P_{pgk} (pCHn_Ppgk_F1 and pCHn_Ppgk_R1, template *P. hiranonis* gDNA) and P_{secG} (pCHn_PsecG_F1 and pCHn_PsecG_R1, template *P. hiranonis* gDNA) were inserted into the pRECas1_CD6_MCS backbone digested with NotI and SalI. The final plasmid, pCHnRE1, was confirmed by two colony PCRs (Cas9scr1 and ColE1_tra_F2, Cas9scr1R and pCD6_R1), followed by Sanger sequencing. To allow genome editing to be assessed, the application specific module was digested from pRECas1_MCS_CD by SalI And AscI, and ligated into the pCHnRE1 backbone digested with the same restriction enzymes. The final plasmid, pCHnRE1_CD, was confirmed by colony PCR (Cas9scr1R and pCD6_R1), followed by Sanger sequencing.

To test if there were any improvements in toxicity or genome editing by using the *ncas9*, a conjugation efficiency assay was carried out for: pCHRE1, pCHnRE1, pCHRE1_CD and pCHnRE1_CD (Figure 3.21).



Conjugation Efficiency	3.9×10^{-8}	3.7×10^{-9}	1.8×10^{-9}	1.3×10^{-8}	2.2×10^{-8}
± SD	2.0×10^{-8}	2.5×10^{-9}	3.1×10^{-9}	6.6×10^{-9}	7.0×10^{-9}

Figure 3.21 Testing conjugation efficiency of pCHRE1 and pCHnRE1. Comparison of the conjugation efficiency in *P. hiranonis* of the native promoter RiboCas backbone with both *cas9* and *ncas9*, with and without the application specific module for baiCD deletion. Data represent mean values of the conjugation in technical triplicate ± SD.

For conjugation efficiency, statistical analysis was carried out using one-way ANOVA with Dunnett's test for multiple comparisons against the pMTL84151 control: pCHRE1 (p=0.0003), pCHnRE1 (p=0.0002), pCHRE1_CD (p=0.0057), pCHnRE1_CD (p=0.0304). Not shown in table.

Two-way ANOVA with Tukey's test for multiple comparisons of conjugation efficiency, significant p-values. 84151 vs: pCHRE1, pCHnRE1 and pCHRE1_CD. pCHRE1 vs pCHnRE1_CD2 (p=0.0464). pCHnRE1 vs pCHnRE1_CD (p=0.0276). Not shown in table.

All four test plasmids exhibited a reduced conjugation efficiency when compared to the pMTL84151 control (one-way ANOVA with Dunnett's test: pCHRE1, p=0.0003; pCHnRE1, p=0.0002; pCHRE1_CD, p=0.0057; pCHnRE1_CD, p=0.0304). This is as expected due to the increase in size (approximately 4.4 kb for backbones, and 6.5 kb with CD editing templates), and possible toxicity from *cas9*.

There was limited conjugation for both of the backbone plasmids, pCHRE1 and pCHnRE1, and there was no significant difference between the two. When compared to their CD knockout counterparts, there was no difference for pCHRE1, but pCHnRE1_CD had a significantly higher conjugation efficiency than pCHnRE1 (p=0.0276). Colony PCR was carried out on transconjugants for pCHRE1_CD, and pCHnRE1_CD using primers to assess for a *baiCD* KO (Ch_scr_F3 and Ch_scr_R2), but they were all WT.

For the backbones containing *cas9* or *ncas9* only, the result was unexpected if the RiboCas system was working. For example, Li et al. (2016) carried out the same comparison in *C. acetobutylicum*, of *cas9* vs *ncas9* with no sgRNA or editing template, and demonstrated a large reduction in transformation efficiency of *cas9* vs the control plasmid, but little reduction of the *ncas9* plasmid compared to the control. This is likely due to the reduced lethality of off targets effects of the nickase, being single stranded nicks and not double stranded breaks. However, these results demonstrate no difference between *cas9* or *ncas9*, and both had a large reduction in conjugation efficiency compared to the control. It is therefore unlikely that the reduction is due to off-target effects of a functioning CRISPR system, and may be explained by the effects of plasmid size on conjugation efficiency.

The lack of increase in conjugation efficiency for pCHnRE1_CD compared to pCHRE1_CD also suggests that the CRISPR system is non-functional. As previously discussed, one would expect editing efficiency to increase using *ncas9* and therefore transfer efficiency due to increased mutant survival. This was observed by Li et al. (2016). Combining this with a lack of *baiCD* KO generation in either of the pCHRE1_CD or pCHnRE1_CD transconjugants suggests that there is no selection for the mutant genotype and likely no functioning Cas9. If homologous recombination was the sole issue one would expect to obtain no transconjugants with pCHRE1_CD and see an improvement with pCHnRE1_CD. The increase in conjugation efficiency pCHnRE1_CD compared to pCHnRE1 is also difficult to explain.

Overall, it is difficult to identify the cause of the failure of the new RiboCas system with native promoters. The use of native promoters with confirmed expression in *P. hiranonis* should have ensured *cas9* or sgRNA expression, but this did not enable genetic modification as hypothesised.

3.4.3 Development of inducible native promoters

In a functioning CRISPR system, using an inducible promoter to drive *cas9* can allow an increase in editing efficiency by reducing toxicity and improving transfer, as shown by Cañadas et al. (2019). It was therefore decided to develop inducible promoters to use in the RiboCas system in *P. hiranonis*. Although previous assays suggest that this is unlikely to overcome the lack of genome editing by RiboCas, results have been inconclusive as to the cause of this. Moreover, developing inducible promoters could be beneficial by improving the tools available for future work in genetic modification of *P. hiranonis*.

Previous work successfully generating inducible promoters in clostridia has involved the use of theophylline inducible riboswitches, evidenced by Topp et al. (2010) and Cañadas et al. (2019). Riboswitches are a class of regulatory RNAs that regulate gene expression in response to small molecule ligands through conformational switching, commonly through transcription termination, translation initiation or splicing. They are *cis*-acting elements that are found in the 5'-UTRs of mRNAs (Breaker, 2018). In addition to natural riboswitches, synthetic aptamer-metabolite combinations have been developed for their binding to non-natural ligands. The theophylline inducible riboswitch was first developed by Jenison, Gill, Pardi, & Polisky (1994), based on their isolated TCT8-4 aptamer, and has a high affinity for the purine theophylline, a stable and cell permeable molecule. Topp et al. (2010) then generated a library of TCT8-4 variants to establish those with high performance in clostridia. These theophylline riboswitches are composed of an aptamer and a synthetic Shine-Delgado sequence, forming a stem loop structure and functioning by translation initiation. The stem-loop structure prevents binding of the RBS in the absence of theophylline. In the presence of theophylline it binds to the aptamer, altering downstream base pairing and causing a conformational change that releases the Shine-Delgado sequence, allowing RBS binding and translation (Lynch, Desai, Sajja, & Gallivan, 2007).

For development of inducible promoters in *P. hiranonis* it was decided to take the two native promoters characterised in 3.4.1, P_{pgk} and P_{secG}, and insert the theophylline inducible riboswitch.

Compatibility with theophylline

Firstly, theophylline toxicity to *P. hiranonis* was determined with an MIC assay (Figure 3.22A), and then with a growth curve at different concentrations (Figure 3.22B).

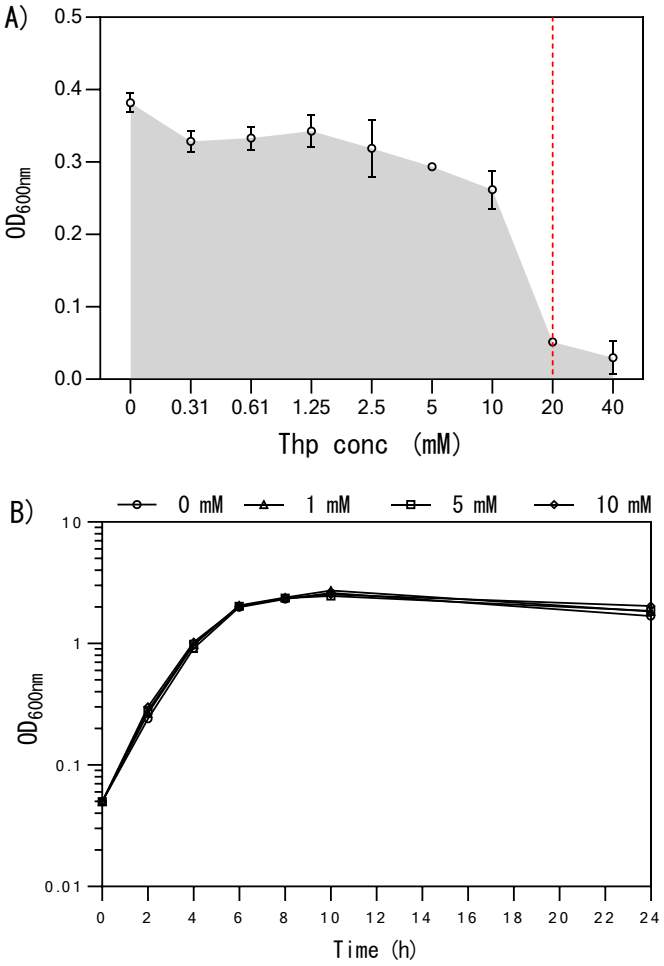


Figure 3.22 Determining suitability of theophylline for use with *P. hiranonis*. A) Minimum inhibitory concentrations of theophylline. *P. hiranonis* WT strain cultured with a range of theophylline concentrations. Growth measured after 24 hours by OD_{600}. B) Growth curves in BHISS + concentrations of theophylline, measured by OD_{600}. Data represent mean values of three independent cultures ± SD.

A dose response was observed at higher concentrations of theophylline, with the estimated MIC at 20 mM (Figure 3.22A). This MIC is aligned with the 5 mM working concentration of theophylline used for induction of the riboswitch in the RiboCas

system in several other clostridia (Cañadas et al., 2019). Moreover, incubation with theophylline at or around the working concentration, 1 mM, 5 mM and 10 mM, confirmed no effect on growth rate over 24 h compared to the 0 mM control, and therefore no toxicity (Figure 3.22B). This allows the exploration of theophylline inducible riboswitches in *P. hiranonis*.

Insertion of a riboswitch

Published work has identified a selection of riboswitches optimised for use in Gram-positive bacteria. Topp et al. (2010) produced five synthetic riboswitches by *in vivo* screening and rational design that demonstrated inducible gene expression across eight diverse species, and identified their 'Riboswitch E and E*' as the most suitable for high levels of expression in Gram-positive bacteria. Furthering this work, Cañadas et al. (2019) made modifications to the sequence of Riboswitch E* based on strong clostridial RBS sequences and generated additional riboswitches, shown to be inducible in *Clostridium sporogenes*. Three of the highly expressed riboswitches, Riboswitch E, F and G, were chosen to test in *P. hiranonis* (Table 3.19).

Table 3.19 Sequences for riboswitches successfully used in Gram-positive bacteria. Riboswitches E, F and G have a constant linker and aptamer sequence but differ in the RBS sequence preceding their **translational start codon.** Riboswitch sequences developed by Cañadas et al. (2019) and Topp et al. (2010)

Ribo switch	Sequence	Source
E	CAAUACGACUCACUAUAGGUGAUACCAGCAUCGUCUUGAU GCCCUUGGCAGCACCCUGCUAAGGAGGUAACAACAAG**AUG**	(Topp et al., 2010)
F	CAAUACGACUCACUAUAGGUGAUACCAGCAUCGUCUUGAU GCCCUUGGCAGCACCCUGCUAAGGAGGUAACUUA**AUG**	(Cañadas et al., 2019)
G	CAAUACGACUCACUAUAGGUGAUACCAGCAUCGUCUUGAU GCCCUUGGCAGCACCCUGCUAAGGAGGUGUGUUA**AUG**	(Cañadas et al., 2019)

Inclusion of a riboswitch in a promoter of a new species involves insertion downstream of the transcriptional start site, adjacent to the core promoter region (-35 and -10 elements), with the 5'-UTR sequence removed. Although preliminary *in silico* characterisation of the *P. hiranonis* promoters has been carried out (3.4.1), this has not been confirmed experimentally, for example identification of the transcriptional start

site by 5'-rapid amplification of cDNA ends (5'-RACE) (Matteau, D. & Rodrigue S., 2015). This was determined to be too complex for the aims of this assay, therefore the alternative protocol by Topp et al. (2010) was followed; for the chosen *P. hiranonis* promoters, the entire native 5'-UTR was taken up until the putative RBS. This produced the putative inducible promoters with the structure outlined in Table 3.20.

Table 3.20 Structure of putative *P. hiranonis* inducible promoters. The three chosen theophylline inducible riboswitches: E, F and G, (illustrated by **Riboswitch**) will be individually inserted into each of the two native *P. hiranonis* promoters, upstream of the putative RBS.

Promoter	Structure
P_{secG}	TTTGAAAAAATGCTATTTAATAAAAAAGATAGAAAATGTCATATAATGTATGAATA TTTTAACAATTTTTGTCGCAGTAAATTATAACCAAATGTAAAATTTGTGACGTATAT GTTGTTGAACTTTTTTAACAGAAGTAGATAATTTTTGAATTAAAAATTATTATTCAA GTTTATAATTTGTTGAAAAATAGCGAAATATATGTTACACTAAAATGTAAGAATATA TTCGATAAAC**RIBOSWITCH**AUG
P_{pgk}	AATATAAAACTTGAATACATTTAAACCTTAATATGATGTATGTACCCCCATACATAC AGAGTTATTAATGTACTTCAGACGAAAAAAGTCCTAGTTTGTAGTTTTGTTACTGCG GACTCGGGCTTTTTTTGTAAATACGGAAAATAGAAATAATGTACATAAAATATAAA TAATATGGTATATTTATAATAGGGAAGTGTGATTAAGTCATACTTTTAAAAAATACT TCAAATAAA**RIBOSWITCH**AUG

To assess the hypothesised inducibility of the new promoter/riboswitch combinations in *P. hiranonis*, FAST reporter plasmids were constructed using hifi PCR. Fragments for the promoter, riboswitch and FAST were generated using the primers in Table 3.21 and inserted into the backbone pMTL84151, digested with SalI and AatII. Insertions were confirmed by colony PCR (ColE1_tra_F2 and pCD6_R1), followed by Sanger sequencing. Final plasmids were then transformed into *E. coli* CA434.

Table 3.21 Construction of *P. hiranonis* inducible promoter reporter plasmids. The two native promoters, P_{secG} and P_{pgk} were inserted into pMTL84151 with one of the three chosen riboswitches, E, F or G, along with FAST, for assessment in a reporter assay. RB=riboswitch

Final Plasmid	Fragment	Primers	Template
pMTL84151_PsecG_RBE	P_{secG}	SecGRB_prom_F1 SecGRB_prom_R1	*P. hiranonis* DNA
	RBE	SecGRBE_RBE_F1 SecGRBE_RBE_R1	ML_Gapdh_RBE
	FAST	SecGRBE_FAST_F1 PromRB_FAST_R1	pMTL8315_ptcdB_FAST

pMTL84151_PsecG_RBF	P_{secG}	SecGRB_prom_F1 SecGRB_prom_R1	*P. hiranonis* DNA
	RBF	SecGRBF_RBF_F1 SecGRBF_RBF_R1	ML_Gapdh_RBF
	FAST	SecGRBF_FAST_F1 PromRB_FAST_R1	pMTL8315_ptcdB_FAST
pMTL84151_PsecG_RBG	P_{secG}	SecGRB_prom_F1 SecGRB_prom_R1	*P. hiranonis* DNA
	RBG	SecGRBG_RBG_F1 SecGRBG_RBG_R1	ML_Gapdh_RBG
	FAST	SecGRBG_FAST_F1 PromRB_FAST_R1	pMTL8315_ptcdB_FAST
pMTL84151_Ppgk_RBE	P_{pgk}	PgkRB_prom_F1 PgkRB_prom_R1	*P. hiranonis* DNA
	RBE	PgkRBE_RBE_F1 PgkRBE_RBE_R1	ML_Gapdh_RBE
	FAST	PgkRBE_FAST_F1 PromRB_FAST_R1	pMTL8315_ptcdB_FAST
pMTL84151_Ppgk_RBF	P_{pgk}	PgkRB_prom_F1 PgkRB_prom_R1	*P. hiranonis* DNA
	RBF	PgkRBF_RBF_F1 PgkRBF_RBF_R1	ML_Gapdh_RBF
	FAST	PgkRBF_FAST_F1 PromRB_FAST_R1	pMTL8315_ptcdB_FAST
pMTL84151_Ppgk_RBG	P_{pgk}	PgkRB_prom_F1 PgkRB_prom_R1	*P. hiranonis* DNA
	RBG	PpgkRBG_RBG_F1 PpgkRBG_RBG_R1	ML_Gapdh_RBG
	FAST	PpgkRBG_FAST_F1 PromRB_FAST_R1	pMTL8315_ptcdB_FAST

Assessment of new inducible promoters

The six reporter plasmids (pMTL84151_PsecG_RBE, pMTL84151_PsecG_RBF, pMTL84151_PsecG_RBG, pMTL84151_Ppgk_RBE, pMTL84151_Ppgk_RBF, pMTL84151_Ppgk_RBG) were conjugated into *P. hiranonis*, allowing 24 h for conjugation. Transconjugants were selected for with thiamphenicol and re-streaked to confirm presence of the plasmid, to generate strains for the FAST reporter assay.

As the theophylline inducer is reconstituted in DMSO, a preliminary FAST assay was first carried out to ensure that the DMSO vehicle did not interfere with the fluorescence signal, and that inducibility is a result of theophylline only (Figure 3.23). There has been anecdotal evidence within the research group of false positives from DMSO (Matthew Lau, personal communications), so the risk required eliminating. The native P_{secG} was compared to the P_{secG} with Riboswitch E (RBE), at three concentrations of theophylline in DMSO, or with the same volume of vehicle (water or DMSO). It was not possible to use 10 mM theophylline in water due to solubility.

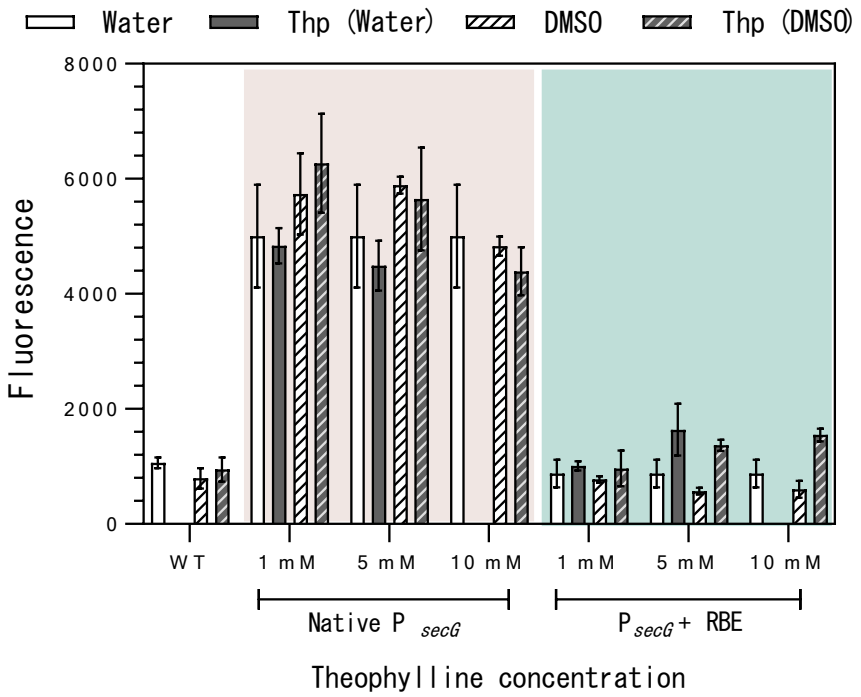


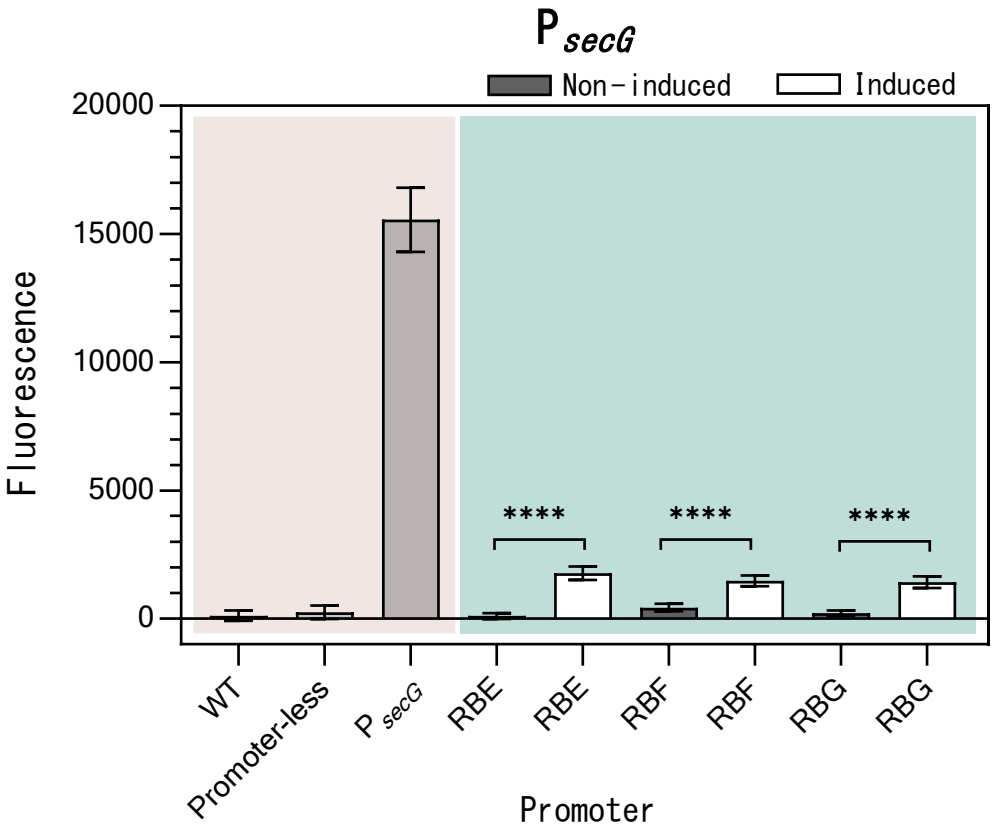
Figure 3.23 FAST assay to assess DMSO vehicle control. *P. hiranonis* harbouring pMTL84151_FAST_secG or pMTL84151_PsecG_RBE was cultured in BHISS liquid medium + thiamphenicol. After 4 h theophylline was added at 1, 5 or 10 mM, dissolved in either water or DMSO. Equivalent volumes of the water or DMSO were also added. At 7 h samples were washed and the TF-Amber ligand added. Fluorescence was measured immediately at an excitation wavelength of 499 nm and emission at 558 nm using a microplate reader. Data represent mean values of one culture in three technical triplicates ± SD.

When studying the native P_{secG} there should be no increase in fluorescence on theophylline addition as there is no riboswitch. There is no significant difference between any of the variables (theophylline in water, theophylline in DMSO, water, DMSO) at any concentration (One-way ANOVA with Tukey's test for multiple comparisons at each concentration, all p-values>0.05). This eliminates the risk of false positives when there is no level of induction.

When studying a functioning riboswitch promoter there should be an increase in fluorescence on theophylline addition, whether dissolved in water or DMSO. There should be no difference in signal depending on which vehicle is used, for the control or theophylline addition. For theophylline in water there is no significant difference

between the vehicle control and 1 mM, but there is at 5mM (unpaired two-tailed Student's t-test, p=0.4623 and p=0.0430 respectively). For theophylline in DMSO there is no significant difference at 1 mM, but there is at 5 mM and 10 mM (unpaired two-tailed Student's t-test, p=0.3612, p=0.0003 and p=0.0008 respectively). There is also no significant difference between fluorescence between the vehicle controls at either 1 mM or 5mM, or between theophylline in water or DMSO at either 1 mM or 5 mM (unpaired two-tailed Student's t-tests).

This preliminary assay confirms that the DMSO vehicle of theophylline does not impact on the fluorescence signal. The FAST assay was therefore carried out to study the inducibility of P_{secG} and P_{secG} with the three riboswitches (Figure 3.24), using theophylline at its usual working concentration of 5 mM in DMSO.


P secG
Non-induced
Induced
20000
15000
10000
5000
0
Fluorescence
****
****
****
WT
Promoter-less
P secG
RBE
RBE
RBF
RBF
RBG
RBG
Promoter

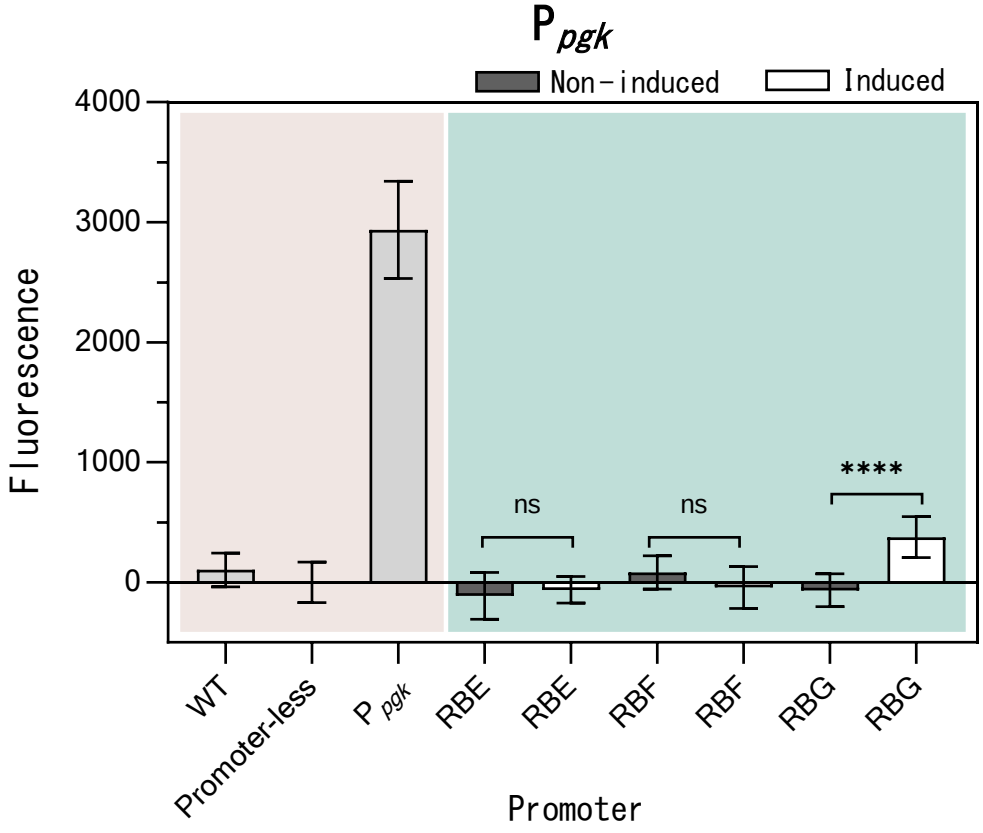

P pgk
Non-induced
Induced
4000
3000
2000
1000
0
Fluorescence
ns
ns
****
WT
Promoter-less
P pgk
RBE
RBE
RBF
RBF
RBG
RBG
Promoter

Figure 3.24 FAST reporter assay of theophylline inducible P_{secG} and P_{pgk}. *P. hiranonis* harbouring the following reporter plasmids was cultured in BHISS liquid medium + thiamphenicol: pMTL84151_PsecG_RBE, pMTL84151_PsecG_RBF, pMTL84151_PsecG_RBG, pMTL84151_Ppgk_RBE, pMTL84151_Ppgk_RBF, pMTL84151_Ppgk_RBG. Theophylline (5 mM in DMSO) was added at 4 h and samples taken at 7 h. Statistical analysis was carried out to compare non-induced and induced values using a non-paired, two-tailed t-test. P_{secG}: RBE, RBF and RBG all $p < 0.0001$ (****). P_{pgk}: RBE = 0.5173, RBF = 0.1138, RBG $p < 0.0001$ (****). Values are blanked to BHISS liquid medium. Data represent mean values of three independent cultures in technical triplicate ± SD.

The introduction of riboswitches into P_{secG} and P_{pgk} has afforded inducibility in both cases. All three riboswitches introduced to P_{secG} allowed induction with theophylline, as there was a significant difference between the fluorescence signal with and without the inducer (unpaired two-tailed Student's t-test, $p < 0.0001$ for RBE, RBF and RBG). However, only RBG allowed induction for P_{pgk} ($p < 0.0001$) as there was no significant difference for RBE and RBF ($p=0.5173$ and $p=0.1138$ respectively). The RBS sequence is the only differential between the riboswitches, but it is unlikely to be responsible for the failure of RBE and RBF; the RBS is organism dependent and therefore RBE and RBF should not only function in P_{secG}.

The induction of RBE, RBF and RBG in P_{secG} with 5 mM theophylline gave a range of activation ratios, that is average expression levels with theophylline/average expression levels without theophylline, similar to results in other clostridia (Cañadas et al., 2019). In *P. hiranonis* for P_{secG} the ratios were 16, 3.5 and 6.7 respectively, compared to 12, 13.5 and 16 in *C. sporogenes*. A range of results was also observed for P_{pgk}, albeit not all riboswitches were inducible. Unfortunately, the activation ratio for P_{pgk} RBG could not be calculated due to negative fluorescence values from subtracting the blank values.

The constitutive P_{pgk} was shown to be a weaker promoter when directly compared to P_{secG} in the initial FAST assay (Figure 3.15), demonstrating a 2.5-fold lower fluorescence signal, and this was shown again in this assay. Both promoters demonstrated a significant reduction in protein expression after riboswitch insertion,

and this is not unusual according to results by Topp et al. (2010). However, it is possible that the lower baseline of expression of P_{pgk} prevented adequate expression following riboswitch insertion. The promoters themselves must also be considered, as the influence of promoter elements can affect riboswitch inducibility, as shown by Cañadas et al. (2019) and Topp et al. (2010), and therefore the positioning of the riboswitches downstream of these elements may have had an influence; appropriate positioning was only estimated. It is also possible that the FAST assay is not sensitive enough at lower fluorescence levels, as evidenced by the error bars throughout the assay as well as the negative values after blank subtraction, to detect a small increase in expression.

From these results, P_{secG}RBE was chosen to further characterise its dynamic expression, as it produced the highest activation ratio. The FAST assay was repeated using a range of theophylline concentrations to determine the dose response (Figure 3.25).

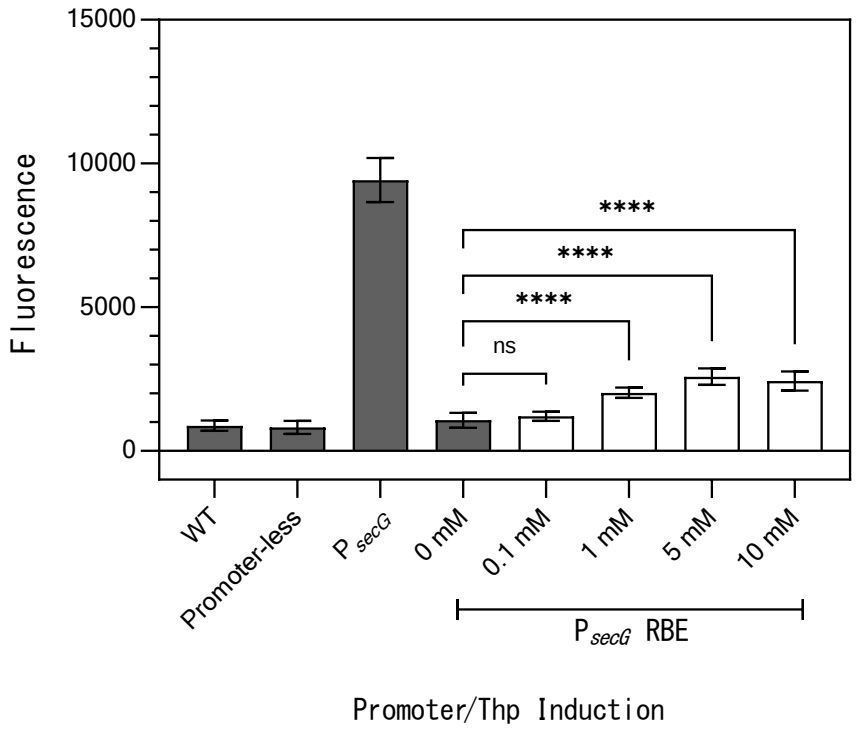


Figure 3.25 P_{secG}RBE dose response to theophylline induction. *P. hiranonis* harbouring the pMTL84151_PsecG_RBE reporter plasmids was cultured in BHISS liquid medium + thiamphenicol. Theophylline at varying concentrations was added at 4 h and samples taken at 7 h. Statistical analysis was carried out using one-way ANOVA with Dunnett's test for multiple comparisons against the non-induced (0 mM); p-values are indicated as non-significant (ns), or p<0.0001 (****). Values are blanked to BHISS liquid medium. Data represent mean values of three independent cultures in technical triplicate ± SD.

Induction of P_{secG}RBE with increasing concentrations of theophylline demonstrated a dose response up until 10 mM; the activation ratios were 1.8, 2.4 and 2.3 for 1 mM, 5 mM and 10 mM respectively. There was no significant difference between non-induced and 0.1 mM (p=0.5642), which may either be due to the sensitivity of the riboswitch, or the assay. The activation ratio for 5 mM of 2.4 was comparably lower than 16 in the same conditions in Figure 3.24. This is likely due to a combination of overall lower fluorescence values, for example P_{secG} at 9425 vs 15559, but also higher background levels, shown in the WT and promoter-less controls, and in 0 mM. The level of fluorescence can vary between each assay depending on the parameters used on the microplate reader, although this was kept consistent throughout. Moreover, this

should not affect the activation ratio, and again highlights the possible influence of assay sensitivity on results. Regardless, this assay confirms that induction of P_{secG}RBE is as a result of theophylline addition, and that this inducible promoter can provide a range of expression levels. This widens the scope for its applications in genetic engineering in *P. hiranonis*.

3.4.4 Use of inducible native promoters in RiboCas

Construction and conjugation of inducible RiboCas

With P_{secG}RBE having the highest activation ratio and having confirmed a dose response with theophylline, this inducible promoter was chosen to drive the expression of *cas9* in the modified *P. hiranonis* RiboCas system in a final attempt to allow genetic modification. To avoid a duplicate promoter and possible recombination of the plasmid, the P_{secG} driving the sgRNA was also replaced with P_{pgk} (Figure 3.26). This new backbone, pCHRE1_RB, was generated by hifi assembly of the fragments listed in Table 3.22, inserted into the backbone generated by digestion of pRECas1_CD6_MCS with NotI and SalI. The final plasmid was confirmed by two colony PCRs (Cas9scr1 and ColE1_tra_F2, Cas9scr1R and pCD6_R1), followed by Sanger sequencing.

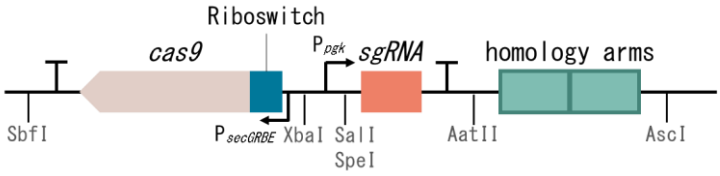


Figure 3.26 Application specific module of the modified RiboCas system, pCHRE1_RB. The promoters for *cas9* and the sgRNA module in pCHRE1 have been exchanged to include a riboswitch; P_{secG}RBE for *cas9* and P_{pgk} for the sgRNA. The generates the new RiboCas backbone pCHRE1_RB. The other components of the application specific module and plasmid backbone remain the same as pCHRE1_RB.

Table 3.22 Fragments used to generate pCHRE1_RB. Primers used to generate the RiboCas backbone with P_{secG}RBE and P_{pgk} promoters for *cas9* and the sgRNA. Assembled by hifi PCR.

Fragment	Length (bp)	Primers	Template
cas9	4109	pCHRB_cas9_F1 pCHRB_cas9_R1	pRECas1_MCS
P_{secG}RBE	314	pCHRB_sec_R1 pCHRB_Sec_F1	pMTL84151_PsecG_RBE
P_{pgk}	255	pCHRB_pgK_F1 pCHRB_pgK_R1	*P. hiranonis* DNA

The pCHRE1_RB backbone was then used to generate a sample of RiboCas vectors for *bai* gene knockouts. These were generated for *baiA, baiCD* and *baiF* by digestion of pRECas1_MCS_A, pRECas1_MCS_CD and pRECas1_MCS_F with SalI and AscI to generate the sgRNA/homology arm editing cassette, for insertion by ligation into the MCS of pCHRE1_RB digested with SalI and AscI. This generated pCHRB_A, pCHRB_CD and pCHRB_F. Final plasmids were confirmed by colony PCR (Cas9scr1R and pCD6_R1), followed by Sanger sequencing. The three plasmids were then transformed into *E. coli* CA434.

In addition to further attempts at *bai* gene deletion, a KO vector was also generated for *pyrE*, to eliminate the possibility that *bai* operon targeting was the issue. The *pyrE* mutation creates an uracil auxotroph and has been successfully used as a selection marker in clostridia (Ng et al., 2013), thus making it a reliable target in *P. hiranonis*. The vector pCHRB_PE was generated by hifi assembly of the fragments listed in Table 3.23, assembled into the backbone pCHRE1_RB digested with SalI and AscI. Insertion was confirmed by colony PCR (Cas9scr1R and pCD6_R1), followed by Sanger sequencing. The final plasmid was then transformed into *E. coli* CA434.

Table 3.23 Fragments used to generate pCHRB_PE. Primers used to generate the RiboCas pCHRE1_RB with application specific elements for a *pyrE* deletion. Assembled by hifi PCR. LHA = Left homology arm. RHA = Right homology arm.

Fragment	Primers	Template
sgRNA (SEED = CTTGGAAGCGAACTTCAGGA)	pCHRB_PE_sg_F1 pCHRB_PE_sg_R1	pCHRB_CD
LHA (775 bp)	pCHRB_PE_LHA_F1 pCHRB_PE_LHA_R1	*P. hiranonis* DNA
RHA (743 bp)	pCHRB_PE_RHA_F1 pCHRB_PE_RHA_R1	*P. hiranonis* DNA

Attempts at gene knockouts

All four KO plasmids based on the inducible RiboCas system (pCHRB_A, pCHRB_CD, pCHRB_F, pCHRB_PE) were conjugated into *P. hiranonis*, selecting for transconjugants with thiamphenicol. Transconjugants were then confirmed by re-streaking and colonies were plated onto 5 mM theophylline for RiboCas induction.

Colony PCR was carried out to screen for deletions across the four genes, with primers used as follows: *baiA*, Ch_scr_F5 and Ch_scr_R4; *baiCD*, Ch_scr_F3 and Ch_scr_R2; *baiF*, Ch_scr_F6 and Ch_scr_R5; *pyrE*, CHIR_PEscr_F1 and CHIR_PEscr_R1. Unfortunately, as seen previously, all genotypes for *baiA*, *baiF* and *pyrE* were WT in size. This confirms that the *bai* operon itself is not responsible for an inability to generate knockouts, and is likely to be the system itself. For *baiCD* a mixed population was identified but, despite repeated restreaking and colony PCR, the KO could not be isolated.

Following the attempts to use inducible promoters in the RiboCas system, it was felt that most avenues available to troubleshoot the inability to develop genetic tools and generate *bai* gene knockouts in *P. hiranonis* was rationally exhausted. It was therefore decided to focus on mechanisms of colonisation resistance by *P. hiranonis* against *C. difficile* for the remainder of this chapter.

3.5 The effect of 7α-dehydroxylation on competition with *Clostridioides difficile*

Whilst unable to establish a reproducible CRISPR/Cas system for mutant generation in *P. hiranonis*, the work undertaken has improved the characterisation and understanding of this lesser known organism, and this will be harnessed to investigate its mechanisms of 7α-dehydroxylation. The production of a *ΔbaiCD* strain in 3.3.4 provides an opportunity to study the impact of 7α-dehydroxylation on resistance against *C. difficile*, and could allow insights into the debated mechanisms that allow this resistance.

3.5.1 The barAB regulatory region

The expression of *bai* genes was discovered to be bile acid inducible as early as 1990 (Darrell H Mallonee & White, 1990), and since then work has been carried out to understand the regulatory elements involved. Early work demonstrated conserved regions in the promoters upstream of *baiB* in 7α-dehydroxylating organisms such as *C. scindens* VPI 12708, and has been extended to *P. hiranonis* and *C. hylemonae* (Darrell H Mallonee & White, 1990; Ridlon et al., 2010; Wells & Hylemon, 2000). The identification of regulatory genes followed, with Ridlon, Kang, & Hylemon (2006) identifying the *barB* and *barA* gene in *C. scindens*, upstream of the start of the operon on the opposite strand (Table 3.24). These genes have also been found in *P. hiranonis* and *C. hylemonae* (Ridlon et al., 2020).

Table 3.24 BarA and BarB identified in *C. scindens* by Ridlon et al. (2006)

Protein	Molecular mass (kDa)	Catalytic activity/function	Gene family
BarA	46	Transcriptional regulation	AraC/XylS
BarB	22	Transcriptional regulation	RpoB; permeases of the major facilitator superfamily

Several studies have demonstrated the upregulation of *bai* genes with CA but no upregulation or downregulation with DCA in *C. scindens*, *P. hiranonis* and *C. hylemonae* (Devendran et al., 2019; Reed et al., 2020; Ridlon et al., 2020). However, it is not clear from these studies on the involvement of BarA and BarB, or the promoter region, in this CA regulation. Whilst knockout mutations of the regulatory region would be most beneficial, tools are not available to do this. Therefore, it was decided that as a functional reporter assay has now been validated in *P. hiranonis* as part of

this report, the FAST assay would be used to study the regulatory region and its induction.

Comparison with C. scindens

To further understanding of the *bai* regulatory region, particularly for *P. hiranonis*, BProm was used to predict the promoter elements, and these were compared to previously published predictions, and to *C. scindens* VPI 12708. Alignments were carried out using ClustalW for the 160 bp region upstream of *baiB* and 160 bp upstream of *barA* (opposite strand) and annotated with various promoter elements for comparison (Figure 3.27).

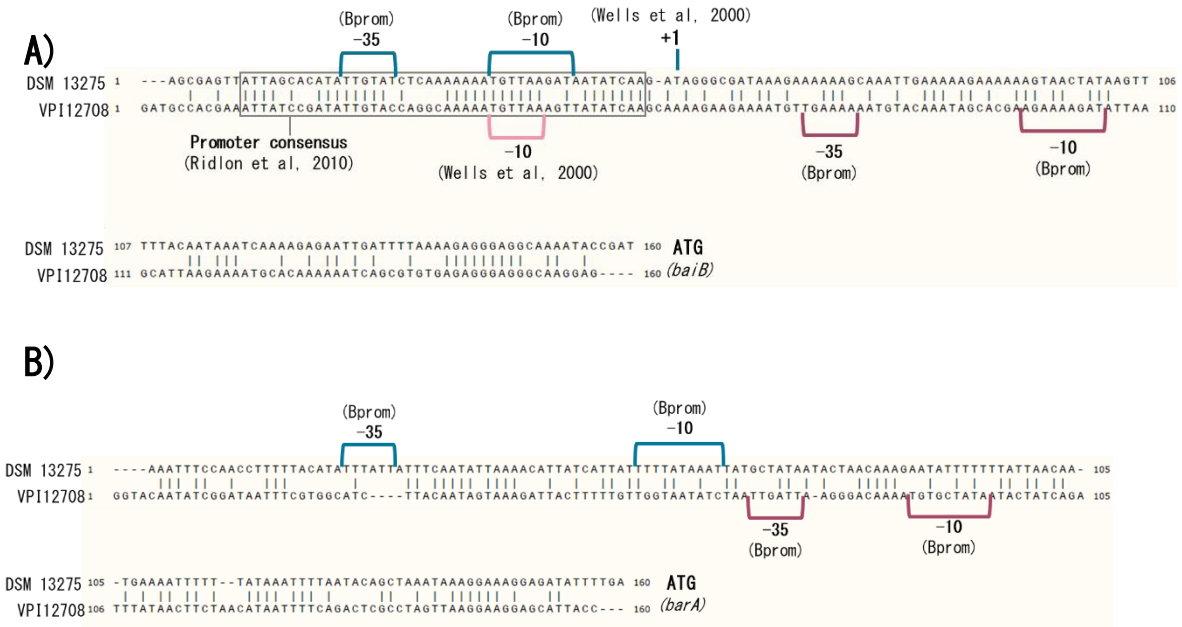


Figure 3.27 Comparison of the putative promoter sequences of *baiB* (A) and *barA* (B) in *C. scindens* VPI 12708 and *P. hiranonis* DSM 13275. Predictions from BProm. *C. scindens*, *baiB*: ttgaaa (-35) and agaaaagat (-10). *C. scindens*, *barA*: ttgatt (-35) and tgctataat (-10). *P. hiranonis*, *baiB*: ttgtat (-35) and tgttaagat (-10). *P. hiranonis*, *barA*: tttatt (-35) and ttttataaa (-10). Consensus sequence for *baiB* promoter suggested by Ridlon et al. (2010): ATxTxxtaxcxxxxxxAAxTGTTAAxxTtaTATCAA. Sequence alignment by ClustalW.

A comparison of the *baiB* upstream region (Figure 3.27A) for *P. hiranonis* and *C. scindens* demonstrated a 53.1% homology overall. *P. hiranonis* contains the predicted

promoter consensus sequence suggested by Ridlon et al. (2010), ATxTxxtaxcxxxxxxAAxTGTTAAxxTtaTATCAA, with one nucleotide difference (highlighted), whereas *C. scindens* is an exact match. The -35 and -10 elements predicted by BProm for *P. hiranonis* lie within this region, increasing confidence in the predictions, and the -10 element aligns with that suggested for *C. scindens* by Wells & Hylemon (2000). The predicted promoter elements for *P. hiranonis* also align well with the predicted TSS by the same authors. The BProm predictions for *C. scindens* do not align with the consensus sequence or with *P. hiranonis*, questioning their validity.

Fewer studies have been made for the promoter of *barA*, with no previous consensus region identified. The 160 bp upstream region has a lower homology than for that of *baiB*, at 46.4% (Figure 3.27B) and there is no consensus for the -35 and -10 element sequences, or position in relation to the start codon. This could suggest that transcriptional control of *barA* is less important than for *baiB*, or that the confidence in predictions by BProm is low.

Following comparisons of the promoter regions, the BarA and BarB protein sequences were compared for *P. hiranonis* and *C. scindens*. BLASTP was used to compare identity levels of *C. scindens* ATCC 35704 and *P. hiranonis* DSM 13275 with that of *C. scindens* VPI 12708 (Table 3.25). Comparing the two strains of *C. scindens*, both proteins had a high % identity over 95%, with low E values. Similarities were lower for *P. hiranonis*, at 65% for BarA and 58% for BarB, with low E values.

Table 3.25 Comparison of BarA and BarB in C. scindens VPI 12708. BLASTP comparison to the same protein in *C. scindens* ATCC 35704 and *P. hiranonis* DSM 13275.

Strain	% identity	E value	Query cover	Accession
		BarA		
C. scindens ATCC	95.6	0	99%	WP_004607866.1
P. hiranonis DSM	64.9	0	98%	WP_006438993.1
		BarB		
C. scindens ATCC	97.8	2e-132	99%	WP_004607867.1
P. hiranonis DSM	57.9	8e-71	96%	WP_006438992.

Following sequence comparisons, protein domains were predicted for BarA and BarB in *P. hiranonis* and *C. scindens* VPI 12708. BarB did not yield any domain hits for either organism, but the BarA protein demonstrates a pair of helix-turn-helix domains of the AraC transcription factor family. The results were identical between *P. hiranonis* and *C. scindens.* Results for *P. hiranonis* are shown in Figure 3.28.

A)

Name	Accession	Description	Interval	E-value
AraC	COG2207	AraC-type DNA-binding domain and AraC-containing proteins	190-290	5.2e-19
HTH_ARAC	Smart00342	Helix_turn_helix, arabinose operon control protein	201-284	5.8e-16
HTH_ARAC	Smart00342	Helix_turn_helix, arabinose operon control protein	320-402	9.8e-16
AraC	COG2207	AraC-type DNA-binding domain and AraC-containing proteins	314-406	4.4e-15

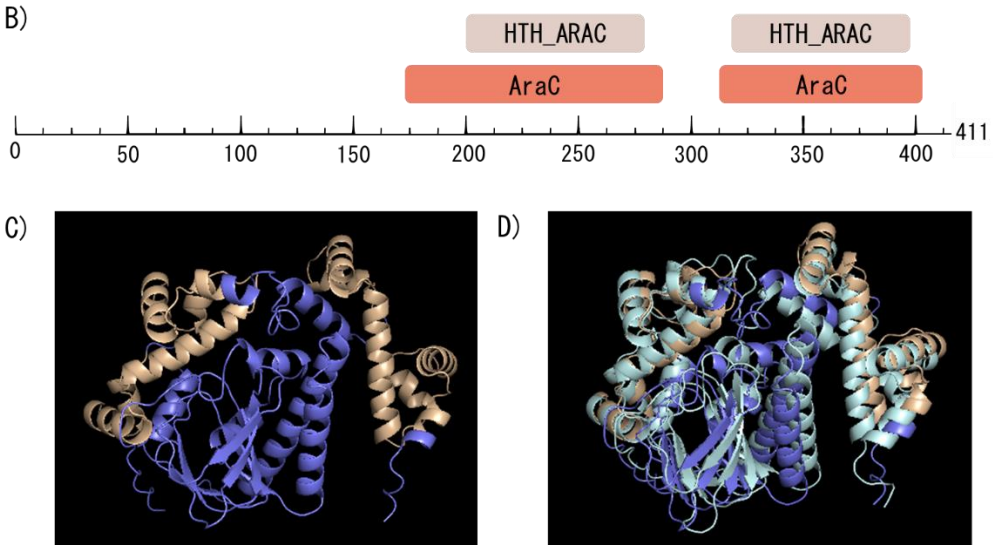


Figure 3.28 Predicted structures of the *P. hiranonis* BarA protein. A) BLASTP domain hit results. B) BLASTP domain hit illustration. C) Alphafold predicted 3D structure (dark blue). Predicted HTH_AraC domains highlighted in beige. D) Overlap of BarA from *P. hiranonis* and *C. scindens* VPI 12708 (light blue). Predictions were also attempted for BarB but yielded no domain hits.

The BarA protein for *P. hiranonis* contains two helix-turn-helix (HTH) motifs, identifying it as an AraC family protein, and there is a high similarity of the 3D predicted structure to that of *C. scindens* (Figure 3.28D). The protein architecture demonstrated in Figure 3.28A-D, is typical of the family; a HTH motif towards the

end of the approximately 300 amino acid long protein, with an additional HTH motif. The HTH motifs allow interactions with DNA, with most of the family facilitating DNA looping that allows transcriptional activation (Egan, 2002; Gallegos, Schleif, Bairoch, Hofmann, & Ramos, 1997). Many of the AraC proteins demonstrate a ligand response, through binding of their non-HTH domain, and this could explain the function of BarA; binding to a bile acid, likely CA, to induce transcriptional activation of the *baiB* promoter (Egan, 2002). Previous work from our group has shown that these genes are constitutively expressed in *C. scindens* even with CA induction (A. Dempster 2017, unpublished). The high sequence homology of the promoter and transcription factor in *P. hiranonis* and *C. scindens* further suggests the importance of these regions and presence of similar regulation. With no identified protein domains for BarB it is not possible to suggest its role in this process, but its presence in most major 7α-dehydroxylators suggests it is worthy of further study.

Construction of barAB reporter plasmid

To assess the induction and regulation of the putative *bai* operon promoter, a FAST reporter plasmid was constructed. The region including *barB, barA* and the intergenic region downstream of *barA* ending at the ATG of *baiB,* was assessed. The plasmid was generated using hifi assembly, with the barAB fragment from *P. hiranonis* (84bar_bar_F1 and 84bar_bar_R1) and FAST (84bar_FAST_F1 and 84bar_FAST_R1) inserted into pMTL84151 digested with SalI and XhoI. The final plasmid, pMTL84151_barAB_FAST, was confirmed by colony PCR (ColE1_tra_F2 and pCD6_R1) and Sanger sequencing.

Promoter induction with bile acids

The pMTL84151_barAB_FAST plasmid was conjugated into *P. hiranonis*, selecting with thiamphenicol and confirming through re-streaking on selection. The final strain was used to carry out a FAST reporter assay to study bile acid induction with cholic acid and deoxycholic acid (Figure 3.29). The overnight culture of the strain was sub-cultured into 10 mL of cholic acid (CA) or deoxycholic acid (DCA) reconstituted in BHISS, including thiamphenicol selection, at the given concentration, and samples were taken after 7 h of growth.

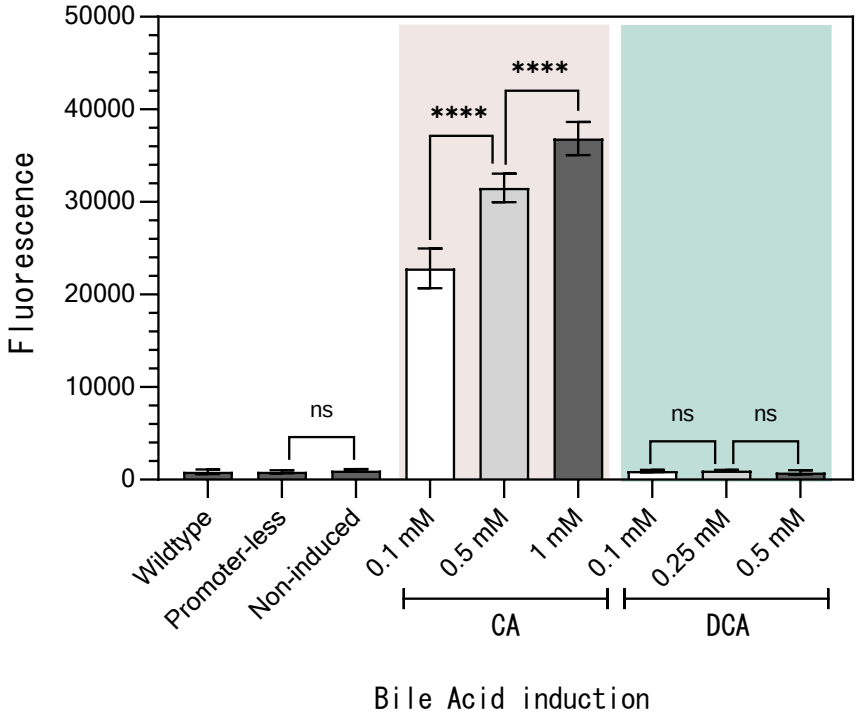


Figure 3.29 FAST reporter assay of bile acid induction of the *bai* operon promoter and regulatory *barAB* region. *P. hiranonis* harbouring the pMTL84151_barAB_FAST reporter plasmid was cultured in BHISS liquid medium + thiamphenicol + cholic acid (CA) or deoxycholic acid (DCA) at varying concentrations. Samples were taken at 7 h and fluorescence measured on addition of TF-Amber ligand. Statistical analysis was carried out by one-way ANOVA with Tukey's test for multiple comparisons, p-values are indicated as non-significant (ns), or p<0.0001 (****). Values are blanked to BHISS liquid medium. Data represent mean values of three independent cultures in technical triplicate ± SD.

From the FAST assay there was a clear induction of the *bai* operon promoter with CA but not with DCA. CA induced fluorescence even at the lowest concentration of 0.1 mM, and increased at both concentrations of 0.5 mM and 1 mM (One-way ANOVA with Tukey's test for multiple comparisons, p<0.0001 for both comparisons), with a 1.5-fold increase in fluorescence from 0.1 mM to 1 mM. There was no difference in fluorescence between non-induced and any of the three DCA concentrations (0.1 mM p>0.9999, 0.25 mM p>0.9999, 5 mM p=0.9997). There was no expression in the non-induced state, with no significant difference in fluorescence to that of the promoter-less control (p>0.9999).

These results agree with the hypothesised function that BarA binds to CA and activates the *baiB* promoter, and are consistent with previous transcriptomic analyses in *P. hiranonis* and *C. scindens* that show induction of expression with CA. Ridlon et al. (2020) showed significant upregulation of the expression of all 7 major *bai* genes in *P. hiranonis* by CA at 100 µM, ranging from a 177-fold increase for *baiH* to 1416-fold increase for *baiE*. Significant upregulation with CA (50 µM) was also observed in *C. scindens* ATCC 35704 by Devendran et al. (2019). If tools are ever available, KO strains of *barA* and the operon promoter in *P. hiranonis* would further confirm that the BarA/CA interaction is responsible for *bai* gene induction.

The lack of induction with DCA suggests that the 7α-hydroxyl group could play a role in ligand binding and BarA activation. However, the data by both Ridlon et al. (2020) and Devendran et al. (2019) suggest a downregulation of the seven *bai* genes with DCA compared to non-induced when measuring transcripts per kilobase million. This difference could be due to assay methodologies as it was not possible to detect a decrease in FAST expression in this assay as there was no difference between the non-induced and the promoter-less control. Ridlon et al. (2020) have suggested that there could be a feedback loop during the accumulation of DCA. The availability of the *ΔbaiCD P. hiranonis* mutant would allow this to be investigated by focussing on induction by CA alone without conversion to DCA. In the WT strain DCA could also be added after CA addition and FAST expression measured.

Future work to obtain a crystal structure or suggested binding domain could allow the prediction of BarA interactions with the binding BA and confirm this hypothesis. In the meantime this study could be expanded to look at intermediate BA structures to gain potential structure/function analysis data between different functional groups and promoter activation. Previous work in *C. scindens* and *C. hylemonae* has suggested that substrates with either the cis (5b-bile acids) or trans (5a-bile acids) orientations about the A/B ring junction are able to induce the operon (Ridlon et al., 2010), so it would be interesting to expand these studies to the *P. hiranonis bai* operon promoter.

This study can only demonstrate gene regulation at the transcriptional level for the *bai* promoter, in combination with the one RBS of *baiB*. Whilst this has illustrated transcriptional activation of the promoter, to fully understand the regulation of *bai* gene expression, translational control should also be scrutinised. Whilst the

transcriptomic data currently illustrates differing expression levels and increases/decreases in expression with CA and DCA (Devendran et al., 2019; Ridlon et al., 2020), further data, such as RBS studies, could help explain these differences and deepen understanding of *bai* gene regulation as a whole. There is currently no work studying the RBS's for the individual *bai* genes, and they are difficult to identify and characterise as they are within upstream coding regions.

3.5.2 Assessment of the ΔbaiCD *strain*

One of the main aims of this project was to generate a strain of *P. hiranonis* that is incapable of 7α-dehydroxylation, through the deletion of one or several *bai* genes. Whilst reliable genetic tools were not able to be developed to reproducibly generate KOs, a single *ΔbaiCD* strain was generated. BaiCD has been identified as a NAD+-dependent 3-oxo-Δ4- cholenoic acid oxidoreductase for 7α-hydroxy bile acids (D. J. Kang et al., 2008), acting at two steps of the pathway; it is thought to convert 3-oxo-cholyl-CoA to 3-oxo-4,5-dehydro-cholyl-CoA, and 3-oxo-4,5-dehydro-deoxycholic acid to 3-oxo-deoxycholic acid (Funabashi et al., 2020). It is hypothesised that deletion of *baiCD* would remove 7α-dehydroxylation and the conversion of CA to DCA, and this was therefore assessed in the *ΔbaiCD* strain.

Whole Genome Sequencing

Whole genome sequencing was carried out for the *ΔbaiCD* strain and analysed (2.6.5) to identify any off-target SNPs or indels (Table 3.26).

Table 3.26 Resequencing analysis of *P. hiranonis ΔbaiCD* strain. SNPs and other genomic variants listed. Gene and function identified according to annotations of the reference (Ref.) genome (NZ_CP036523.1) (IG=intergenic). For variants, A=sequence at locus, C=coverage of reads, F=frequency in reads. Amino acid changes arising from variants are listed (AA). Blank cells denote no change from reference sequence.

				***ΔbaiCD* strain**			
Gene (KGNDJEFE)	**Function**	**Position**	**Ref.**	**A**	**C**	**F**	**AA**
RS00265	tuf, elongation factor Tu	38537	G	**A**	243	100	Tyr201Tyr
RS00815	23S ribosomal RNA	118529	C	**A**	17	100	Asn70Lys
RS11860	hypothetical protein	160482	CACTA	-	215	96.28	Pro102FS
IG		257528	C	**T**	28	96.43	IG
IG		257697	C	**G**	45	95.56	IG
IG		257699	G	**A**	43	95.35	IG
IG		290272	-	**C**	40	100	IG

IG		291754	-	**T**	103	100	IG
RS02065	HAMP sensor histidine kinase	365624	A	**G**	162	100	Lys371Arg
RS04415	sensor histidine kinase	940864	T	**C**	146	100	Ile533Thr
RS11715	ATP-binding protein	1031995	C	**T**	138	100	Gly379Glu
IG		1033041	T	**C**	108	100	IG
RS05430	ABC transporter ATP-binding protein	1172909	A	**G**	138	100	Phe146Phe
IG		1411488	C	**T**	130	100	IG
RS06950	hypothetical protein	1471447	TATGTAGTA	**-**	125	82.4	Tyr46FS
RS07105	IS256 family transposase	1502740	A	**G**	210	99.52	Val168Val
RS07170	hypothetical protein	1516088	ATTGATG	**-**	174	76.44	Asn332FS
RS07320	oligoribonuclease/PAP phosphatase nrnA	1554336	-	**GC**	161	99.38	Ile78FS
RS07320	oligoribonuclease/PAP phosphatase nrnA	1554337	A	**CT**	164	99.39	Thr77FS
IG		1625392	C	**-**	343	94.17	IG
RS09725	hemolysin family protein	2076305	A	**G**	182	100	Phe297Leu
RS10410	MerR transcriptional regulator	2237283	G	**A**	147	100	Gly55Gly
RS11235	nucleoside transporter	2428142	A	**G**	192	100	Ile342Thr

A total of 23 variants were identified with SNPs, insertions and deletions throughout the whole genome. Of the 23 variants, 15 were within coding regions and consist of 4 synonymous, 6 missense and 5 frameshift mutations. Unfortunately, data was not available to analyse SNPs present in the parental WT strain so SNPs unique to the *ΔbaiCD* strain could not be identified. This would isolate SNPs that may be having an impact when comparing WT and *ΔbaiCD* strains in subsequent assays. Fortunately there are no SNPs in genes pertaining to bile acid metabolism, allowing an assessment of *ΔbaiCD*. However, if unique SNPs are present these need to be considered in future assays comparing the strains. In an area of possible impact for this study, the overnight growth was assessed and both strains reached a similar OD_{600}, suggesting minimal impact of SNPs on growth.

7α-dehydroxylation

Both the WT and *ΔbaiCD* strains of *P. hiranonis* were incubated in BHISS with CA at 100 µM for 24 hours, and the concentration of DCA, plus intermediates, was measured (2.3.9) (Figure 3.30). Bile acid analyses were carried out by Eduard Vico Oton (École Polytechnique Fédérale de Lausanne (EPFL)).

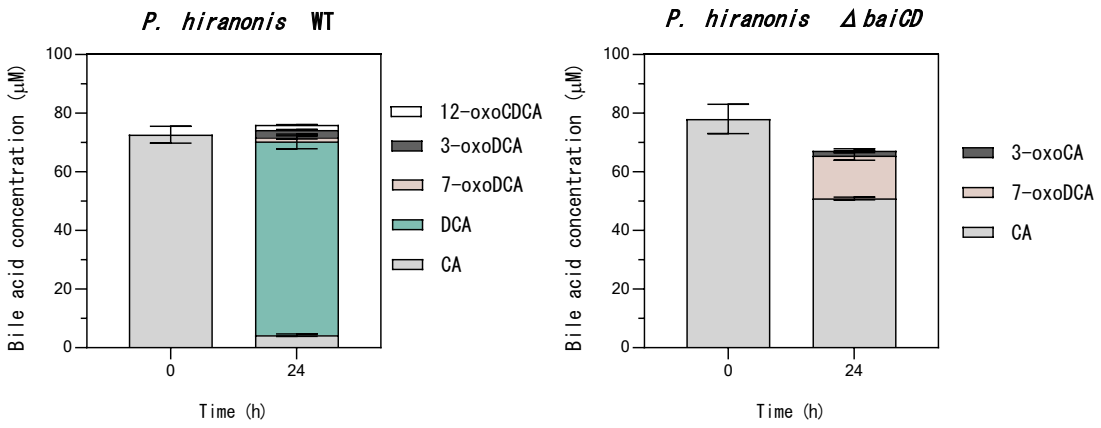


Figure 3.30 7α-dehydroxylation of *P. hiranonis* WT vs *ΔbaiCD*. Both WT and *ΔbaiCD* strains of *P. hiranonis* were grown with cholic acid (CA; 100 μM) for 24 h and bile acid intermediates analysed. These include intermediates of chenodeoxycholic acid (CDCA) and deoxycholic acid (DCA). Samples were prepared, dried and injected on the LC-HRMS system.

These results demonstrate that *baiCD* in *P. hiranonis* is essential for the conversion of CA to DCA by 7α-dehydroxylation, as there is no DCA produced in the KO strain. The conversion in the WT is almost 100%, with an average of 72.67 μM CA measured at 0 h, and 4.29 μM CA and 66.11 μM DCA measured at 24 h. However, in *ΔbaiCD* there was 78.03 μM CA measured at 0 h compared to 50.9 μM at 24 h, but no DCA was detected. This suggests that instead of 7α-dehydroxylation CA has entered an alternative pathway; conversion to 7-oxoDCA is carried out by 7α-HSDH, and is the first step of the conversion to 7-epicholic acid (Ridlon et al., 2006). CA can also undergo oxidation and epimerisation at the 3- and 12-oxy groups by HSDHs. The intermediate and final products of these three pathways were not explored in full in this assay, and this could explain why there is a discrepancy in the total BA concentration between 0 h and 24 h in *ΔbaiCD*.

The abolition of 7α-dehydroxylation in the *P. hiranonis ΔbaiCD* strain provides an exciting opportunity to examine further the role of 7α-dehydroxylation in colonisation resistance against *C. difficile*, and is the first of its kind in any classical 7α-dehydroxylating strain.

3.5.3 Complementation of the ΔbaiCD *strain*

To confirm that the effects observed in future assays are due to the deletion of *baiCD*, a complemented strain is required. Ideally the BM8 bookmark that was inserted during the deletion would be used as a target for CRISPR for repair. However, as genetic tools were unable to be developed for *P. hiranonis*, this is not possible. The alternative is to use a plasmid-based complement system and express the deleted gene on a plasmid.

As *baiCD* is part of an operon, and this operon uses the bile acid inducible promoter with regulatory elements, it was decided to instead use a native constitutive promoter whose expression had been confirmed in the FAST assay. The P_{pgk} promoter was chosen, as the lowest of the two confirmed promoters, to reduce overexpression as much as possible. The expression vector was assembled using hifi PCR, with the fragments P_{pgk} (CDcomp_prom_F1 and CDcomp_prom_R1) and *baiCD* (CDcomp_bai_F1 and CDcomp_bai_R1) inserted into the pMTL84151 backbone digested with XhoI and SalI. The final vector, pMTL84151_CDcomp, was confirmed by colony PCR (ColE1_tra_F2 and pCD6_R1) and Sanger sequencing. It was then transformed into *E. coli* CA434.

The transfer of pMTL84151_CDcomp plasmid into *P. hiranonis ΔbaiCD* was attempted via conjugation and selection with thiamphenicol. Unfortunately, despite three attempts at this conjugation, transfer was not achieved. The pMTL84151 control did conjugate, so it was not an issue with the strain. It could be that the constitutive expression of *baiCD* is toxic, or that the levels expressed from the P_{pgk} promoter are higher than native levels and this is causing toxicity. Whilst the *barAB*/*baiB* promoter could overcome this issue as it is inducible, CA as the inducer would likely interfere with bile acid assays being studied in the strain, so is not an option. Moreover, the use of a plasmid complemented strain would be limited as the antibiotic required to maintain the plasmid would interfere in competition assays.

3.5.4 Preliminary work to determine conditions for competition assays

Whilst the proposed mechanism of colonisation resistance against *C. difficile* by 7α-dehydroxylating bacteria has been suggested as inhibition of germination, secondary

bile acids have been shown to inhibit both germination and vegetative growth (Buffie et al., 2015; Giel et al., 2010; Sorg & Sonenshein, 2008, 2009, 2010). Because of this, it was decided to carry out competition assays using *P. hiranonis* and both spores and vegetative cells of *C. difficile*. The conditions of these assays were first determined to ensure suitability for both organisms.

C. difficile spore preparation

Firstly, suitable strain(s) of *C. difficile* had to be chosen, as it has been shown that different isolates respond to BAs at varying levels. There are different rates and extents of germination in response to taurocholate and glycine between isolates (Heeg et al., 2012; Rajani Thanissery et al., 2017), and also differences in the effects of a variety of BAs on germination, outgrowth and toxin activity (Rajani Thanissery et al., 2017).

The strains R20291 and 630 were chosen based on their responses to DCA, as this will be the main focus of these assays. The two strains cover the range of responses in areas of importance for this assay. Both strains were shown to germinate in response to taurocholate (Heeg et al., 2012; Rajani Thanissery et al., 2017), but R20291 had approximately a four-fold higher percentage of germination compared to 630 (Heeg et al., 2012). The two strains also cover the range of responses to DCA, with R20291 having a higher fold reduction in germination than 630, but 630 having a larger reduction in outgrowth (Heeg et al., 2012). Spore preparations of at least 95% purity were carried out for each strain, as detailed in 2.3.10, with purity confirmed under the microscope.

BA MICs

To determine suitable concentrations to use in the competition assays, the toxicity of BAs to all strains of *P. hiranonis* and *C. scindens* needed to be established. To do this, an MIC assay was carried out for CA and DCA in *P. hiranonis* WT and *ΔbaiCD*, measuring the OD_{600} after 24 h incubation in different concentrations of BAs in BHISS, inoculated with a 1 in 10 dilution of an overnight culture (Figure 3.31). The same was also carried out for both strains of *C. difficile*, 630 and R20291 (Figure 3.32 and Figure 3.33). A spore outgrowth MIC assay was additionally carried out for both strains, inoculating the BAs with a 1 in 10 dilution of spore preparation and taurocholic

acid (TCA) to induce germination, and growth measured after 24 h (Figure 3.32 and Figure 3.33). Both 0.1% and 1% TCA were compared for germination as they have both been used in previous germination assays (Giel et al., 2010; Sorg & Sonenshein, 2008, 2009, 2010), but if germination is being studied in this assay then the concentration must be suitable to demonstrate inhibition with DCA.

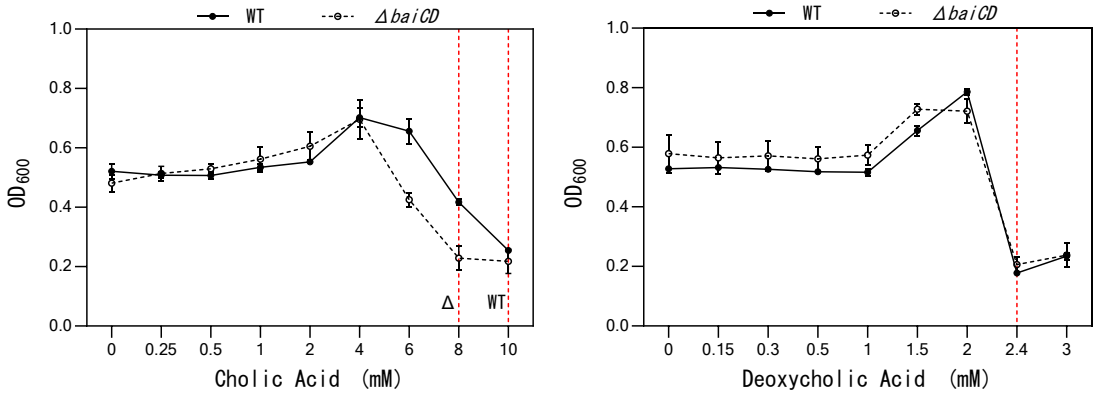


Figure 3.31 Minimum inhibitory concentrations of bile acids in *P. hiranonis*. The WT and *ΔbaiCD* strains of *P. hiranonis* were cultured with a range of concentrations of cholic acid (CA) and deoxycholic acid (DCA). Starting from a 1 in 10 dilution of overnight culture, growth was measured after 24 h by OD_{600}. Data represent mean values of three independent cultures ± SD.

The dose response for each bile acid was largely similar between the two strains of *P. hiranonis* (Figure 3.31). For CA there was similar growth up until 4 mM, and between 4 mM and 10 mM the *ΔbaiCD* strain had reduced growth at each concentration, with an approximate MIC value of 8 mM, compared to 10 mM for the WT. With no conversion of CA to DCA in the *ΔbaiCD* strain, this is a truer reflection of CA toxicity, and likely explains the reduced growth at higher concentrations as there will be a higher concentration of CA present than in the WT. This also suggests that the level of conversion to DCA at the concentrations of CA tested does not produce high enough concentrations of the toxic DCA to impact growth; if this were the case then the WT strain would have lower growth at higher concentrations. The dose response for DCA was almost identical between the two strains, and both had an estimated MIC of 2.4 mM. This is expected given that the *baiCD* KO should have no impact on processing of DCA, and therefore its toxicity.

The MIC values estimated from this assay vary to those published for the WT by Reed et al. (2020); they estimate CA at 2.5 mM, and DCA at 0.78 mM. Moreover, the increase in OD_{600} observed in both assays is unexpected as maximum growth should occur at 0 mM, and it is unlikely that CA and DCA are beneficial for growth at the mid-range concentrations tested. However, this increase was observed in further repeats of the assay (data not shown). To ascertain if this is real or an artefact of the assay, a growth curve could be carried out for both strains at the concentrations used in the MIC assay. In addition to being of interest for the *ΔbaiCD* strain as it has not been studied before, visualising growth over 24 h would clarify BA toxicity. Unfortunately, this assay requires a plate reader in an anaerobic chamber due to the number of concentrations required, and this was not available for this study.

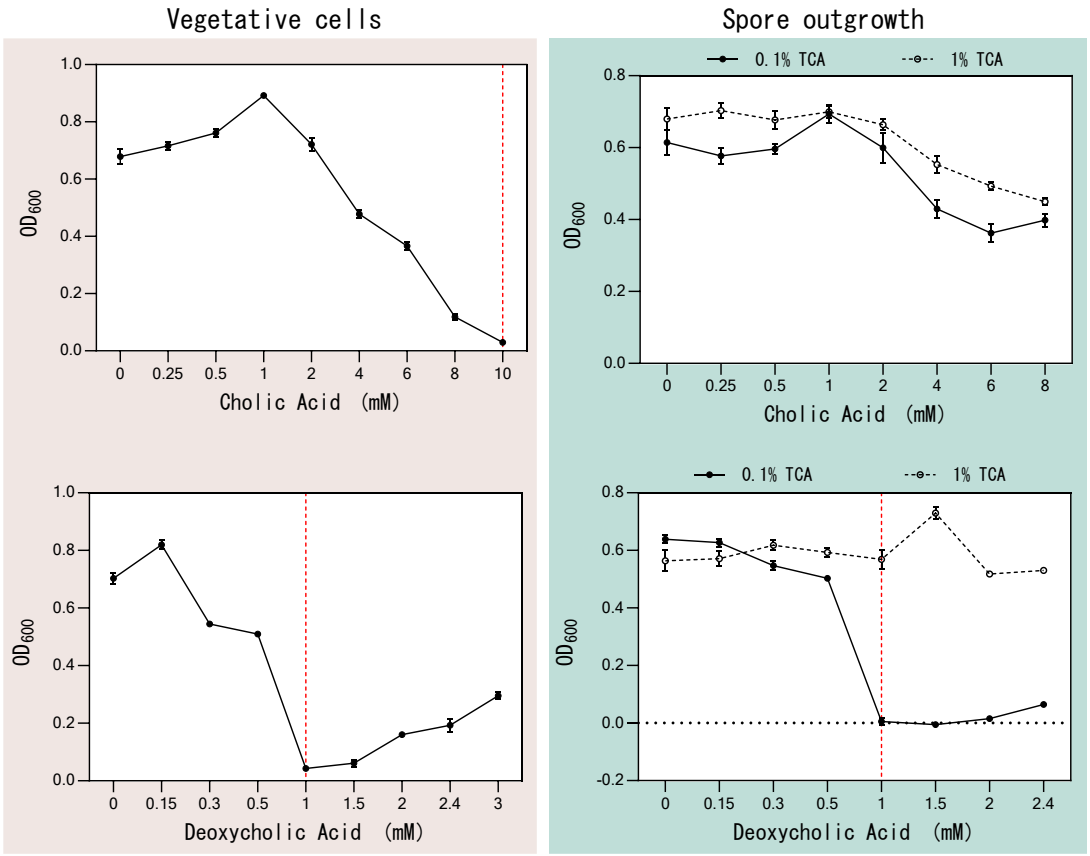


Figure 3.32 Minimum inhibitory concentrations of bile acids in *C. difficile* 630. *C. difficile* 630 was cultured with a range of (CA) and deoxycholic acid (DCA) concentrations, and growth was measured after 24 h by OD_{600}. Vegetative cell cultures were inoculated with a 1 in 10 dilution of overnight culture. Spore cultures were inoculated with a 1 in 10 dilution of a spore preparation and included the germinant taurocholic acid (TCA) at either 0.1% or 1%. Data represent mean values of three independent cultures ± SD.

As shown in Figure 3.32, for *C. difficile* 630 vegetative cells the MIC values are estimated at 10 mM (CA) and 1 mM (DCA), the same values determined by Reed et al. (2020). There was an increase in OD for DCA at concentrations above the MIC value, but this could be due to DCA forming micelles in solution as this was observed in further repeats of the assay.

For *C. difficile* 630 spore outgrowth there was a similar dose response to CA with both 0.1% and 1% TCA. There was a reduction in OD at higher concentrations, but the MIC could not be determined as the concentrations used did not abolish growth

completely. For DCA the MIC using 0.1% TCA was 1 mM, but could not be established for 1 % TCA as the OD was consistent across the concentrations tested.

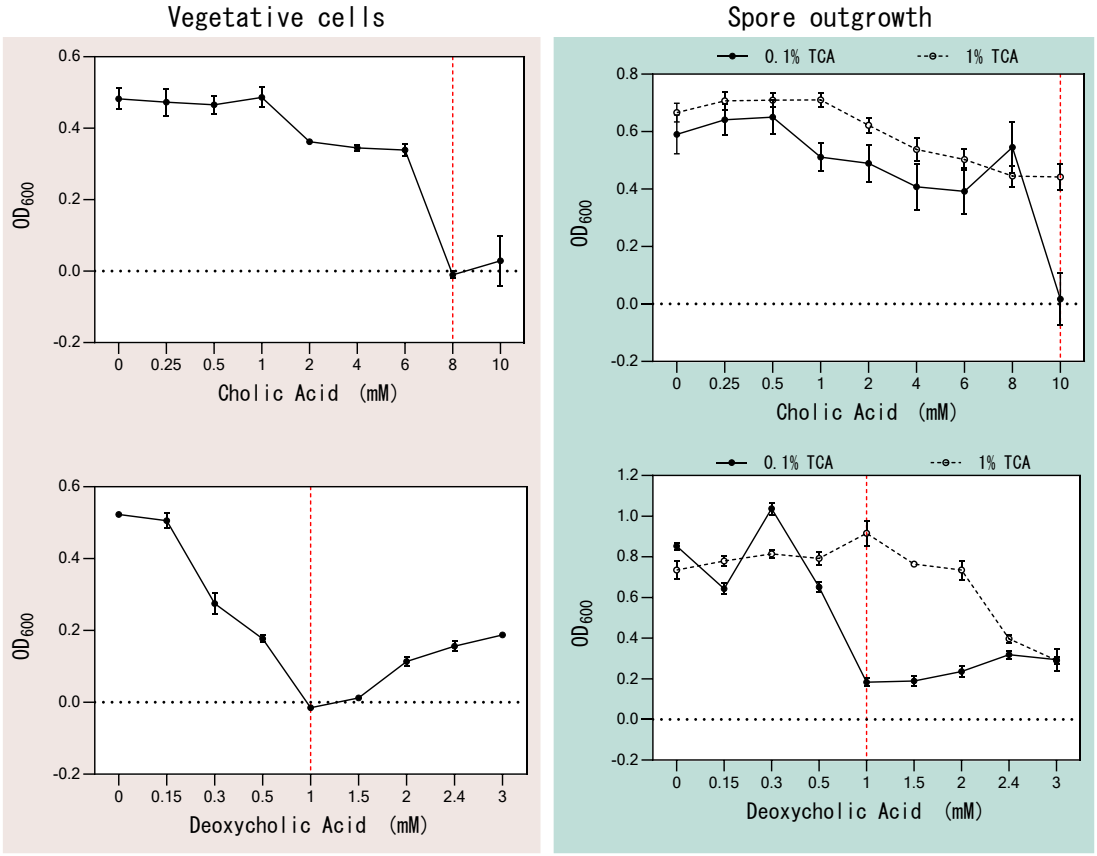


Figure 3.33 Minimum inhibitory concentrations of bile acids in *C. difficile* R20291. *C. difficile* R20291 was cultured with a range of cholic acid (CA) and deoxycholic acid (DCA) concentrations, and growth was measured after 24 h by OD_{600}. Vegetative cell cultures were inoculated with a 1 in 10 dilution of overnight culture. Spore cultures were inoculated with a 1 in 10 dilution of a spore preparation and included the germinant taurocholic acid (TCA) at either 0.1% or 1%. Data represent mean values of three independent cultures ± SD.

The results for *C. difficile* R20291, shown in Figure 3.33, were very similar to that of 630 for both vegetative cells and spores. In vegetative cells the MICs were estimated at 8 mM for CA and 1 mM DCA, which is comparable to the values of 10 mM and 1 mM respectively by Reed et al. (2020). For spore outgrowth the MIC for CA with 0.1% TCA was estimated at 10 mM, but could not be estimated for 1% TCA. Again, for DCA, there was a reduction in OD at the higher concentrations with 0.1% TCA,

with an estimated MIC of 1 mM, but a reduction in growth was only observed at 2 mM with 1% TCA.

These results suggest that to ensure sufficient growth in the presence of BAs, appropriate concentrations include: 4 mM CA and 2 mM DCA for *P. hiranonis*, and 2 mM CA and 1 mM DCA for *C. difficile*. The concentration of TCA was decided as 0.1%.

3.5.5 Supernatant assays

To identify possible colonisation resistance *in vitro*, the species of interest can be grown in the presence of the supernatant of another bacterial culture. This removes the impact of direct growth competition, for example for Stickland metabolism substrates by Stickland fermenting clostridia, a proposed method of protection against CDI (Aguirre et al., 2021; Arrieta-Ortiz et al., 2021; Girinathan et al., 2021), and allows the impact of metabolites produced to be studied. It also removes the need for selection against one or both organisms being studied, as only the species of interest is cultured. Accordingly, this method was chosen to study the effect of 7α-dehydroxylation in *P. hiranonis* on the growth of *C. difficile*, with the hypothesis being that DCA production may inhibit germination and/or growth, and that this inhibitory effect would be absent in the *ΔbaiCD* strain.

Spore Outgrowth

To study the impact of *P. hiranonis* bile acid production on the germination of *C. difficile* spores, spore outgrowth was assessed (Figure 3.34). The WT and *ΔbaiCD* strains of *P. hiranonis* were grown overnight in BHISS + 4 mM CA, as decided by the MIC assay. Unfortunately, neither strain grew so lower concentrations of CA were trialled, and 1 mM was the highest concentration that allowed growth of both strains to the same OD_{600} overnight (data not shown). Supernatants were taken for both strains with and without CA, and were prepared according to 2.3.11.

The spore outgrowth assay was carried out according to 2.3.12, with growth measured with an OD_{600} reading every hour (Figure 3.34).

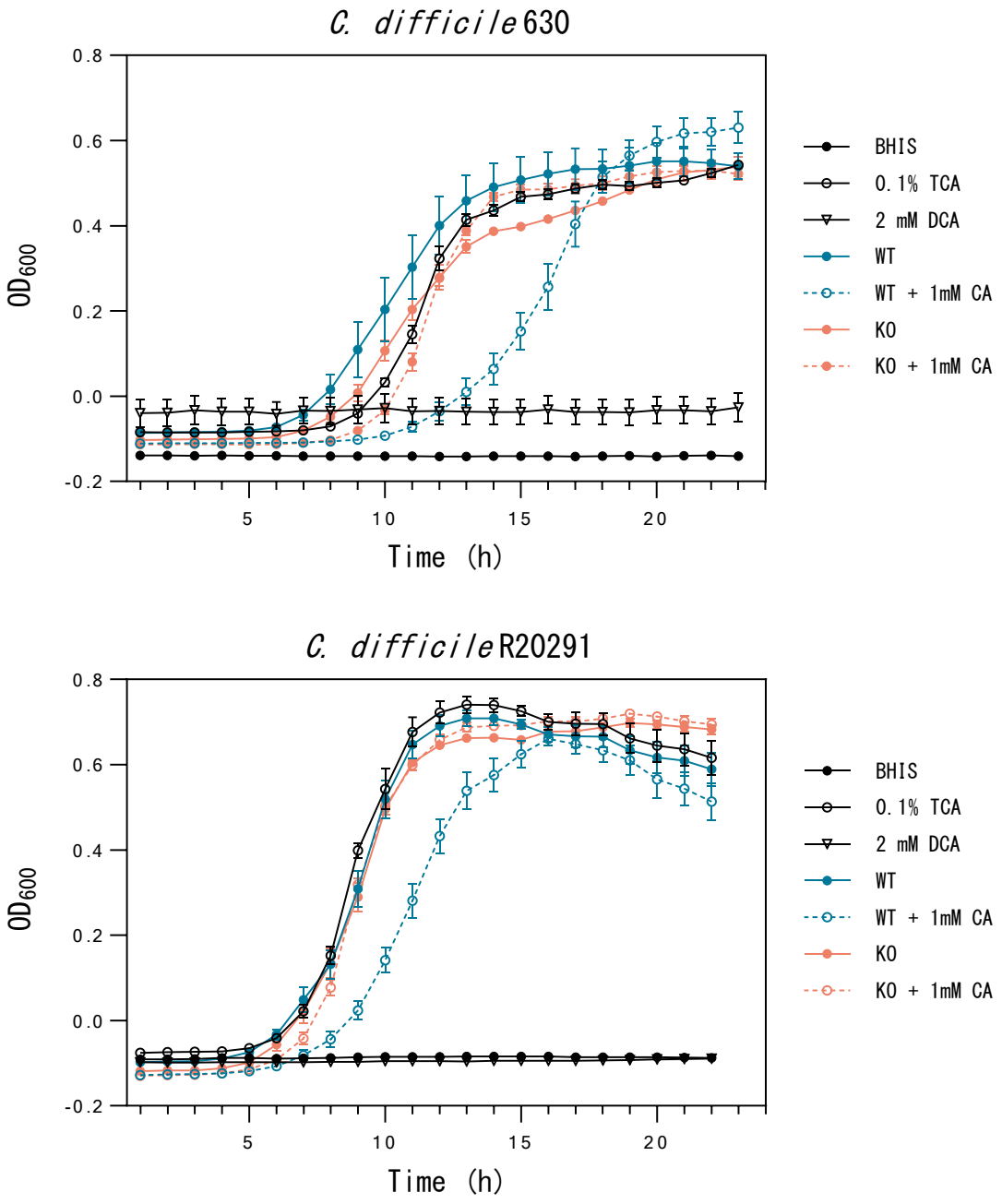


Figure 3.34 Spore outgrowth of *C. difficile* 630 and R20291. Spores were incubated with deoxycholic acid (DCA) and prepared supernatants from *P. hiranonis* WT and *ΔbaiCD* (KO) strains incubated with and without cholic acid (CA). The germinant TCA at 0.1% was also incubated with all spores. Growth was measured by OD_{600} in a plate reader every hour for 22 h. Data represent mean values of spores incubated with three independent supernatant preparations, each in technical triplicate ± SD.

A similar trend in spore outgrowth was observed in both strains of *C. difficile*. As expected, no spore growth was seen without the germinant TCA, or in the presence of DCA, shown to be inhibitory to *C. difficile* germination. All other supernatants studied followed a similar growth profile to that of the positive control, BHIS + 1% TCA, except for the WT + 1 mM CA, which had a similar but delayed profile. R20291 spores had a quicker onset of outgrowth than 630, at approximately 5 h compared to 8-10 h, with a more rapid exponential phase and earlier stationary phase. Both strains reached a similar final OD_{600} of approximately 0.6. The supernatants studied produced more consistent growth profiles in R20291 than 630.

The delayed growth profile of WT + 1 mM CA in both *C. difficile* strains is likely due to the conversion to DCA by 7α-dehydroxylation in *P. hiranonis*. This conclusion is confirmed by the removal of the growth delay in the *ΔbaiCD* strain; this strain is incapable of 7α-dehydroxylation, and therefore there should be no DCA present to inhibit germination. This result further strengthens the theory that 7α-dehydroxylation is the mechanism of colonisation resistance against *C. difficile*, although a combination with other mechanisms cannot be disregarded.

The reduction of spore outgrowth with the WT + 1 mM CA supernatant suggests that the levels of 7α-dehydroxylation in *P. hiranonis* do not produce sufficient concentrations of DCA to completely eliminate spore outgrowth. Previous work has demonstrated that supernatants from *C. scindens* incubated with 2 mM CA were able to eliminate spore outgrowth of six different strains of *C. difficile* (Dempster, 2017). This could be due to a higher level of 7α-dehydroxylation in *C. scindens,* which has been shown previously (Maki Kitahara et al., 2000), or due to a higher CA MIC. Varying incubation times and CA concentrations could be explored for *P. hiranonis* to determine if complete elimination of spore outgrowth could be achieved, and compared to *C. scindens*.

Germination Initiation

Following the spore outgrowth assay a germination initiation assay was carried out using the same supernatant samples to establish whether it is a change in germination initiation that is resulting in the delayed spore outgrowth observed with DCA

production. The germination assay was carried out according to 2.3.13, detecting germination as an initial drop in OD_{600} when the spores rehydrate and become less optically dense (Figure 3.35).

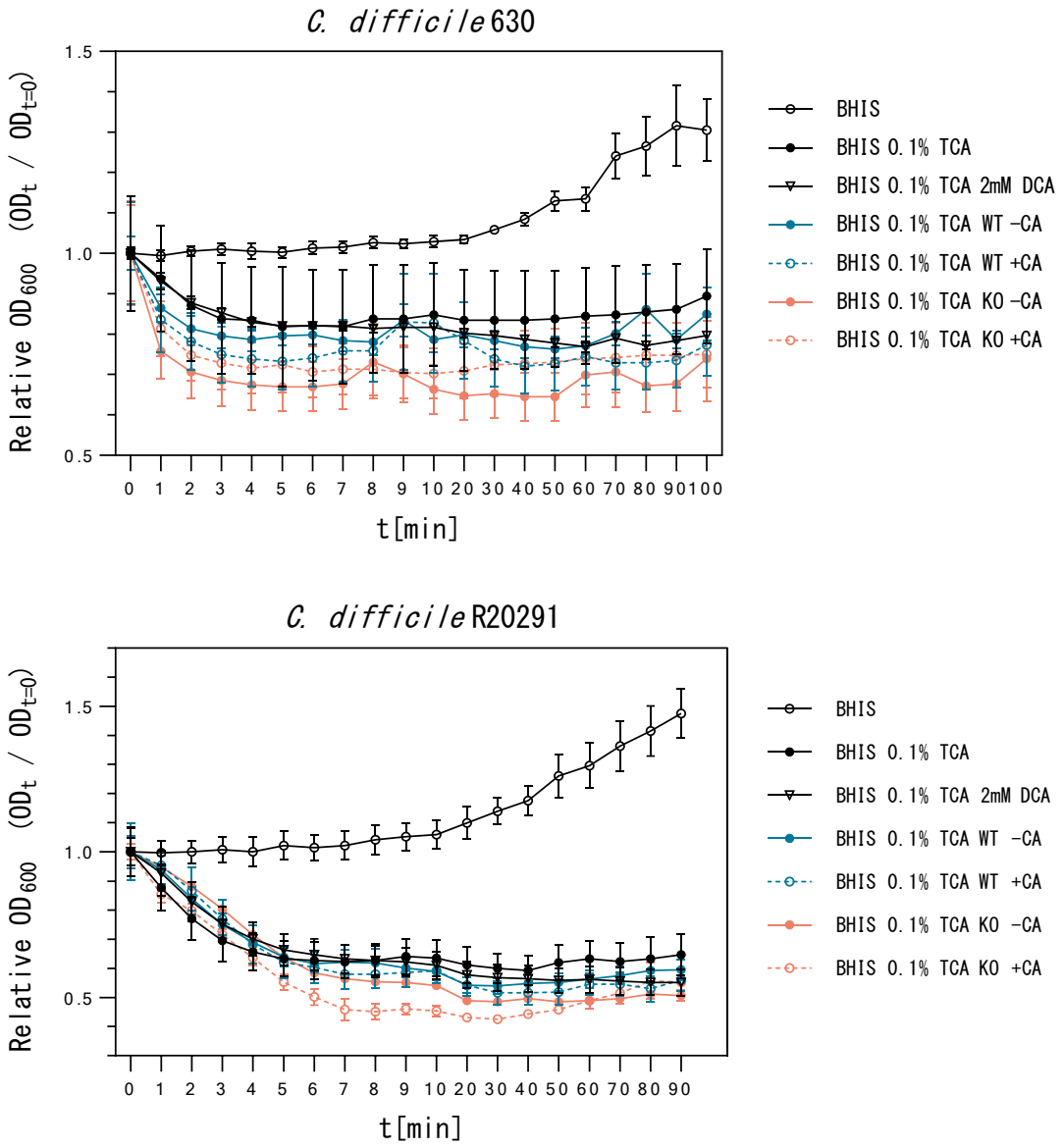


Figure 3.35 Germination initiation of *C. difficile* 630 and R20291. Spores were incubated with deoxycholic acid (DCA) and prepared supernatants from *P. hiranonis* WT and *ΔbaiCD* (KO) strains incubated with and without cholic acid (CA). The germinant TCA at 0.1% was also incubated with all spores. Growth was measured by OD_{600} in a plate reader to capture early decreases. Data represent mean values of spores incubated with three independent supernatant preparations, each in technical triplicate ± SD.

For both strains of *C. difficile*, except for the negative BHISS control without germinant, germination initiation occurred in all supernatant conditions as indicated by the initial decrease in OD_{600}, following the same or similar profile to that of the TCA positive control. The R20291 spores exhibited less variation between the supernatant conditions than 630 spores, and also dropped to a lower relative OD_{600}, approximately 0.5 compared to 0.7.

Given that the 2 mM DCA control exhibited the same profile as the positive control in both cases, this indicates that the effect of DCA and supernatants + CA on spore outgrowth in Figure 3.34 is likely due to effects on downstream germination mechanisms, or on the growth of vegetative cells following germination; DCA is not affecting germination initiation. As the WT + CA supernatant also did not inhibit germination initiation, this increases confidence in DCA production being the mechanism for delay in spore outgrowth.

The variation between spore strains is likely to be due to their differing germination rates to TCA; R20291 was shown to have five times higher the percentage germination rate than 630 (Rajani Thanissery et al., 2017). This could explain why there is a larger fall in OD_{600}, and less variation within the assay overall.

3.5.6 Competition assays

Transwell plate assays

Given that the role of bile acids in colonisation resistance by 7α-dehydroxylation has been debated, it was decided to also study direct growth competition between *P. hiranonis* and *C. difficile*. The *ΔbaiCD* strain should help decipher what influence, if any, the production of DCA is having on *C. difficile* germination and growth. Consequently, it was decided to study both spore outgrowth and vegetative cell growth where possible.

To allow the concurrent growth of both *P. hiranonis* and *C. difficile* to be studied, it was decided to carry out a transwell plate assay. Transwell plates contain two chambers (inner and outer) per well, separated by a membrane filter that allows diffusion of proteins and small molecules but is impermeable to bacteria. They have previously been used to studied peptide signalling in quorum sensing (Ma, Li, &

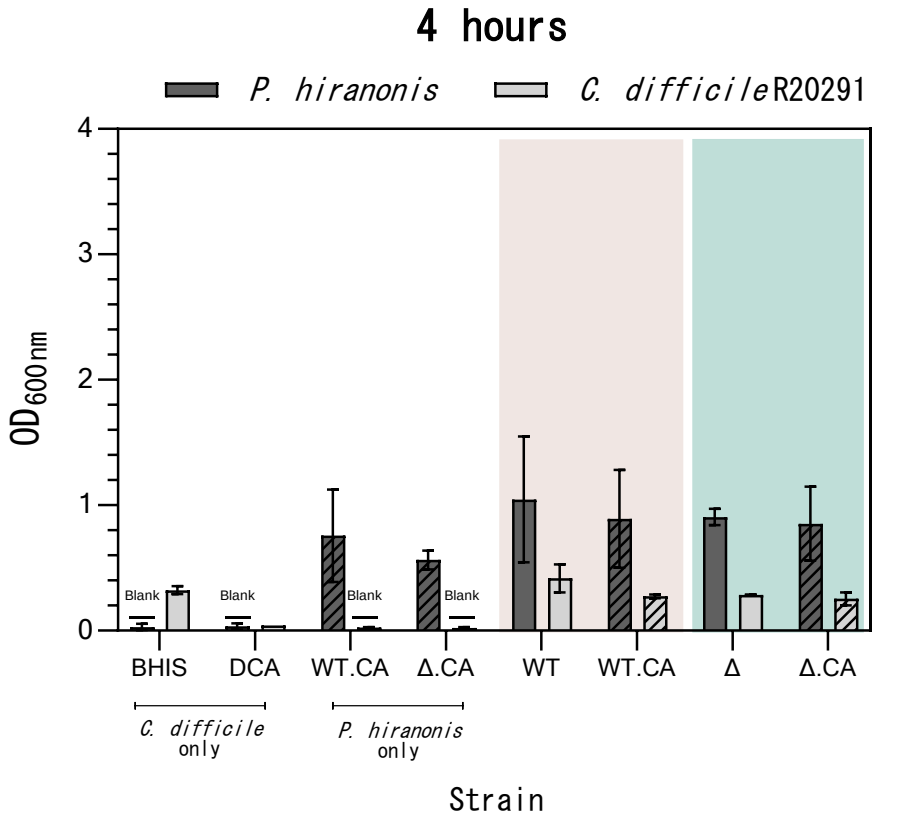
McClane, 2015; Vidal et al., 2012; Yu et al., 2017), and should allow the impact of bile acid production to be studied; bile acids from the *P. hiranonis* well should diffuse into the *C. difficile* well to elicit a response, but as the two bacteria are separated their individual growth can be monitored.

Two separate assays were carried out, one for spores and one for vegetative cell growth, and as the number of plates and wells was limited only *C. difficile* R20291 was studied. The assay was carried out as described in 2.3.14. The wells were loaded with strains and bile acid conditions as detailed in Table 3.27. As the wells are of limited size, only two OD_{600} readings could be taken per assay, chosen as 4 and 8 hours for the spore assay, and 8 and 24 hours for the vegetative cell assay.

Table 3.27 Conditions used in transwell plate studies. *P. hiranonis* WT or *ΔbaiCD* is incubated in BHISS liquid medium in the outer chamber. *C. difficile* spores or vegetative cells are incubated in BHISS liquid medium in the inner chamber, with 0.1% TCA for spore germination.

Well	***P. hiranonis* (vegetative)**	***C. difficile* (spores or vegetative)**
1	BHISS control. *P. hiranonis* -ve.	*C. difficile*.
2	BHISS + 2 mM DCA control. *P. hiranonis* -ve.	*C. difficile*.
3	*P. hiranonis* WT.	BHISS control. *C. difficile* -ve.
4	*P. hiranonis ΔbaiCD*.	BHISS control. *C. difficile* -ve.
5	*P. hiranonis* WT.	*C. difficile*.
6	*P. hiranonis* WT + 1mM CA.	*C. difficile*.
7	*P. hiranonis ΔbaiCD*.	*C. difficile*.
8	*P. hiranonis ΔbaiCD* + 1 mM CA.	*C. difficile*.

The results for the spore assay are shown in Figure 3.36 and for the vegetative assay in Figure 3.37.


4 hours
P. hiranonis
C. difficile R20291
OD600nm
Blank
BHIS
DCA
WT.CA
Δ.CA
WT
WT.CA
Δ
Δ.CA
C. difficile only
P. hiranonis only
Strain

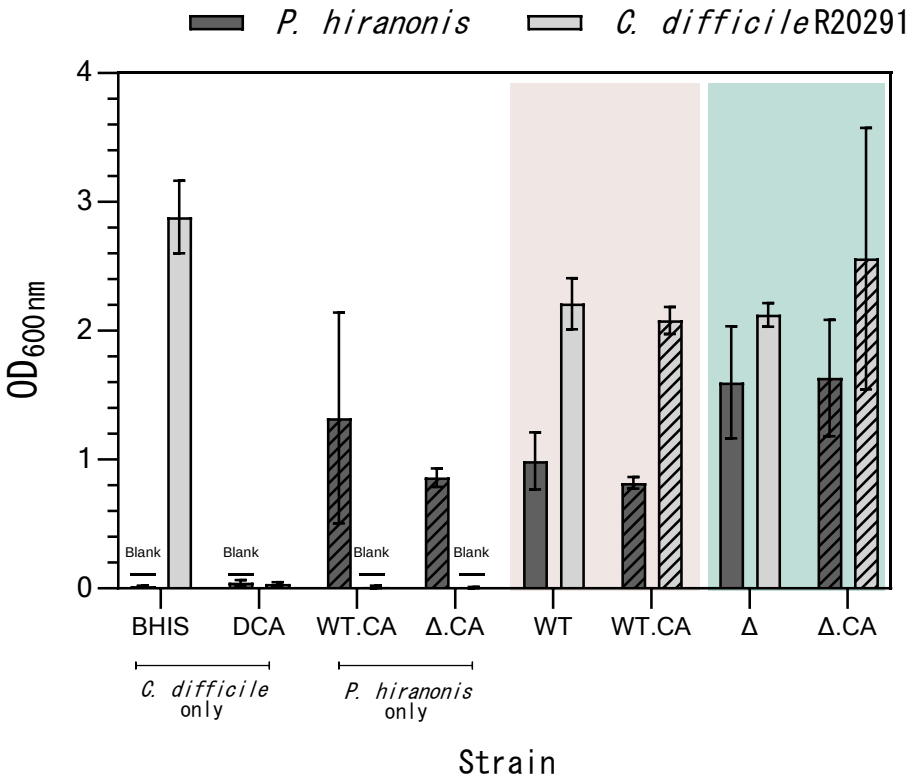

8 hours
P. hiranonis
C. difficile R20291
OD600nm
Blank
BHIS
DCA
WT.CA
Δ.CA
WT
WT.CA
Δ
Δ.CA
C. difficile only
P. hiranonis only
Strain

Figure 3.36 Transwell plate assay for *C. difficile* spore outgrowth. Spores of *C. difficile* R20291 were incubated with 0.1% TCA in the inner well of the transwell plate. *P. hiranonis* WT or *ΔbaiCD* (KO) strains were incubated with and without cholic acid (CA) in the outer well. Samples were taken at 4 and 8 h and growth measured by OD_{600}. Statistical analysis was carried out with one-way ANOVA with Tukey's test for multiple comparisons, one test for *P. hiranonis* and one test for *C. difficile*. Data represent mean values of three independent culture pairs (three wells) ± SD.

For the spore assay at 4 h there was little difference in growth ratio between any of the four assay conditions, *P. hiranonis* had a higher OD_{600} throughout but it was roughly the same across the four conditions, at approximately 1.0. The OD_{600} values of *C. difficile* were all lower, at approximately 0.3, and this is expected due to the time required for *C. difficile* spores to germinate and begin to grow. In a one-way ANOVA with Tukey's test for multiple comparisons of *P. hiranonis* grown in wells where *C. difficile* was present, there were no significant differences amongst the OD_{600} readings. In the same test for *C. difficile,* growth with WT *P. hiranonis* was higher than with WT.CA ($p=0.0217$), Δ ($p=0.0365$) and Δ.CA ($p=0.0075$).

At 8 h the OD_{600} of *P. hiranonis* had increased slightly from the reading at 4 h and *C. difficile* readings increased to approximately 2.1. Again for *P. hiranonis* there was no significant differences amongst the OD_{600} readings in the wells where *C. difficile* was also present. This was the same for *C. difficile*.

At both time points there was no growth of *C. difficile* with the presence of 2 mM DCA in the *P. hiranonis* well, which is as expected and demonstrates proper diffusion across the membrane. The growth of each species in combination was also similar to that of the individual control well, suggesting that nutrient competition is unlikely to be affecting growth. Whilst it would have been beneficial to have a *C. difficile* + CA control, there were not enough wells available.

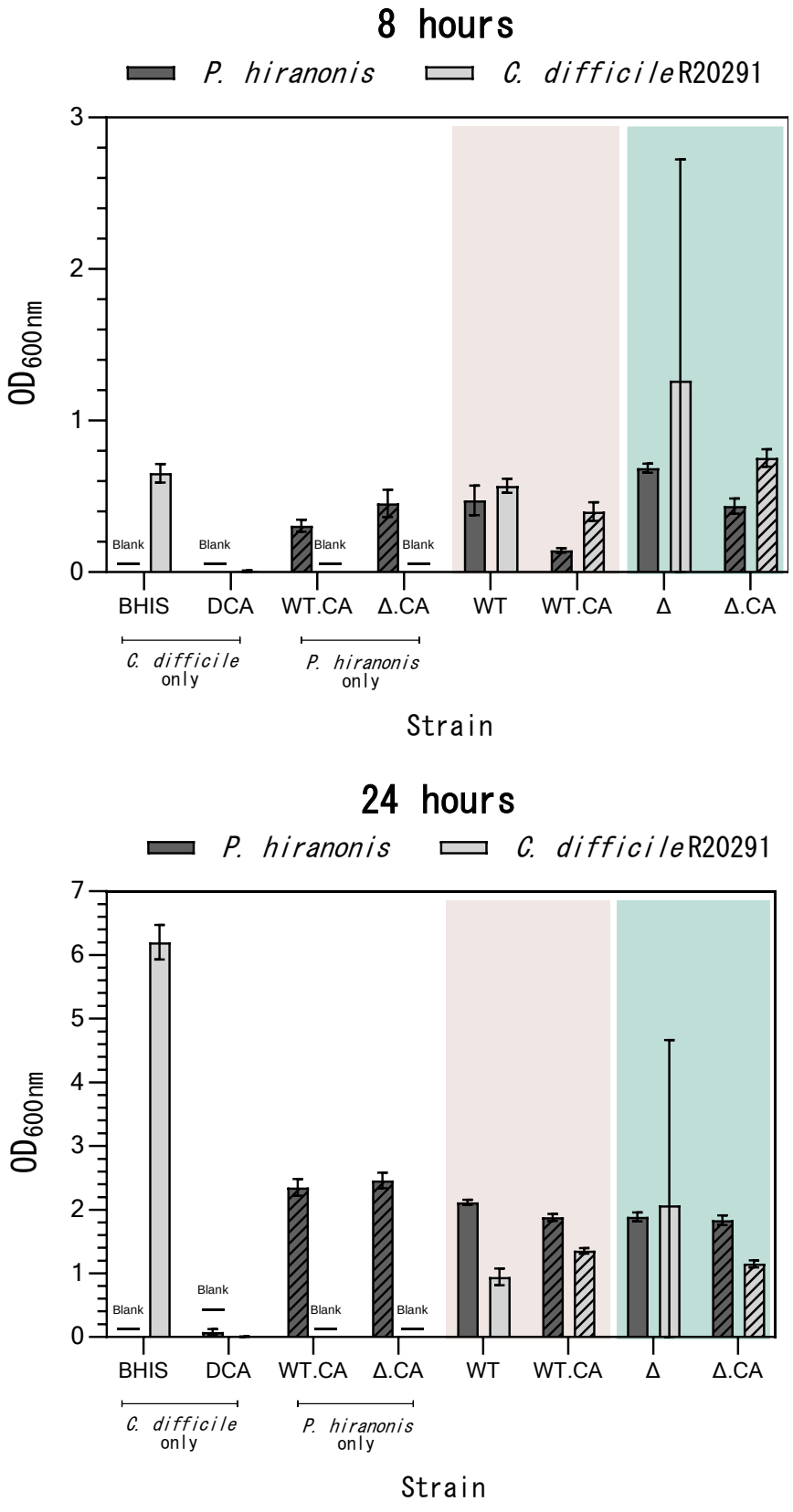

8 hours
P. hiranonis
C. difficile R20291
OD600nm
3
2
1
0
Blank
Blank
Blank
Blank
BHIS
DCA
WT.CA
Δ.CA
WT
WT.CA
Δ
Δ.CA
C. difficile only
P. hiranonis only
Strain
24 hours
P. hiranonis
C. difficile R20291
OD600nm
7
6
5
4
3
2
1
0
Blank
Blank
Blank
Blank
BHIS
DCA
WT.CA
Δ.CA
WT
WT.CA
Δ
Δ.CA
C. difficile only
P. hiranonis only
Strain

Figure 3.37 Transwell plate assay for *C. difficile* vegetative growth. *C. difficile* R20291 was inoculated in BHISS in the inner well of the transwell plate. *P. hiranonis* WT or *ΔbaiCD* (KO) strains were incubated with and without cholic acid (CA) in the outer well. Samples were taken at 8 and 24 h and growth measured by OD_{600}. Statistical analysis was carried out with one-way ANOVA with Tukey's test for multiple comparisons, one test for *P. hiranonis* and one test for *C. difficile*. Data represent mean values of three independent culture pairs (three wells) ± SD.

At 8 h the growth of *P. hiranonis* was very inconsistent, with no observable pattern between the two strains, or in the presence of CA. In a one-way ANOVA with Tukey's test for multiple comparisons of *P. hiranonis* grown in wells where *C. difficile* was present, the WT was higher than WT.CA ($p<0.0001$) but lower than Δ ($p=0.0031$), the WT.CA was lower than Δ ($p<0.0001$) and Δ.CA ($p=0.0001$), and the Δ was higher than Δ.CA ($p=0.0006$). The controls suggest that there is no difference in growth individually of the WT.CA and Δ.CA ($p=0.0599$). For *C. difficile* there was no significant differences amongst the OD_{600} readings in the wells where *P. hiranonis* was also present.

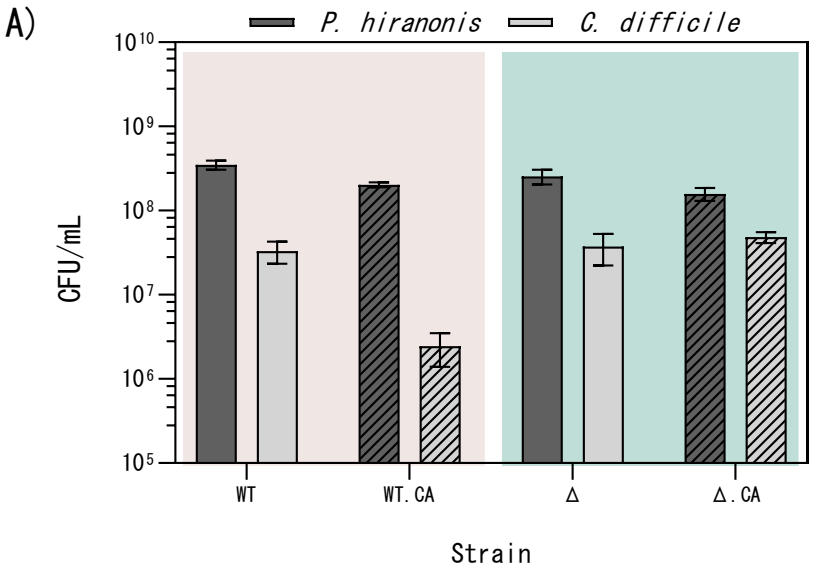
A similar trend is observed at 24 h. For *P. hiranonis*, the wildtype has a higher OD_{600} than WT.CA ($p=0.0335$), Δ ($p=0.0408$) and Δ.CA ($p=0.0091$). Again for *C. difficile* there was no significant differences amongst the OD_{600} readings in the wells where *P. hiranonis* was also present.

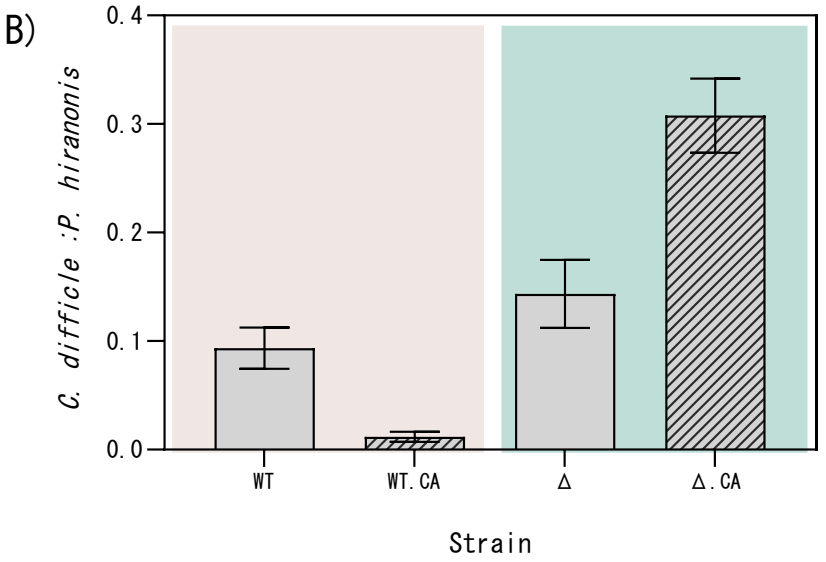
Overall, between the two transwell assays no clear conclusions can be drawn with regards to the impact of *P. hiranonis* metabolite usage and bile acid production on the growth of *C. difficile*, from spores or vegetative cells. With regards to 7α-dehydroxylation one would have expected that the WT.CA *P. hiranonis* could have slowed down or prevented growth of *C. difficile* due to the production of DCA, and this effect would have been removed in the *ΔbaiCD* strain. This was not observed at either time point with spores or vegetative cells, and there was no clear trend. Different OD_{600} readings of *P. hiranonis* in the different test conditions across timepoints also makes it difficult to compare the effects on *C. difficile*, but there was no trend across this either.

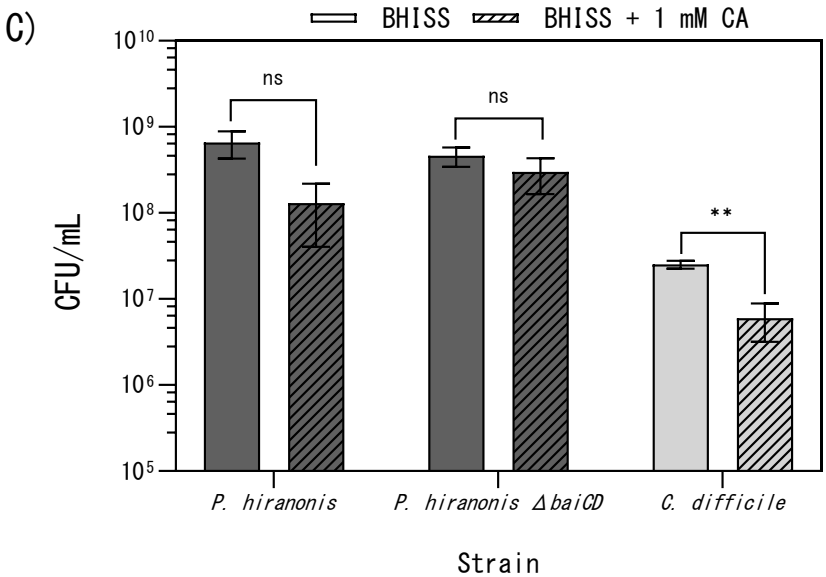
One of the biggest issues with these assays was the variability in data recorded. For example, comparisons of each strain when grown in the presence of the other species to the respective individual control at each time point could have indicated if there was competition for resources. However, the very large range in OD_{600} across many data points has reduced the reliability of such comparisons, and of comparisons in the assay in general. This is perhaps due to the transwell plates themselves as, despite undergoing constant gentle agitation and manual resuspension before sampling, visually there appeared to be clumping of bacteria at the bottom of each section of the well. Moreover, this clumping may have blocked the membrane and prevented free and consistent diffusion of the bile acids, which could explain why the expected results were not seen. For both these reasons the transwell plate assay format was deemed unsuitable for future use, and no conclusions can be confidently drawn from either of these assays.

Co-culture experiment

Due to the inconclusive data of the transwell plates it was decided to carry out a co-culture assay (2.3.15) with *P. hiranonis* and *C. difficile* to assess the same conditions, using vegetative cells only (Figure 3.38). Species were distinguished by colony morphology as unfortunately the usual antibiotic selection for *C. difficile*, d-cycloserine (250 µg/mL) and cefoxitin (8 µg/mL) did not select against *P. hiranonis*.


A)
P. hiranonis
C. difficile
CFU/mL
WT
WT.CA
Δ
Δ.CA
Strain


B)
C. difficle :P. hiranonis
WT
WT.CA
Δ
Δ.CA
Strain


C)
BHISS
BHISS + 1 mM CA
ns
ns
**
CFU/mL
P. hiranonis
P. hiranonis ΔbaiCD
C. difficile
Strain

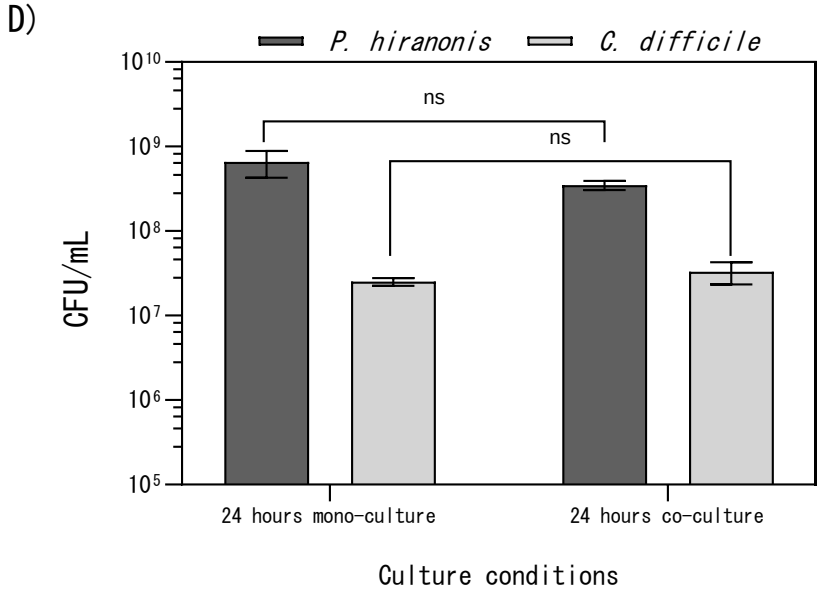


Figure 3.38 Co-culture of *C. difficile* R20291 and *P. hiranonis*. *C. difficile* R20291 and *P. hiranonis* WT or *ΔbaiCD* (Δ) were inoculated into BHISS with and without cholic acid (CA). After 24 h of growth samples were taken and plated onto BHISS plates to determine each CFU/mL. Data represent mean values of three independent co-cultures ± SD. A) CFU/mL measured for the co-culture. B) Ratio of *C. difficile*: *P.hiranonis*. C) Individual control cultures with and without CA. Statistical analysis was carried out to compare CA- vs CA+ using an unpaired two-tailed Student's t-test, p-values are indicated as non-significant (ns), or p<0.005 (**). D) Comparison of WT *C. difficile* R20291 and *P. hiranonis* growth after 24 h in mono- or co-culture. Statistical analysis was carried out using an unpaired two-tailed Student's t-test, p-values are indicated as non-significant (ns).

When comparing the CFU/mL of *C. difficile* when co-cultured with *P. hiranonis*, there was a reduction in growth with WT.CA compared to WT (One-way ANOVA with Tukey's test for multiple comparisons, p=0.0146) of approximately 90%. There was also a similar reduction compared to Δ (p=0.0129) and Δ.CA (p=0.0015). This reduction with CA was not observed in the Δ strain, as there was no significant difference between *C. difficile* growth with Δ and Δ.CA.

The growth of *P. hiranonis* was not consistent throughout all conditions tested. In a one-way ANOVA with Tukey's test for multiple comparisons of *P. hiranonis* only, there was a significant difference between WT vs WT.CA (p=0.0049), with a

reduction of approximately 42%, and WT vs Δ.CA (0.001), with a reduction of approximately 55%. However, there was no difference between Δ and Δ.CA. This makes it more difficult to directly compare the impact of 7α-dehydroxylation on *C. difficile* growth, as the growth of *P. hiranonis* may also have had an impact.

Whilst it is the effect on CFU/mL of *C. difficile* that is of the most interest, the ratio of *C. difficile*: *P. hiranonis* was calculated to consider the effects of changes in *P. hiranonis* growth too (Figure 3.38B). There was reduction with WT vs WT.CA, but an increase in Δ and Δ.CA.

The mono-culture growth of *P. hiranonis* strains and *C. difficile* was also measured, with and without CA (Figure 3.38C). Neither strain of *P. hiranonis* exhibited a reduction in growth on addition of CA, and in both conditions had similar CFU/ml values as in the co-cultures. *C. difficile* did demonstrate a reduction in CFU/mL with CA ($p=0.0044$). Without CA it demonstrated similar values to those with WT and Δ in the co-cultures. With CA it had a similar value to that with the WT, but a lower value than that with Δ.

With regards to bile acid dependent inhibition, overall this assay appears to demonstrate reduction in *C. difficile* growth from the 7α-dehydroxylation of *P. hiranonis*. Accordingly, when the WT is incubated with CA there is a reduction in *C. difficile* growth, but no such reduction is seen with the Δ strain. This suggests that 7α-dehydroxylation by *P. hiranonis* could impact *C. difficile* vegetative growth, as well as the rate of germination (shown in 3.5.5). Whilst *C. scindens* has been shown to impact vegetative growth of *C. difficile* in a bile acid dependent manner (Reed et al., 2020), there is limited literature investigating CA containing co-cultures of *C. difficile* and 7α-dehydroxylating bacteria, including *P. hiranonis*.

The reduction in CFU/mL for the individual culture of *C. difficile* on addition of CA could explain the apparent inhibition by *P. hiranonis* + CA. However, if this was the case it would also be observed in the Δ.CA co-culture, which it is not. Moreover, the impact of CA on *C. difficile* in the individual culture is surprising as 1 mM CA is far below the predicted 10 mM MIC for R20291 vegetative cells (Reed et al., 2020).

The absence of a reduction in CFU/mL for both *P. hiranonis* strains with 1 mM CA therefore cannot explain the variation in growth of *P. hiranonis* observed in the co-

cultures. The reduction from WT vs WT.CA could therefore be due to an effect from *C. difficile*, but this is unlikely as it had substantially reduced growth.

With regards to bile acid independent effects of co-culture of *P. hiranonis* and *C. difficile*, neither strain demonstrated a change in growth when comparing their mono-culture to the co-culture (Figure 3.38D)(unpaired two-tailed Student's t-test, p=0.1411 *P. hiranonis*, p=0.2960 *C. difficile*). This could suggest that there is not a competition for Stickland metabolites, despite both strains having been shown to consume them (Battaglioli et al., 2018; Ridlon et al., 2020). However, this is contradictory to previous studies. In a co-culture by R. Thanissery et al. (2020), the CFU/mL of *P. hiranonis* was significantly reduced in the co-culture compared to the monoculture for both a 1:1 and 1:10 ratio of *C. difficile*:*P. hiranonis*, and there was no difference for *C. difficile*. Only in the 1:100 was *P. hiranonis* not reduced, and neither was *C. difficile*. Reed et al. (2020) also demonstrated significant inhibition of *P. hiranonis* in a 1:1 co-culture with *C. difficile* compared to the monoculture. These differences could be due to differing media used, with BHIS containing various supplements used; Hromada et al. (2021) observed differing levels of *C. difficile* inhibition by *P. hiranonis* with restoration of different compounds. However, the incongruous results of inhibition by *C. difficile* or *P. hiranonis*, across differing assays for Stickland competition, co-cultures, media supplementation and mice colonisation (Aguirre et al., 2021; Hromada et al., 2021; Reed et al., 2020; R. Thanissery et al., 2020) suggests that further work needs to be done before a consensus can be reached on the prevention of CDI by Stickland competition.

3.6 Key Outcomes

The overall aim of this chapter was to develop a reliable method of gene deletion in the poorly studied *P. hiranonis*. DNA transfer was first established in the organism via conjugation, and conjugation efficiency was further explored with different Gram-positive replicons and *in silico* prediction of RM systems. The establishment of DNA transfer allowed the examination of existing genetic tools for use in *P. hiranonis*, including Clostron, endogenous CRISPR and the exogenous inducible CRISPR system, RiboCas, but none of the tools tested proved functional. The RiboCas system, whilst generating a *baiCD* knockout strain, did not allow reproducible and reliable gene knockouts to be generated, so efforts were then focused on its optimisation for

use in *P. hiranonis*. A FAST promoter assay suggested that the exogenous promoters commonly used in clostridial tools, including in RiboCas, were non-functional in *P. hiranonis*, which led to the identification and characterisation of native promoters. Despite many iterations of the RiboCas system being explored with insertion of native promoters, including the use of those rendered inducible through insertion of a riboswitch, further gene deletions could not be achieved.

Although it was not possible to establish a reproducible method of mutagenesis, the *baiCD* knockout strain did provide the opportunity to examine the bile acid-mediated role of *P. hiranonis* in colonisation resistance against *C. difficile*. Mass spectrometry analysis confirmed that the *baiCD* knockout strain was not capable of 7α-dehydroxylation as it did not convert CA to DCA, and the strain was therefore studied in competition assays with *C. difficile*. Supernatant assays demonstrated that the knockout strain, when incubated with CA, did not result in the delayed germination of *C. difficile* spores observed with the wildtype strain. Co-culture assays also demonstrated a reduction in vegetative growth of *C. difficile* when incubated with the wildtype *P. hiranonis* and CA, but not with the knockout strain and CA. These results illustrate the presence of bile-acid mediated resistance to *C. difficile* by *P. hiranonis*, and the *baiCD* knockout strain provides a basis for furthering understanding of the mechanisms involved.

Chapter 4 Establishing a *Clostridium butyricum* strain for heterologous gene expression through generation of auxotrophic gene knockouts

4.1 Introduction

4.1.1 C. butyricum *background*

Clostridium butyricum is an anaerobic, Gram-positive, spore-forming bacillus that is the type species for the genus *Clostridium*. It is a common commensal bacterium first isolated from pig intestines in 1880 and is found in both the animal and human gut, as well as a range of environments including soil (Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, 2009). Due to the broad range of niches it inhabits, *C. butyricum* colonisation has a wide-reaching relevance in many settings, including healthcare, biotechnology and the food industry.

Named for its high level of butyric acid production, production of butyrate by *C. butyricum* has been beneficially harnessed, however it also poses a threat to food hygiene. The bacterium can grow in foods, including canned foods, where the production of butyric acid causes spoilage. This is made more persistent through the survival of spores during heat treatment (Ghoddusi & Sherburn, 2010). Moreover, far more serious than the inconvenience of food spoilage, *C. butyricum* further poses a challenge to human health due to the production of neurotoxins by select strains. *C. butyricum* was classified as a non-pathogenic organism until 1986, when it was first linked to two cases of botulism in Italy (Aureli et al., 1986). Botulism is a potentially fatal, acute paralytic disease caused by botulinum neurotoxin (BoNT). Whilst most commonly produced by *Clostridium botulinum*, there are some atypical strains of *C. butyricum* that produce BoNT type E and these neurotoxinogenic strains have been implicated in infectious botulism, caused by wound or intestinal colonisation and subsequent toxin secretion (Aureli et al., 1986; Cassir, Benamar, & La Scola, 2016; Dykes, Lúquez, Raphael, McCroskey, & Maslanka, 2015). Type E *C. butyricum* has also been implicated in food-bourne cases of botulism (Meng et al., 1997). Whilst posing a serious threat, toxigenic strains are relatively rare, for example in a study of 93 tested isolates from 978 environmental samples, none were found to be toxigenic (Ghoddusi & Sherburn, 2010).

4.1.2 C. butyricum *as a probiotic*

C. butyricum presents a multi-faceted opportunity for its use in treating gut-related diseases. Firstly, as a spore-forming anaerobe it is able to survive the acidic environment of the stomach and could therefore be used as a chassis strain for expression and delivery of beneficial therapies to the gut, for a range of conditions. Its potential as a delivery vehicle is already evidenced by the *C. butyricum* MIYAIRI 588 (CBM588) strain. First isolated in 1933 this strain has been used as a probiotic for decades in Japan, illustrating its relative safety, and has already been approved as a novel food ingredient in Europe (The European Commission, 2014). As a probiotic it has demonstrated an acceptable safety profile across several factors, including antimicrobial sensitivity and toxin production (Isa et al., 2016). Whilst its use as a probiotic currently exploits endogenous benefits, it has meant that the strain has been established for use in humans which will enable it to be harnessed for novel applications in CDI treatments.

The potential of *C. butyricum* as a CDI therapy is further elevated by its endogenous mechanisms of resistance against *C. difficile*. *C. butyricum* CBM588 has historically been used as an anti-diarrhoea medicine in Japan and was shown to have a preventive effect for antibiotic-associated diarrhea (AAD) in children (Seki et al., 2003). Since then it has been implicated in resistance against CDI. For example CBM588 treatment in mice significantly improved clinical symptoms associated with CDI by elevation of the immune response (Hagihara et al., 2018), and enhanced the antibacterial activity of fidaxomicin against *C. difficile* by metabolic and immune modulation (Hayashi, Nagao-Kitamoto, Kitamoto, Kim, & Kamada, 2021). Combining this with engineered resistance pathways therefore presents the opportunity for a multi-mechanistic probiotic therapy against CDI, and the ability to engineer these pathways in *C. butyricum* will be explored in this dissertation.

4.1.3 Gene insertion by ACE and orthologous expression

As discussed in 1.3.2, ACE technology can be utilised for insertion of cargo into the genome. The *pyrE* deletion mutant, and theoretically other similar auxotrophic mutants, can be used to select for insertion of application specific elements by concomitant gene repair and restoration of prototrophy. Repair vectors contain the

LHA and RHA as a continuous 1500 bp sequence, homologous to the host, and restore *pyrE* to wildtype. In expression vectors, application specific elements are located between the homology arms, allowing both restoration of *pyrE* to wildtype and gene insertion (Minton et al., 2016). ACE allows insertions of larger cargo than CRISPR technologies in Clostridia due to the smaller sized plasmid backbone required.

Generation of multiple loci for ACE insertion could allow the insertion of multiple genes from one operon, where the size of the operon prevents insertion at one site. These genes will be under the control of an inducible promoter system for controlled, orthologous expression, and this system will be inserted into the genome via the ACE expression vector. This system will comprise of two parts, shown in Figure 4.1, a lactose inducible promoter system controlling expression of *tcdR*, and P_{tcdB} controlling the expression of the chosen *bai* genes. The lactose inducible promoter system is from *Clostridium perfringens* and consists of a gene encoding a transcriptional regulator, BgaR, and a divergent upstream promoter, P_{bgaL}; BgaR is a transcriptional activator of P_{bgaL}, and P_{bgaL} expression is induced by lactose (Hartman, Liu, & Melville, 2011). P_{bgaL} will control the expression of TcdR, a specialised sigma factor of *C. difficile*, which recognises and exclusively allows transcription at P_{tcdB} (Dupuy & Matamouros, 2006; N. P. Minton et al., 2016). The *bai* genes will be under the control of P_{tcdB}, and their expression will therefore be tightly controlled and induced with addition of lactose.

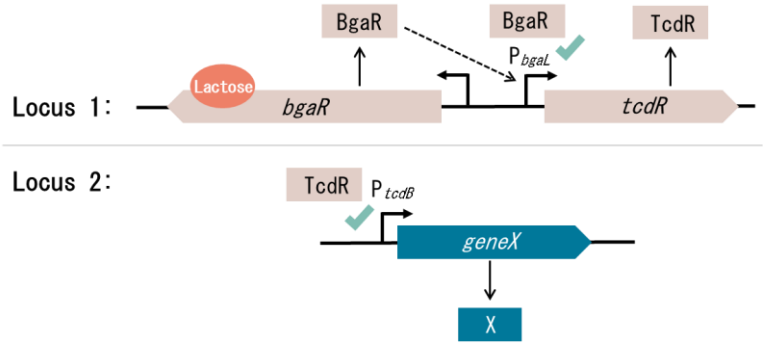


Figure 4.1 Illustration of the lactose inducible P_{tcdB} system for orthogonal expression of exogenous genes. Expression of the *bgaR* gene, a transcriptional activator of P_{bgaL}, is induced by lactose P_{bgaL} is placed upstream of *tcdR*, a sigma factor of P_{tcdB}, which is induced by BgaR. Exogenous genes (*geneX*) are inserted at other chromosomal regions under the control of P_{tcdB}, which are then activated by TcdR.

4.1.4 Objectives

-Characterise and optimise the conditions required for DNA transfer into *C. butyricum*.

-Establish existing clostridial genetic manipulation techniques for use in *C. butyricum*, including RiboCas and Allele-Coupled Exchange.

-Generate and characterise three auxotrophic mutants and produce a final triple knockout strain ready for cargo insertion into the genome at multiple loci.

-Characterise a promoter system suitable for heterologous gene expression in *C. butyricum*.

4.2 General characterisation of *C. butyricum* NCTC 7423 and improvement of DNA transfer

4.2.1 Growth rate

To optimise work with *C. butyricum* NCTC 7423 growth of the wildtype strain was studied (Figure 4.2), according to 2.3.5. Following a lag phase of approximately 5 h, exponential growth occurred until approximately 9 h, reaching a maximum average OD_{600} of 3.71. The OD_{600} remained stable until 24 h.

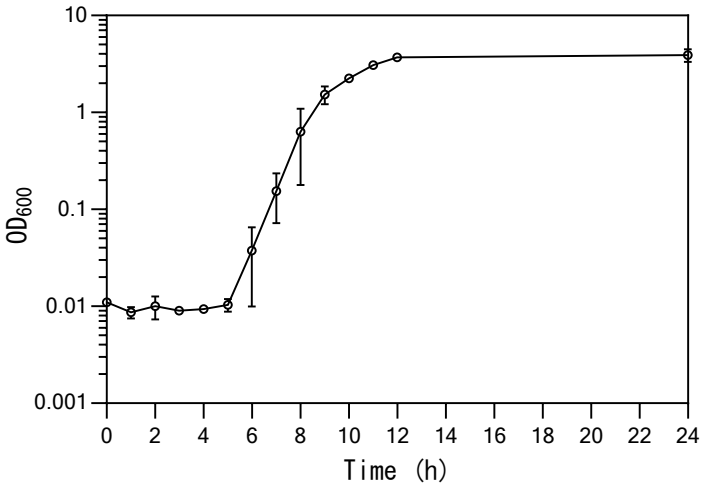


Figure 4.2 *C. butyricum* growth curve. Growth curve of *C. butyricum* WT measured by OD_{600}. Strain grown in liquid 2YTg medium. Data represent mean values of three independent cultures ± SD.

4.2.2 Antibiotic resistance

A minimum inhibitory concentration (MIC) assay was carried out according to 2.3.6, to determine resistance to common antibiotics used in clostridia and those used as markers in the pMTL80000 vector series (Figure 4.3), to assess their use in future plasmids.

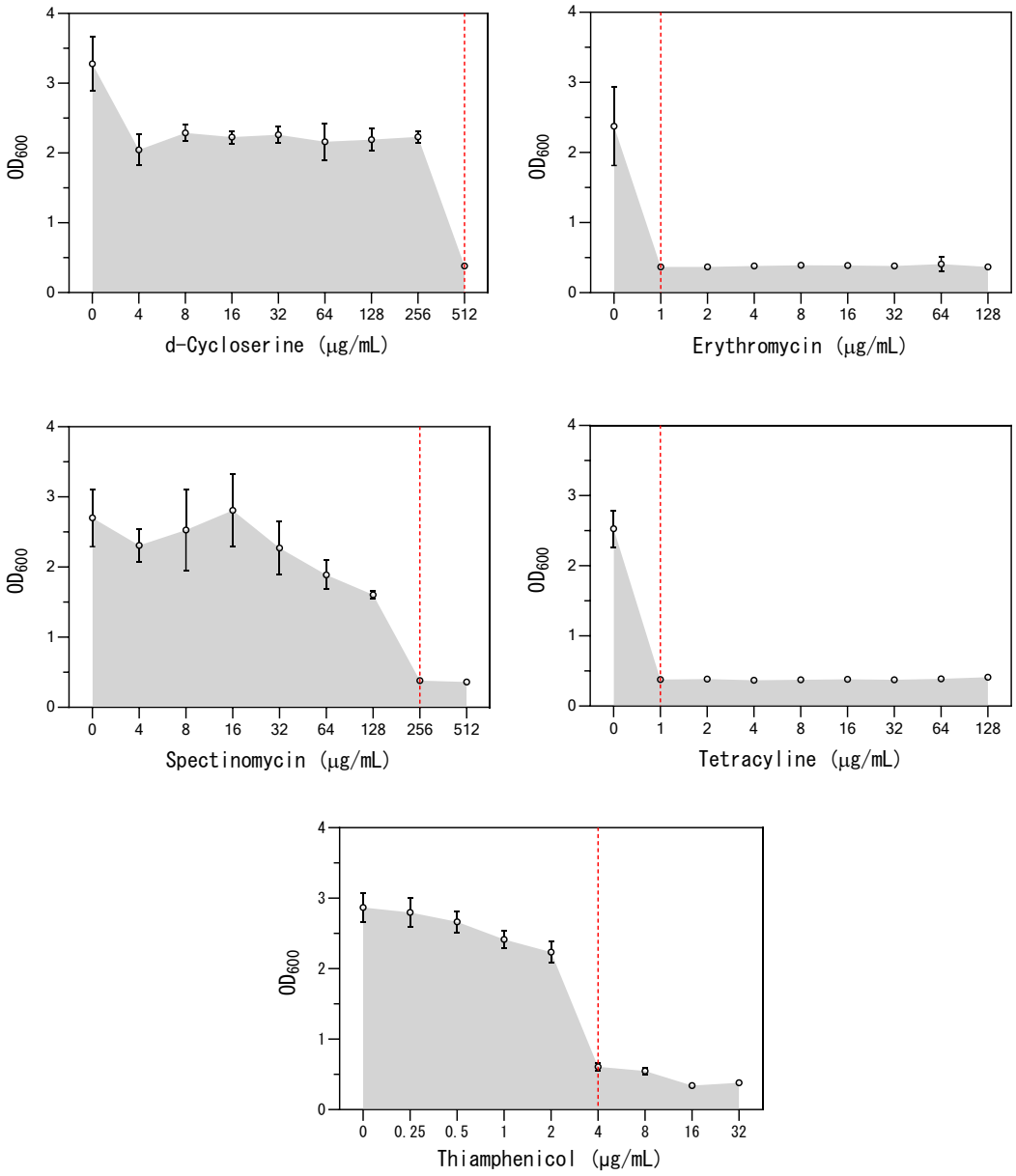


Figure 4.3 Minimum inhibitory concentrations of common antibiotics in *C. butyricum* NCTC 7423. WT strain cultured with a range of concentrations of chosen antibiotics. Starting from a 1 in 10 dilution of overnight culture, growth was measured after 24 h by OD_{600}. Data represent mean values of three independent cultures ± SD.

In accordance with other *Clostridium sp.*, the MIC for d-cycloserine was very high at 512 μg/mL, and will allow its use as a counter selection marker against *E. coli* donors. A dose response was observed for the other four antibiotics tested. The MIC was

estimated at: erythromycin (1 µg/mL), spectinomycin (256 µg/mL), tetracyline (1 µg/mL) and thiamphenicol (4 µg/mL).

4.2.3 Conjugation efficiency

pMTL vectors

To determine the most suitable Gram-positive replicon to maximise DNA transfer in *C. butyricum*, a conjugation efficiency assay (2.3.3) was carried out to compare the pMTL8X151 vectors (Figure 4.4). A 24 h conjugation into *C. butyricum* was carried out using both *E. coli* CA434 and sExpress as the conjugal donor, given the presence of type IV RM systems (4.2.4). Eight transconjugant colonies were re-streaked onto RCM with thiamphenicol selection to confirm plasmid presence, and a glycerol stock was made for three biological replicates of each pMTL vector.

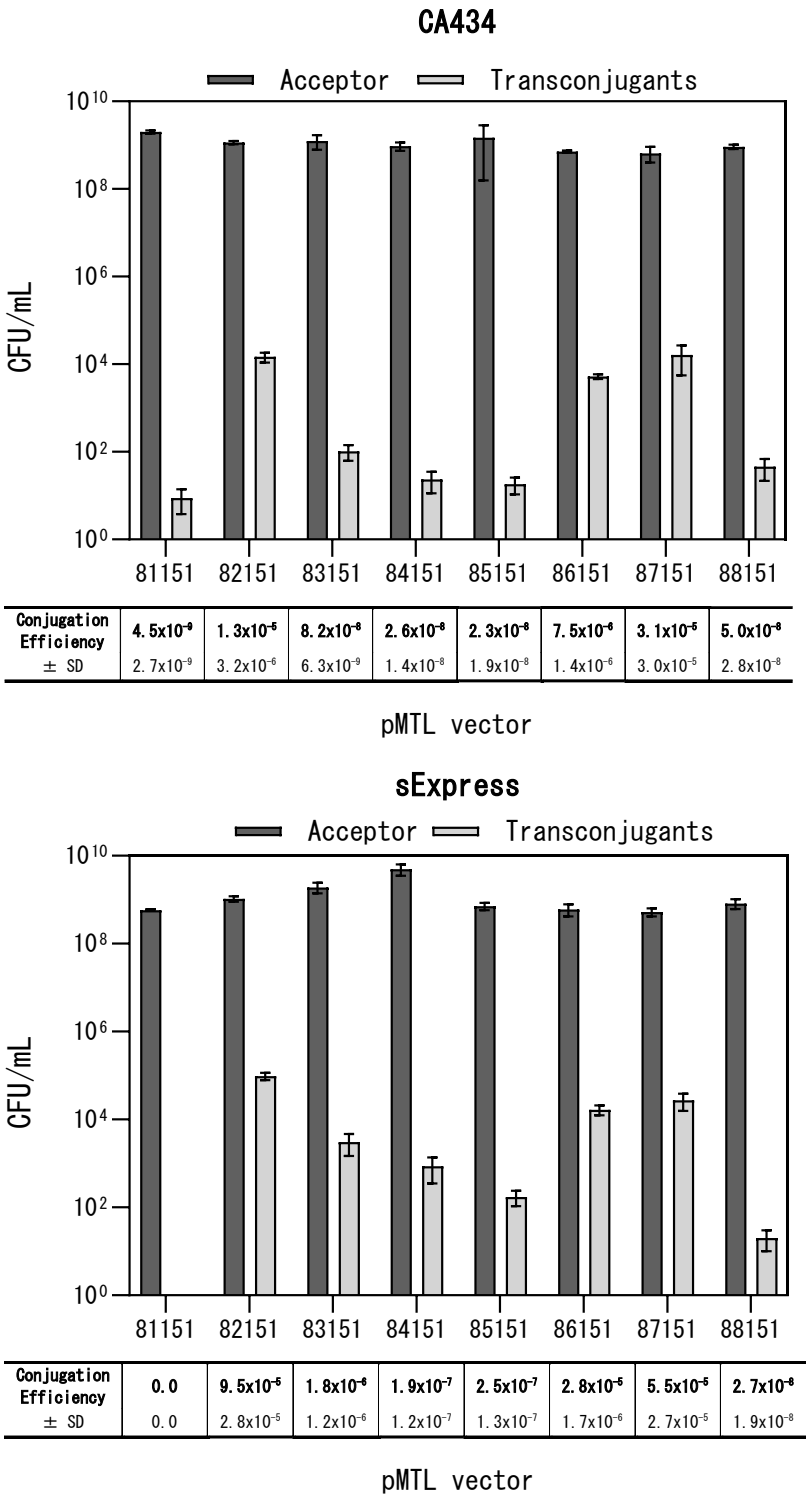


Figure 4.4 Conjugation Efficiency of the pMTL vector series into *C. butyricum*. A comparison of the different Gram-positive replicons (X) in the pMTL vector series, denoted by 8X151. Efficiency calculated as transconjugant colonies (CFU/mL)/ *C. butyricum* conjugal acceptor colonies (CFU/mL). Data represent mean values of the conjugation in technical triplicate ± SD.

Whilst conjugation efficiency appears low across most replicons, the number of transconjugant colonies produced for many was above acceptable levels for further mutagenesis steps. The control pMTL81151 plasmid gave no transconjugants, as expected. The highest efficiency of transfer from *E. coli* CA434 donors was achieved using pMTL82151 and pMTL87151 vectors. Similar results were achieved with sExpress; statistical analysis was carried out to compare the conjugation efficiency of each pMTL plasmid with either the CA434 or sExpress donor, using a two-way ANOVA with Sidak's multiple comparisons test, and there was only a significant increase by using sExpress for pMTL82151 ($p<0.0001$). With two type IV RM systems predicted by Rebase in *C. butyricum* (Figure 4.7), it would be expected that sExpress would improve conjugation efficiency as it does not methylate at Dcm sites and therefore prevents the recognition as foreign DNA by cytosine-specific type IV systems (Woods et al., 2019). However, with the recognition sites of the two type IV systems unidentified, and no sites available for closely related neighbours either, it is not possible to correlate the number of sites in each pMTL plasmid with its conjugation efficiency.

Whilst there was 7-fold increase in conjugation efficiency for pMTL82151 using sExpress, this was not deemed sufficient enough to justify the increased risk of plasmid modifications that have anecdotally been observed, and therefore *E. coli* CA434 was chosen as the conjugal donor for future transfer.

pCB102

Several species of *Clostridium* have been shown to harbour large plasmids and small cryptic plasmids, often encoding factors such as toxins and antibiotic resistance markers (Amy et al., 2018; J. Li et al., 2013; N. Minton & Morris, 1981). This includes *C. butyricum* NCIB 7423, which harbours the megaplasmid pCBU1 (770 kb), and the two smaller plasmids pCB101 and pCB102 at 8 kb and 6 kb respectively (N. Minton & Morris, 1981). Present on pCB102 is the Gram-positive replicon pCB102 which has been utilised in the pMTL vector series for genetic tools, represented by a 3 in the modular plasmid series (for example pMTL83151). The conjugation efficiency of pMTL83151 into *C. butyricum*, shown in Figure 4.4, is low and it is hypothesised that this is due to the duplication of the Gram-positive replicon. As the cell would struggle to maintain both plasmids it is also hypothesised that the native pCB102 could be

cured instead of pMTL83151 in a small number of cases, and the new strain should more easily accept and maintain pMTL83151.

To test the above hypotheses *C. butyricum* WT strain harbouring pMTL83151 was subjected to plasmid loss by culturing in liquid RCM media without selection and then patch plating with and without thiamphenicol selection (2.4.13). Colonies that had been cured of the plasmid were screened by colony PCR for the loss of the native pCB102 replicon (flanking primers pCBU2_F1 and pCBU2_R1), in addition to the presence of pMTL83151 for confirmation of plasmid loss (ColE1_tra_F2 and pCB102_R1) (Figure 4.5).

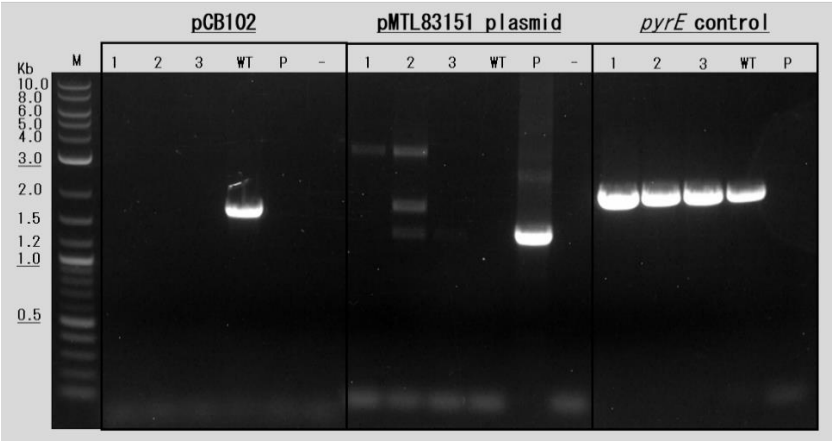


Figure 4.5 Colony PCR screening for loss of pCB102. Agarose gel electrophoresis visualisation of colony PCRs to assess for loss of pCB102 in *C. butyricum*. 1) pCB102 (*C. butyricum*) flanking primers (pCBU2_F1 and pCBU2_R1). 2) pMTL83151 specific primers (ColE1_tra_F2 and pCB102_R1). 3) *C. butyricum* chromosome specific primers (BUT_pyrE_F1 and BUT_pyrE_R3). M = DNA Marker (2-log ladder; NEB). Lanes 1-3 = colonies screened after loss of pMTL83151. WT = *C. butyricum* wildtype. P = pMTL83151 plasmid. - = dH_2O control. Expected fragment sizes (bp): 1) 1667. 2) 1367. 3) 2046.

Colony PCR identified colonies that had been cured of both the pCB102 plasmid and pMTL83151, as demonstrated in Figure 4.5. This ΔpCB102 strain was then assessed for conjugation efficiency of pMTL83151 (Figure 4.6).

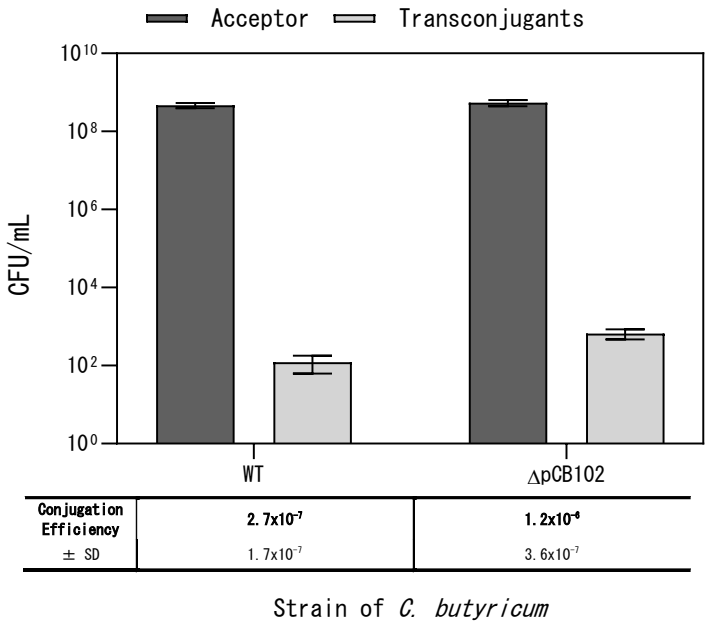


Figure 4.6 Conjugation efficiency in *C. butyricum* ΔpCB102. Conjugation of pMTL83151 into *C. butyricum* wildtype (WT) and ΔpCB102. Efficiency calculated as transconjugant colonies (CFU/mL)/ *C. butyricum* conjugal acceptor colonies (CFU/mL). Statistical analysis was carried out for conjugation efficiency using an unpaired two-tailed Student's t-test; p=0.0453. Data represent mean values of the conjugation in technical triplicate ± SD.

There was a 4.5-fold higher conjugation efficiency of pMTL83151 into the ΔpCB102 strain compared to the WT, and this was statistically significant (unpaired two-tailed Student's t-test; p=0.0453). However, whilst this strain is beneficial for future use of pMTL vectors with the pCB102 replicon, the improved conjugation efficiency is of a similar magnitude to other replicons tested in Figure 4.4. Moreover, there may be loss of beneficial genes encoded by pCB102, for example Cbu04g_42380 encoding a thiol reductase thioredoxin. The bacteriocin thought to be responsible for the probiotic effect of *C. butyricum* CBM588 is also present on pCB102 (Nakanishi & Tanaka, 2010). It was therefore decided that the small benefit from ΔpCB102 strain did not outweigh cost given the alternatives available, with the preferred strategy to continue with more efficient replicons in the WT strain.

4.2.4 Restriction Modification systems

To identify possible barriers to DNA transfer in *C. butyricum* the Rebase Genomes database was consulted (http://tools.neb.com/genomes/view.php?view_id=82891, Genbank CP040626), revealing two putative RM systems; 2 type II and 2 type IV (Figure 4.7). Two proteins were identified as constituents of a type II system, a BREX-1 system adenine-specific DNA-methyltransferase Pglx (CbuD10702ORF11735P) and a DNA adenine methylase (CbuD10702ORF12535P). Two proteins were also identified as constituents of a type IV system, a DEAD/DEAH box helicase methyl-directed restriction enzyme (CbuD10702ORF7335P), and a DUF3427 domain-containing methyl-directed restriction enzyme (CbuD10702ORF7435P).

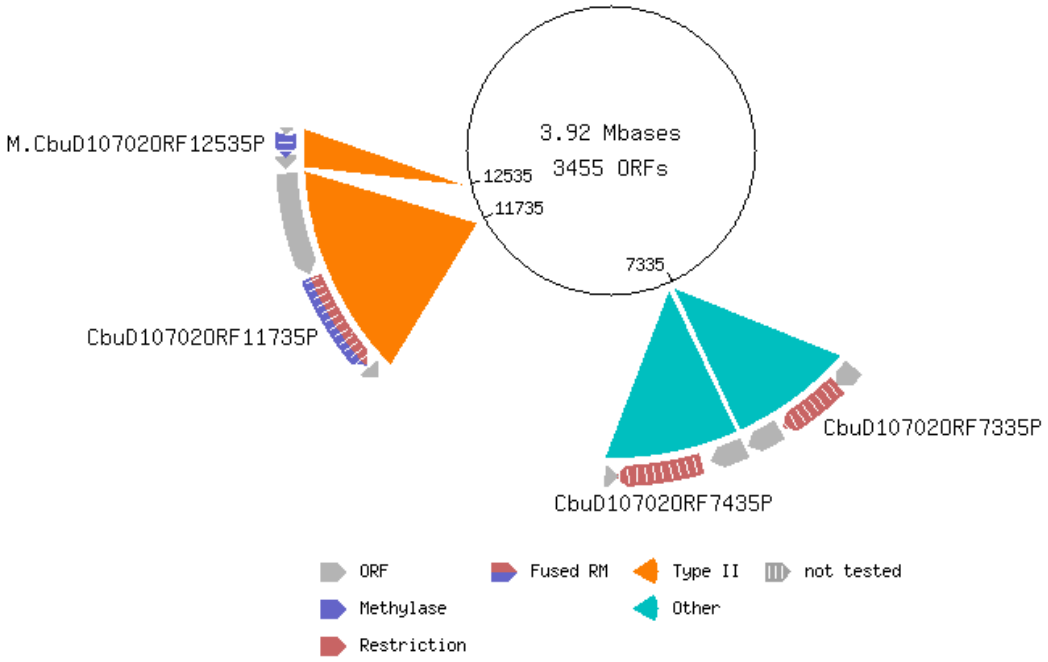


Figure 4.7 Restriction modification (RM) systems present in *Clostridium butyricum* NCTC 7423. Summary data of the RM systems present, Genbank number CP040626. Data retrieved from Rebase Genomes (http://tools.neb.com/genomes/view.php?view_id=82891).

As RM systems can reduce the efficiency of plasmid transfer it was decided to attempt to create an RM system KO using the RiboCas system, not previously used in *C. butyricum*. As the sExpress conjugal donor is available to circumvent type IV systems, it was chosen to target the *pglx* gene of the type II system, predicted as a fused

restriction methylase where a single polypetide can employ both restriction and modification functions. Further analysis of its adjacent genes in the genome (Figure 4.8) revealed a structure classic to the BREX (bacteriophage exclusion) Type I system, a novel phage defense system (Goldfarb et al., 2015).

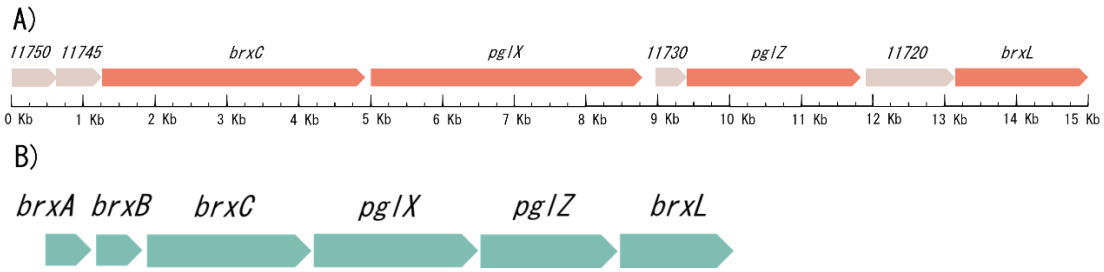


Figure 4.8 BREX 1 system. A) Putative BREX 1 system of *C. butyricum* based on genome annotations. Annotated genes = orange. Genes with unknown products = beige, labelled with locus tag (FF104_label). FF104_11750 = DUF1819 domain-containing protein, a putative inner membrane protein. FF104_11745 = DUF1788 domain-containing protein. FF104_11730 and 11720 = hypothetical proteins. B) Common structure of BREX system subtype 1 (Goldfarb et al., 2015). BrxA and BrxB = function unknown. BrxC = ATPase. PglX = DNA methylase. PglZ = phosphatase. BrxL = protease.

Construction of RiboCas vectors

Firstly, to allow subsequent conjugation into *C. butyricum* the Gram-positive pCB102 replicon of pRECas1_MCS was substituted to the p19 of pMTL87151. This was achieved by digestion with FseI and AscI of both vectors, and ligation of p19 into pRECas1_MCS. This generated pRECas_p19_MCS to use as the backbone for insertion of the application specific module elements.

To generate a *pglx* gene deletion with BM8 bookmark, the vector pRECas_p19_pglx was generated by hifi assembly of the fragments listed in Table 4.1. Assembly was confirmed by colony PCR (Cas9scr1R and p19_R1), followed by Sanger sequencing of the whole plasmid. The final plasmid was then transformed into *E. coli* CA434.

Table 4.1 Fragments used to generate pRECas_p19_pglx. Primers used to generate the RiboCas pRECas_p19_pglx with application specific elements for a *pglx* deletion, and a BM8 bookmark. Assembled by hifi PCR. LHA = Left homology arm. RHA = Right homology arm.

Fragment	Primers	Template
pRECas_p19 1 of 2	pglXKO_RE1a_F1 pglXKO_RE1a_R1	pRECas_p19
pRECas_p19 2 of 2	pglXKO_RE1b_F1 pglXKO_RE1b_R1	pRECas_p19
LHA (835 bp)	pglXKO_LHA_F1 pglXKO_LHA_R1	*C. butyricum* DNA
RHA (787 bp)	pglXKO_RHA_F1 pglXKO_RHA_R1	*C. butyricum* DNA
Guide RNA SEED = GATATTCTATAGATGCAGAA	pglXKO_sg_F1 pglXKO_sg_R1	Primer dimer

The RiboCas vector pRECas_p19_pglx+ was also generated to complement the *pglx* deletion, targeting the BM8 bookmark. The vector was generated by hifi assembly of the guide RNA fragment (PglxComp_sg_F1 and PglxComp_sg_F1, template pRECas_p19_pglx) and the editing template fragment (Pglxcomp_pglx_F1 and Pglxcomp_pglx_R1; template *C. butyricum* DNA; editing template of 430 bp LHA and 417 bp RHA flanking the WT *pglx*) assembled into the backbone pRECas_p19_MCS digested with SalI and AscI. Insertion was confirmed by colony PCR (Cas9scr1R and p19_R1), followed by Sanger sequencing. The final plasmid was then transformed into *E. coli* CA434.

Knockout and complementation

The final pRECas_p19_pglx was transferred into *C. butyricum* by a 24 h conjugation, selecting for transconjugants with thiamphenicol and confirming with a re-streak. Colonies were then streaked onto RCM plates with thiamphenicol and 5 mM theophylline and colonies were then screened for the *pglx* deletion by colony PCR (BUT_pglxscr_F1 and BUT_pglxscr_R1) (Figure 4.9A). A successful deletion was confirmed by Sanger sequencing and the plasmid was then cured by one round of patch plating (2.4.13). The KO genotype was reconfirmed by colony PCR to make the final strain, *C. butyricum Δpglx*.

To generate the complemented strain pRECas_p19_pglx+ was transferred into *C. butyricum ⊿pglx* by conjugation, and the same process was followed. Colonies were screened for gene repair by colony PCR (BUT_pglxscr_F1 and BUT_pglxscr_R1) (Figure 4.9B). The successful repair (lane 2) was confirmed by Sanger sequencing, the plasmid cured and genotype then reconfirmed by colony PCR to generate the strain *C. butyricum pglx+*.

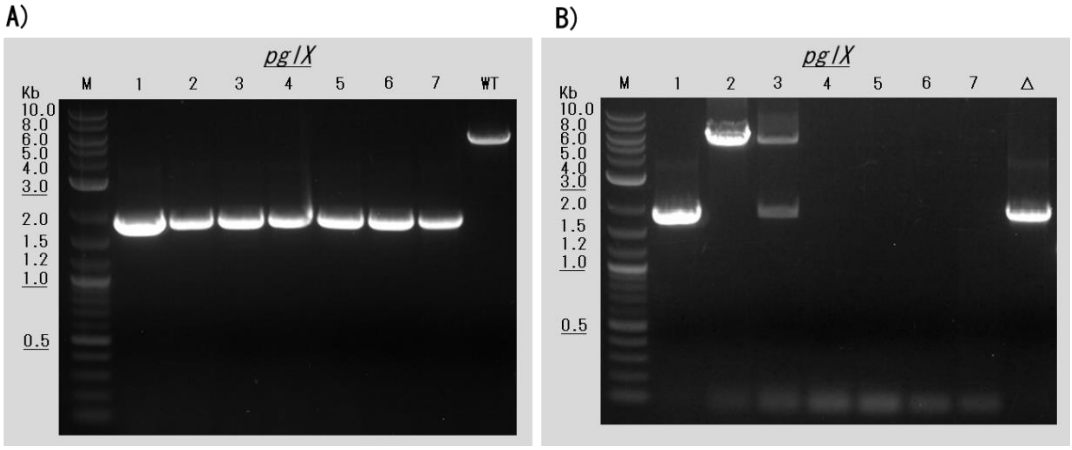


Figure 4.9 Colony PCR screening for *pglX* deletion (A) and complementation (B). Agarose gel electrophoresis visualisation of colony PCRs with flanking primers (BUT_pglxscr_F1 and BUT_pglxscr_R1) to assess for RiboCas deletion of *pglx* (A) and subsequent repair of Δ*pglx* to wildtype (B) in *C. butyricum*. WT= wildtype, Δ = *⊿pglx*, M = DNA Marker (2-log ladder; NEB). Numbered lanes = transconjugant colonies plated on thiamphenicol selection with theophylline inducer. Expected fragment sizes (bp) WT = 5491, Δ = 1859.

Conjugation efficiency

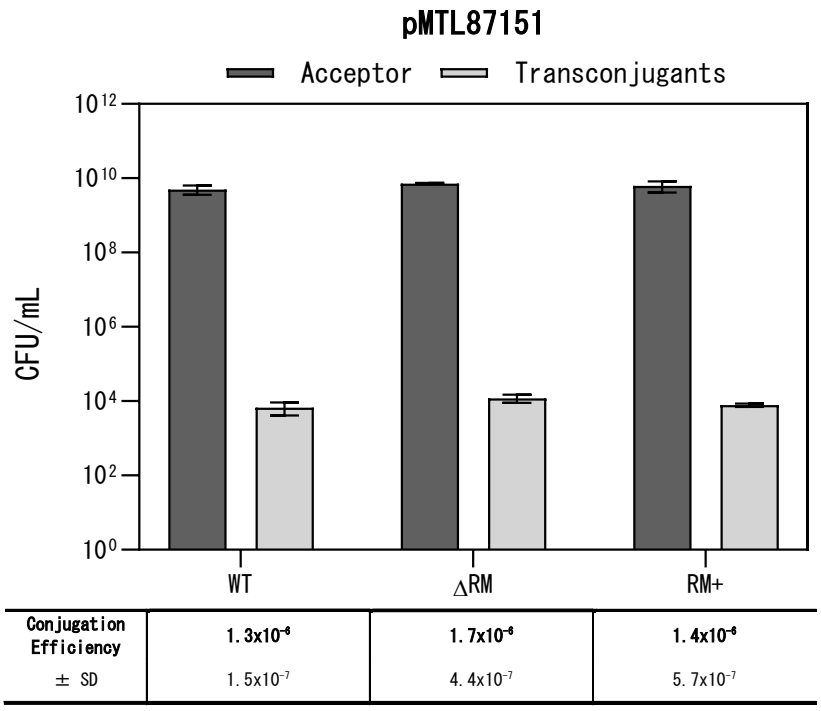
To assess if the RM KO strain allowed for improved DNA transfer a conjugation efficiency assay was carried out; efficiency should increase in the KO strain for plasmids containing unmethylated recognition sites for the *pglx* type II RM system. Unfortunately, the Rebase data does not allow prediction of the recognition site, therefore Pglx proteins were assessed from closely related organisms to allow an estimate. The most highly related Pglx sequences in the Rebase database with recognition sites were those from *C. perfringens* FDAARGOS_904, CPI 75-1 and

NCTC10578, all sharing a 65% identity (E- value=0) with Pglx of *C. butyricum*. These type II methyltransferases produce 6-methyladenosine with a recognition site of GANGAGY. This was used as a proxy recognition site for *C. butyricum*, and the pMTL vector series was assessed for its occurrence (Table 4.2).

Table 4.2 Assessment of pMTL vectors for the Pglx recognition site. Number of GANGAGY sites in each vector.

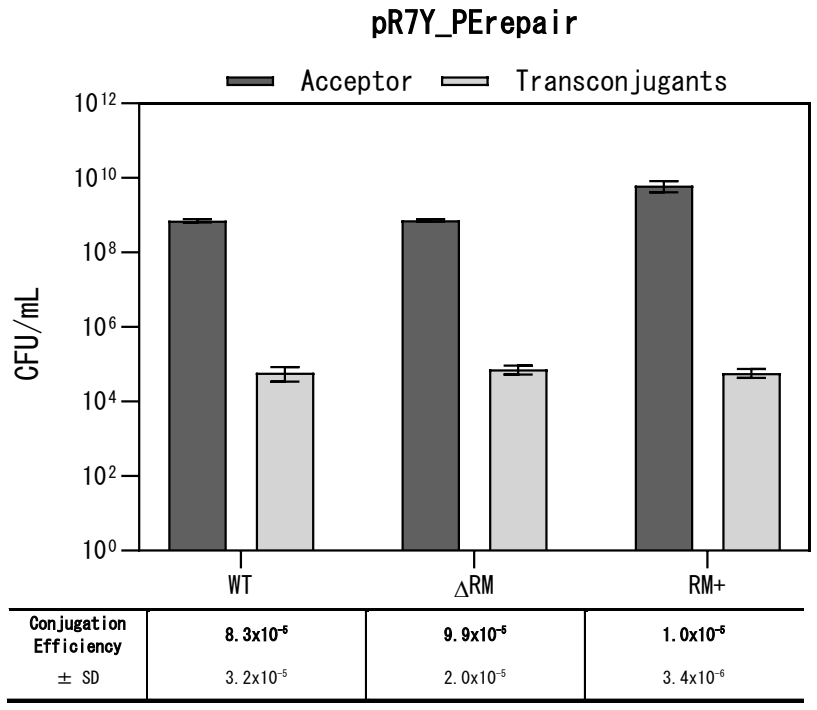
Plasmid	Recognition sites
pMTL82151	7
pMTL83151	7
pMTL84151	6
pMTL85151	6
pMTL86151	7
pMTL87151	6
pMTL88151	7

The pMTL87151 vector was chosen to assess as it had been shown to produce consistent transfer by conjugation previously (Figure 4.4). A further vector, pR7Y_PErepair, was also chosen for comparison as it has the same backbone and is a similar size, but contains more recognition sites (11 sites). Both chosen vectors were transformed into *E. coli* sExpress, conjugated over 24 h into the *C. butyricum* RM strains and efficiency calculated (Figure 4.10).



Conjugation Efficiency ± SD	WT	ΔRM	RM+
	1.3×10^{-6}	1.7×10^{-6}	1.4×10^{-6}
	1.5×10^{-7}	4.4×10^{-7}	5.7×10^{-7}

Strain of *C. butyricum*



Conjugation Efficiency ± SD	WT	ΔRM	RM+
	8.3×10^{-5}	9.9×10^{-5}	1.0×10^{-5}
	3.2×10^{-5}	2.0×10^{-5}	3.4×10^{-6}

Strain of *C. butyricum*

Figure 4.10 Conjugation efficiency in *C. butyricum Δpglx*. Conjugation of two plasmids, pMTL87151 and pR7Y_PErepair into *C. butyricum* strains wildtype (WT), *Δpglx* (ΔRM) and complemented *pglx* (RM+). Efficiency calculated as transconjugant colonies (CFU/mL)/ *C. butyricum* conjugal acceptor colonies (CFU/mL). Statistical analysis for conjugation efficiency was carried out using one-way ANOVA with Dunnett's test for multiple comparisons against the WT control. pMTL87151; ΔRM (p=0.5337), RM+ (p= 0.9665). pR7Y_PErepair; ΔRM (p=0.5886), RM+ (p= 0.0107). Data represent mean values of the conjugation in technical triplicate ± SD.

Unexpectedly there was no improvement in conjugation efficiency in the *Δpglx* strain compared to the WT for either plasmid tested. There was no significant difference between conjugation efficiency in any of the three strains (WT, ΔRM and RM+) for pMTL87151 (one-way ANOVA with Dunnett's test), and for pR7Y_PErepair there was no difference between WT and ΔRM, but there was for WT and RM+ (p=0.0107).

The lack of improvement in conjugation efficiency is surprising. Firstly, the use of the sExpress conjugal donor should have overcome the type IV systems, and was shown to do this by improved conjugation efficiency (Figure 4.4). Moreover, although the exact recognition sequence was not identified for the type II system, using the proxy sequence from *C. perfringens* should give a good estimate. One would expect plasmids containing 7 and 11 recognition sites to experience sufficient barriers to conjugation that an improvement would be observed in the ΔRM. One would also expect there to be a large improvement in transfer for pR7Y_PErepair compared to pMTL87151 due to increased presence of recognition sites, rather than no improvement for either plasmid. Restriction mutants have been generated in other clostridia, for example a type I *Δhsdr1Δhsdr2Δhsdr3* in *Clostridium saccharobutylicum* (Huang, Liebl, & Ehrenreich, 2018), and a ClosTron insertional mutant in a type II endonuclease in *C. acetobutylicum* (Dong, Zhang, Dai, & Li, 2010). Transfer of unmethylated plasmids was improved in both of these strains compared to the WT, and one would therefore expect the same result in *C. butyricum Δpglx*. Moreover, a *Lactobacillus casei Δpglx* strain exhibited a higher rate of transformation of a plasmid containing the recognition site compared to the WT (Hui et al., 2019).

There is, however, conflicting evidence on the role of *pglx* in preventing plasmid transfer. BREX family defence systems are immunity systems against bacteriophage infection via a largely uncharacterised mechanism (Luyten et al., 2022). On integration of a BREX type 1 system of *Bacillus cereus* into the *B. subtillus* genome, transfer of phages was reduced by > 5 orders of magnitude, compared to no reduction or mild reduction (~1 order of magnitude) for plasmid transfer. The *⊿pglx* strain lost defence against all 10 phages tested (Goldfarb et al., 2015). Moreover, the authors suggest that the BREX mechanism of action is not consistent with that of RM systems. Although BREX mediated methylation has been shown to allow identification of non-host DNA (Gordeeva et al., 2019), the *pglx* deletion by Goldfarb et al. (2015) was not toxic to cells. Additionally, whilst the BREX system limited phage propagation, this was not by degradation of phage DNA (Goldfarb et al., 2015).

Given that the mechanism of BREX 1 mediated immunity appears to differ from the classical RM systems, the lack of expected improvement in transfer efficiency is perhaps not so surprising. Future work should establish the overall activity of RM systems in *C. butyricum*, and also that of the BREX 1 system. Methylation patterns could be identified in the ΔRM strain by PacBio sequencing to confirm a loss of *pglx* activity compared to the WT, and digestion of plasmid DNA by whole cell extracts of both strains compared to establish whether this is a mechanism of transfer resistance by *pglx*. Digestion of just the WT with an unmethylated and methylated plasmid could also reveal barriers by RM systems in general.

Overall, the *⊿pglx* strain does not offer the increased conjugation efficiency as had been hoped. However, given the adequately high efficiency achieved for pMTL plasmids in the WT, this should not hinder transfer capabilities for genetic manipulations. Moreover, the *⊿pglx* strain could be studied further with regards to the BREX 1 system defences in *C. butyricum* against phage. If the strain is studied further whole genome sequencing should be carried out to identify any genetic variations that have arisen during the mutagenesis process.

4.2.5 Plasmid stability

Assay

Whilst a high conjugation efficiency of a Gram-positive replicon is extremely beneficial, this also must be balanced with the ability to cure the plasmid after its purpose has been fulfilled in the cell. The four Gram-positive replicons with the highest transfer rate were chosen to assess their stability in *C. butyricum*: pMTL82151, pMTL83151, pMTL84151 and pMTL87151 (Figure 4.11). The strains underwent a plasmid stability assessment (2.4.14).

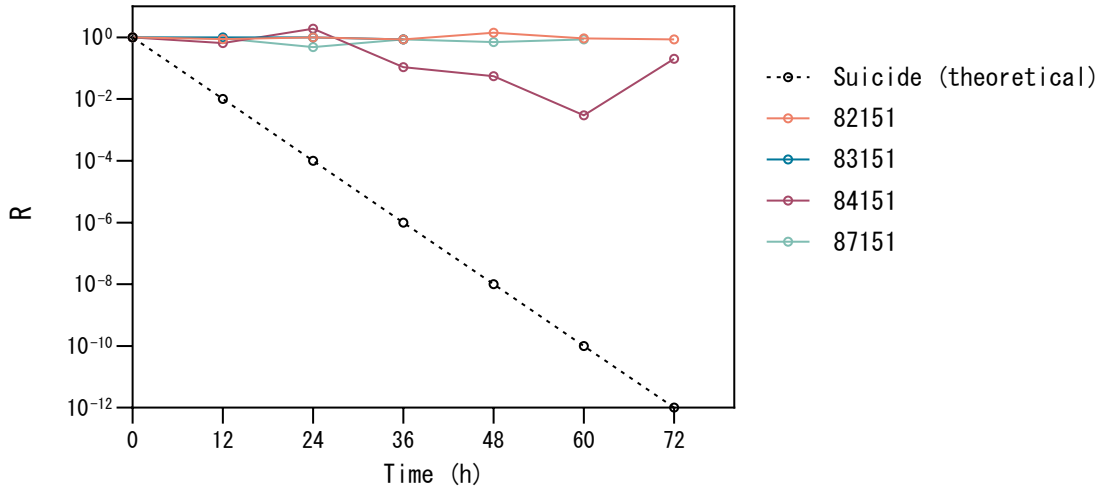


Figure 4.11 Plasmid stability of pMTL vectors in *C. butyricum* without antibiotic selection. *C. butyricum* WT harbouring various pMTL vectors was cultured in liquid RCM media without selection, and repeatedly sub-cultured every 12 h. Samples were taken and plated to calculate the CFU/mL. R = plasmid loss as a function of time, measured by CFU/mL with and without thiamphenicol selection. Data represent mean values of three independent cultures ± SD.

The assay demonstrated that all four plasmids studied were fairly stable, with pMTL84151 being slightly less stable than the other three. Whilst the assay could suggest that plasmid loss could impede their suitability of use, all four plasmids were lost after one round of curing by patch plating (2.4.13), which is practical for their further use.

There were difficulties in this assay that resulted in unexpected data, questioning the validity of the results. This is highlighted in the incomplete data set for pMTL83151, and an 'increase' in stability of pMTL84151. After approximately two rounds of sub-culturing, growth of colonies on RCM agar with and without thiamphenicol revealed two colony morphologies, one classical of the WT (small, circular and convex colonies with an entire margin) and a new morphology (circular but flatter with a lobate margin). The assay was repeated in triplicate and the new morphology appeared each time. It was not specific to a pMTL vector and appeared to increase in frequency with sub-culturing with the classical morphology decreasing. It varied in the sub-culture number in which it first appeared, but it was always within the first 1-3 subcultures. The new morphology was eliminated as a contaminant; colony PCR was carried out for the 16S region (Univ-0027-F and Univ-1492-R) and sent for Sanger sequencing. This confirmed the colonies as *C. butyricum*.

New morphology

Whole genome sequencing and analysis was performed (2.6.3) for the new colony morphology to assess for possible SNPs that may give rise to the new morphology.

Table 4.3 Resequencing analysis of *C. butyricum* morphology mutant strain. SNPs and other genomic variants, unique to those identified in the WT strain (identified in Table 4.8), in the *C. butyricum* NCTC 7423 morphology mutant. Gene and function identified according to annotations of the reference (Ref.) genome (GCA_014131795.1) (IG=intergenic). For variants, A=sequence at locus, C=coverage of reads, F=frequency in reads. Amino acid changes arising from variants are listed (AA). Blank cells denote no change from reference sequence.

				New morphology			
Gene (FF104)	**Function**	**Position**	**Ref.**	**A**	**C**	**F**	**AA**
IG		1573194	G	**T**	117	100	IG
IG		2257309	A	**C**	69	100	IG
14125	P-type ATPase	3122115	C	**A**	66	100	Arg58Met
20205	CPBP metalloprotease	516784	C	**A**	60	100	Val243Phe
20360	glycerol dehydrogenase	555607	C	**A**	68	100	Gly271Val

A total of 5 variants were identified in the mutant colony morphology strain (Table 4.3) of which all were SNPS, located throughout the whole genome and one large plasmid. Of the 5 variants, 3 were within coding regions and consist of 3 missense mutations. As the 5 variants have also been identified in further strains that do not

demonstrate the new colony morphology (*ΔpyrE* strain B, Table 4.9) they are unlikely to be responsible.

As the new morphology cannot be explained by DNA analysis, it is likely to be epigenetic. For example, *C. difficile* has been shown to form two distinct colony morphologies, rough and smooth, and this has been attributed to phase variation of a signal transduction signal, named colony morphology regulators RST. In addition to morphology it has also been shown to effect motility and virulence (Garrett et al., 2019). Future work could explore the possibility of phase variation in *C. butyricum*, for example through carrying out RNAseq to explore the transcriptome of the morphology mutant.

4.2.6 Spore characterisation

The final aim for this project would be to use *C. butyricum* to allow expression of exogenous genes in the gut, and therefore the spore is the important vehicle of delivery that would allow this. Because of this it was decided to carry out characterisation of sporulation and germination in preparation for future studies.

Spore identification and preparation

To test spore properties it was first necessary to generate pure spore preparations, and ensure that spores can be produced in the culture conditions used. Spores were prepared according to 2.3.10, and visualisation under the microscope confirmed spore production.

Germination

To allow future characterisation of germination the required conditions needed to be established. Firstly, various heat treatments were assayed to ensure that they would be sufficient to kill vegetative cells and allow germination alone to be measured (2.3.16)(Figure 4.12). Vegetative cells from a four-hour culture in RCM liquid media were subjected to 20 or 30 mins at 65°C or 80°C.

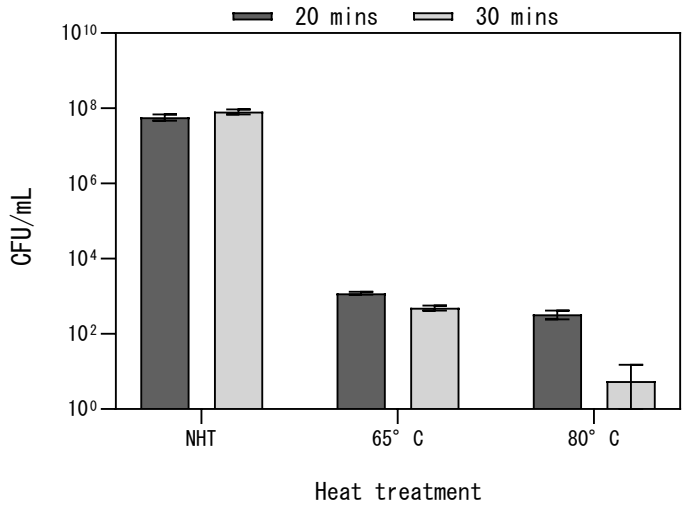


Figure 4.12 Heat treatment of *C. butyricum* vegetative cells. *C. butyricum* wildtype was inoculated to a starting OD_{600} of 0.05 from an overnight dilution series, and cultured in RCM liquid medium. Samples were taken after 4 h and heat treated at 65°C or 85°C (or a non-heat treated control, NHT) for 20 or 30 mins. Samples were plated on RCM media and the CFU/mL calculated after 24 h. Data represent mean values of three independent cultures ± SD.

There was a clear decrease in the CFU/mL for both temperatures compared to the non-heat treated control. At 65°C the CFU/mL was reduced to 0.0021% and 0.00067% of the non-heat treated control for 20 and 30 mins respectively. At 80°C the CFU/mL was reduced to 0.00057% and 0% of the non-heat treated control for 20 and 30 mins respectively. Whilst there was an almost complete reduction in CFU/mL for three of the heat treatments, there were some viable colonies. This could suggest that either the heat treatment isn't sufficient, which is unlikely given that similar conditions are used successfully for other *Clostridia* including *C. difficile* (Ingle, 2017), or that spores are present. The sample taken at 4 h was confirmed to contain a limited number of spores by microscopy. These are unlikely to be carry over from the overnight inoculum as a dilution series was used and visually checked for lack of spores, so this may indicate a rapid onset of spore production by *C. butyricum*.

It was then decided to assess the heat treatments with regards to germination, and to ensure germination could be assayed. Research into the germinants of *C. butyricum* is not extensive but have been previously proposed as L-Cysteine, sodium bicarbonate

and glucose (Sarathchandra, Barker, & Wolf, 1973). Although these are common amino acids and minerals, initial germination assays (2.3.16) sampling a 5-day culture and plating onto RCM following heat treatment did not produce any colonies. Because of this it was decided to also test other media: BHIS and RCM + fetal bovine serum (10%) (Gibo), which has been shown to assist in germination of *Clostridium novyi* (Sundaresan et al., 2023). Results were inconsistent in yielding colonies. It was therefore decided to assay germination of the spore preparation, to ensure presence of spores at a high concentration (Figure 4.13).

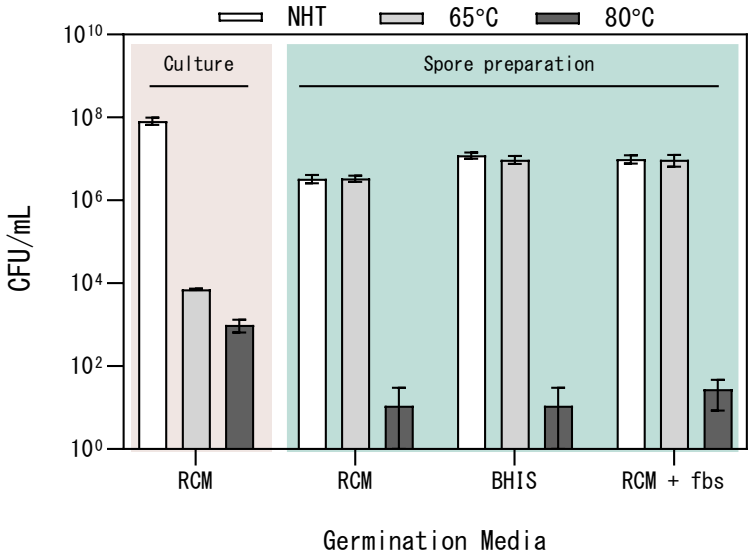


Figure 4.13 Germination of *C. butyricum* on various media. A *C. butyricum* spore preparation was heat treated at 65°C or 85°C (or a non-heat treated control, NHT) for 20 mins. Samples were plated on media: RCM, BHIS or RCM + fetal bovine serum (10%), and the CFU/mL calculated after 24 h. Data represent mean values of three independent spore preparations ± SD. A sample from a *C. butyricum* 5-day culture was also assayed (culture) and plated onto RCM. Data represent mean values of three independent cultures ± SD.

Assaying the spore preparation instead of sampling from the culture appeared to overcome previous issues, and germination was observed on all three media tested, including RCM. This confirms that germinants are present in the media. At 65°C the CFU/mL was similar to the NHT control for all media (102%, 78% and 95% for RCM, BHIS and RCM+fbs respectively), demonstrating that the heat treatment is not required to initiate germination. At 80°C the CFU/mL was below the lower limit of

quantification for all media, suggesting that this temperature is detrimental to the survival of the spores, thus preventing germination.

One of the assay attempts at studying germination directly from a culture sample is also included in Figure 4.13. In this instance germination was achieved on RCM plates and the CFU/mL was reduced to 0.0087% and 0.00012% of the non-heat treated control for 65°C and 80°C mins respectively. The far smaller percentage of germination perhaps reflects the lower concentration of spores found in the culture compared to the spore preparation, and perhaps explains why in some repeats of the assay no germination is observed; the CFU/mL is close to the lower limit of quantification.

Overall the assays determine that the *C. butyricum* prepared spores are able to germinate on the standard RCM media. If vegetative cells are present the sample should be heat treated at 65°C but not 80°C to ensure spore germination and killing of vegetative cells. However, assays to measure spore germination straight from a culture, for example sporulation or germination over time, are not possible due to the variability of germination, which does limit study of germination to spore preparations only. This is an area which could be explored further, but it was deemed outside the scope of this study.

4.3 Generating a triple auxotroph strain

To allow for insertion of exogenous genes into the chromosome of *C. butyricum* auxotrophic mutations need to be generated for subsequent repair and concomitant cargo insertion by ACE.

4.3.1 Generation of auxotrophic mutants using RiboCas

The *C. butyricum* genome was searched for appropriate candidates to generate auxotrophic mutants. Known metabolic pathways were searched for in the genome sequence to identify essential genes located at the end of an operon associated with metabolism; this allows the mutant to be generated without downstream polar effects. Along with the well characterised *pyrE* (pyrimidine metabolism), *purD* (purine

metabolism) and *hisI*(/*hisE*) (L-histidine synthesis) were chosen, shown in Figure 4.14. It was decided to first generate the individual gene truncations with RiboCas to assess for the auxotrophic phenotype.

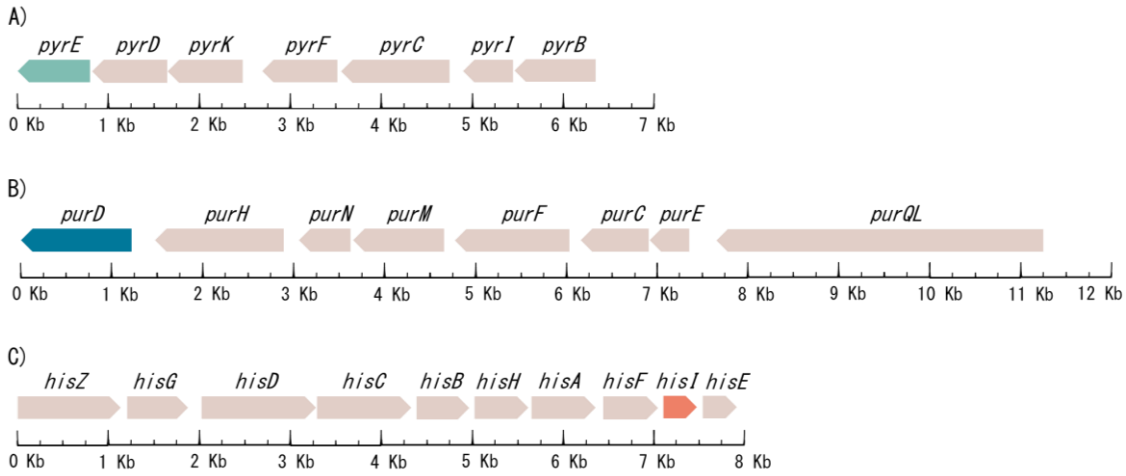


Figure 4.14 Genes chosen to generate a triple auxotroph in *C. butyricum*. Genes essential in amino acid and vitamin metabolism chosen to truncate. A) *pyrE*, Orotate phosphoribosyltransferase involved in UMP biosynthesis, part of pyrimidine metabolism. B) *purD*, Phosphoribosylamine-glycine ligase involved in IMP biosynthesis, part of purine metabolism. C) *hisI*, histidine biosynthesis bifunctional protein, involved in L-histidine synthesis. Genes at the end of the operon were chosen, highlighted, to allow deletion without secondary effects.

Vector construction

Following its success to generate a *pglx* mutant (4.2.4) RiboCas vectors were designed to generate gene truncations at the chosen loci to inactivate gene function, deleting a portion of the gene adjacent to the stop codon (Figure 4.15). Truncation size was chosen to allow deletion of gene function whilst also preventing a functioning gene in the complementation vector in *E. coli* and limiting size of the fragment and plasmid.

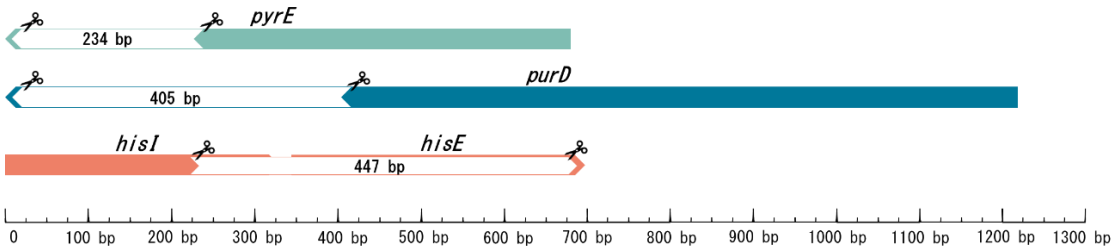


Figure 4.15 Truncations of chosen genes in *C. butyricum* to generate auxotrophic mutants: *pyrE*, *purD* and *hisI*. RiboCas editing templates designed to create truncations at the 3' end before the stop codon.

The RiboCas vector for each gene truncation was generated by a digestion of pRECas_p19_MCS with SalI and AscI and a three-way ligation with the sgRNA and editing template. The sgRNA fragment was generated with the chosen SEED fragment by primer dimer PCR and digested with SalI and AatII. The editing template was generated by SOEing PCR of the LHA and RHA and digestion with AatII and AscI. Fragments were generated using the primers and templates outlined in Table 4.4. Insertion was confirmed by colony PCR (sgRNA_F4 and p19_R1) and Sanger sequencing, to give the final plasmids, pRECas_p19_pyrE, pRECas_p19_purD and pRECas_p19_hisI. Final plasmids were transformed into *E. coli* CA434.

Table 4.4 Fragments used to generate pRECas_p19_pyrE/purD/hisI. Primers used to generate required fragments for insertion into the RiboCas pRECas_p19_MCS backbone by digestion and ligation. Final RiboCas vectors designed to generate truncations in chosen genes.

Target	Fragment	Primers	Template
pyrE	sgRNA (SEED= GATAGAATGGAAAGAGGCCA)	BUT_PyrEtrunc_sg_F1 BUT_RE1_sgRNA_R1	Primer dimer
	LHA (449 bp)	BUT_PyrEtrunc_LF2 BUT_PyrETrunc_LR1	*C. butyricum* DNA
	RHA (499 bp)	BUT_PyrETrunc_RF1 BUT_PyrEtrunc_RR2	*C. butyricum* DNA
purD	sgRNA (SEED= GCTGGAGCTAAATTCGAAGA)	BUT_PurDtrnc_sg_F1 BUT_RE1_sgRNA_R1	Primer dimer

	LHA (526 bp)	BUT_PurDtrnc_LF1 BUT_PurDtrnc_LR1	*C. butyricum* DNA
	RHA (504 bp)	BUT_PurDtrnc_RF1 BUT_PurDtrnc_RR1	*C. butyricum* DNA
hisI	sgRNA (SEED= TTAGTATTGGTAGAGCAAAA)	BUT_HisItrnc_sg_F1 BUT_RE1_sgRNA_R1	Primer dimer
	LHA (528 bp)	BUT_HisItrnc_LF1 BUT_HisItrnc_LR1	*C. butyricum* DNA
	RHA (599 bp)	BUT_HisItrnc_RF1 BUT_HisItrnc_RR1	*C. butyricum* DNA

Gene truncations

Each of the final RiboCas vectors was conjugated separately into *C. butyricum* WT over 24 h, and transconjugants selected for with thiamphenicol. Transconjugants were then re-streaked onto thiamphenicol to confirm plasmid presence, and then onto theophylline (5mM) to induce CRISPR. Colony PCR screening (Figure 4.16) for each gene was then carried out to assess for successful truncation: *pyrE* (BUT_pyrE_F1 and BUT_pyrE_R2), *purD* (BUT_PurD_F1 and BUT_PurD_R1) and *hisI* (BUT_HisI_F1 and BUT_HisI_R1).

Successful truncations were achieved for all three chosen genes. Truncations were confirmed by Sanger Sequencing, the plasmid was then cured by one round of patch plating (2.4.13), and the KO genotype reconfirmed by colony PCR. This resulted in three strains, *C. butyricum ΔpyrE*, *C. butyricum ΔpurD* and *C. butyricum ΔhisI*, and glycerol stocks were made for three biological replicates of each strain.

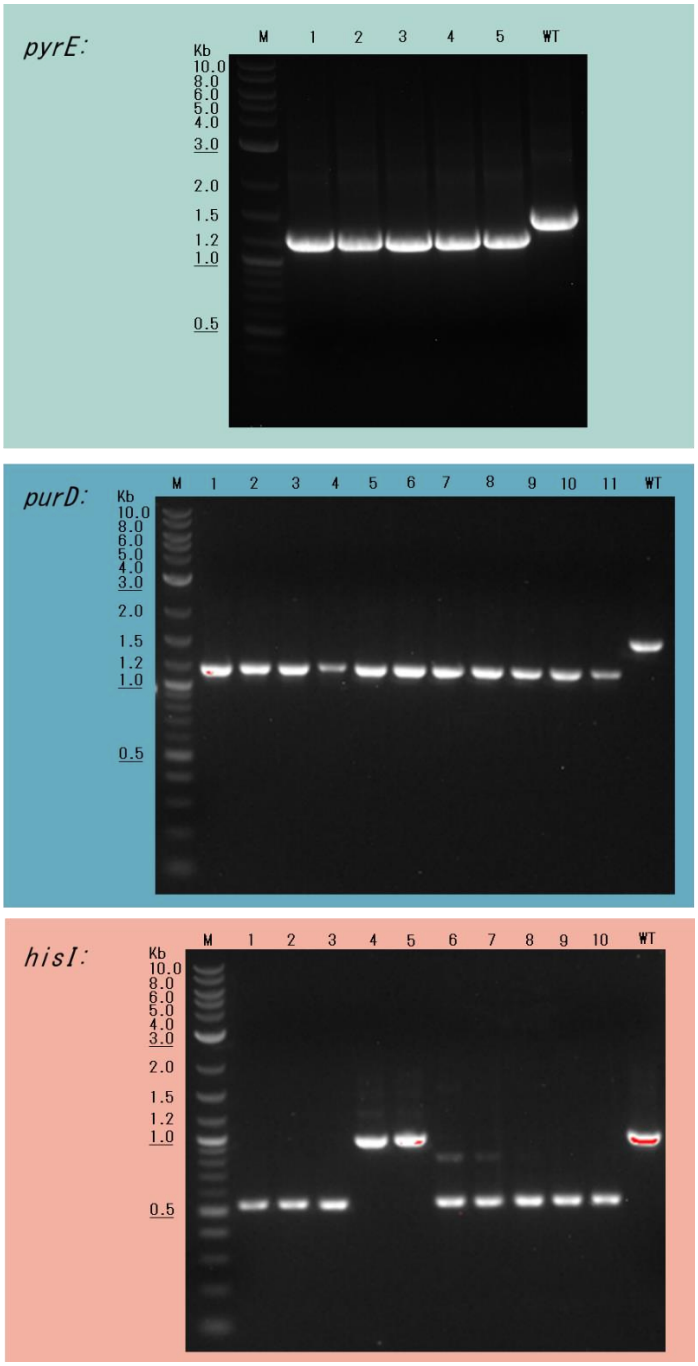


Figure 4.16 Colony PCR screening for *pyrE*, *purD* and *hisI* truncations. Agarose gel electrophoresis visualisation of colony PCRs with primers to assess for deletions by RiboCas in chosen *C. butyricum* genes. WT = wildtype *C. butyricum*, M = DNA Marker (2-log ladder; NEB). Numbered lanes = transconjugant colonies plated on thiamphenicol selection with theophylline inducer. Primers and expected fragment sizes (bp) *pyrE*: BUT_pyrE_F1 and BUT_pyrE_R2. WT = 1513, KO = 1279. *purD*: BUT_PurD_F1 and BUT_PurD_R1. WT = 1600, KO = 1195. *hisI*: BUT_HisI_F1 and BUT_HisI_R1. WT = 994, KO = 547.

As demonstrated in the gel images, RiboCas efficiently generated gene knockouts, with pure knockouts generated and no mixed colonies of wildtype and knockout genotype observed. Moreover, the editing efficiency was high across the colonies screened for each gene: *pyrE* (100%, 5/5), *purD* (80%, 8/10) and *hisI* (100%, 11/11). Only one culture in media without antibiotic supplementation was required to cure the plasmid. Taken together these factors allow efficient use of RiboCas technology in *C. butyricum* and suggest real promise in its successful further use for genetic manipulation.

4.3.2 Characterisation of auxotrophic mutants

Following gene truncation, the three mutant strains *Δpyre*, *ΔpurD* and *ΔhisI* were assessed for their hypothesised auxotrophy. This was achieved by confirming absence of growth on minimal media and attempting to rescue growth with addition of appropriate metabolite.

ΔpyrE

To allow its use in selection for *pyrE* genotypes, the MIC of FOA was first ascertained for wildtype *C. butyricum*; FOA is toxic to wildtype *pyrE*. Single colonies of *C. butyricum* WT were streaked onto increasing concentrations of FOA in minimal media, and grown for 36 h (Figure 4.17).

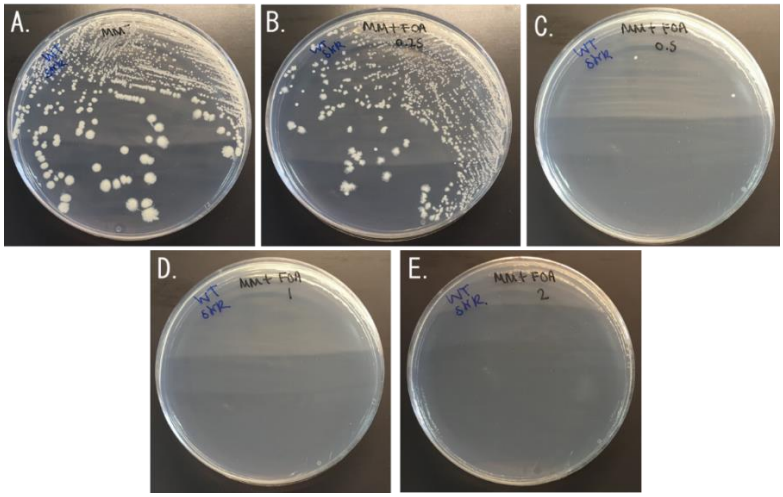


Figure 4.17 Minimum inhibitory concentration assay of fluoroorotic acid (FOA) in *C. butyricum*. *C. butyricum* WT was plated onto increasing concentrations of FOA in minimal media: A) 0 mg/mL B) 0.25 mg/mL C) 0.5 mg/mL D) 1 mg/mL E) 2 mg/mL, and grown for 36 h.

C. butyricum grew well until 0.5 mg/mL of FOA, where there was very limited growth. There was no growth at 1 or 2 mg/mL of FOA, placing the MIC of FOA between 0.5 and 1 mg/mL. This is as expected for clostridia, for example the MIC for *C. difficile* is 1 mg/mL (data not shown). The Δ*pyrE* strain was then plated on minimal media plates with and without the addition of the metabolite uracil (5 µg/mL), and with and without FOA (3 mg/mL) (Figure 4.18).

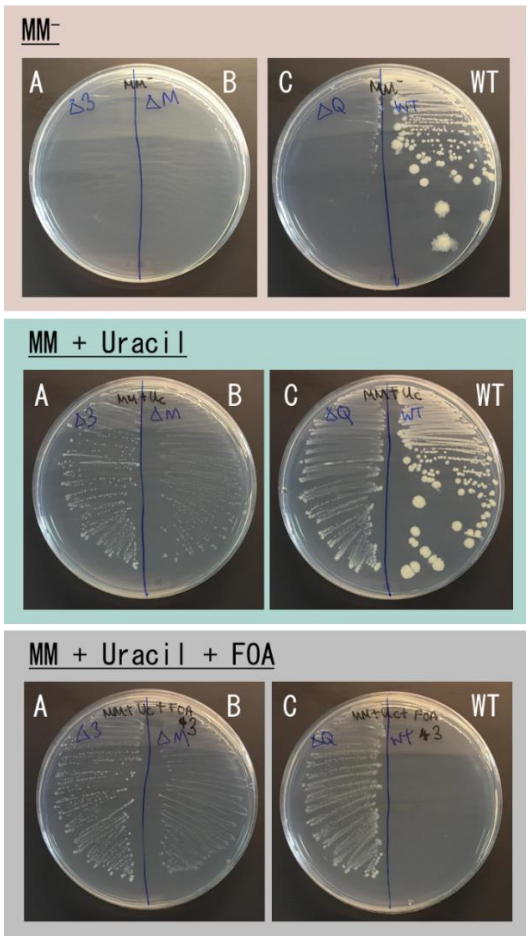


Figure 4.18 Growth of the *C. butyricum ΔpyrE* strain on minimal media (MM). Biological triplicates A-C of the knockout and wildtype (WT) strains. On minimal media with no supplementation (MM-) *ΔpyrE* was unable to grow as it is unable to produce uracil endogenously. On MM with uracil supplementation (5 µg/mL), *ΔpyrE* and WT grew. On MM with uracil and FOA supplementation (3 mg/mL), *ΔpyrE* grew but WT did not, as *ΔpyrE* is unable to convert FOA to the toxic 5-FUMP.

All three biological replicates of *ΔpyrE* displayed the phenotypes expected of a non-functioning *pyrE* gene, under each growth condition. On minimal media (lacking uracil), WT grew but *ΔpyrE* was unable to grow as it is unable to produce uracil endogenously. On minimal media supplemented with uracil, *ΔpyrE* and WT grew. On minimal media supplemented with uracil and FOA, *ΔpyrE* grew but WT did not, as *ΔpyrE* are unable to convert FOA to the toxic 5-FUMP. This confirms that the *pyrE* truncation successfully inactivates gene function, and is therefore suitable for use and selection with repair and complementation vectors.

ΔpurD

The *ΔpurD* strain was plated onto minimal media and was unable to grow, shown in Figure 4.19, demonstrating inactivation of gene function. To attempt to restore growth, it was then plated onto minimal media containing the purine cocktail used by Truong et al. (2015) for their *ΔpurD* strain in *Brucella abortus*, containing adenine, guanine and hypoxanthine all at a final concentration of 1 mM. As shown in Figure 4.19, this cocktail was unable to restore growth of the mutant.

Due to the inability of the purine cocktail to restore growth, the mutant was subsequently plated onto minimal media with the purine cocktail also containing thiamine; addition of thiamine by Truong et al. (2015) increased the growth rate of their *ΔpurD* strain and suggested thiamine dependent growth. In *C. butyricum*, the pathway for purine metabolism that involves *purD* is also shared with the pathway to generate the thiamine precursor 5-aminoimidazole ribotide. If the *purD* knockout is also inhibiting thiamine production, thiamine supplementation is likely due to its requirement in many metabolic processes in its coenzyme form. As shown in Figure 4.19, addition of thiamine (0.05 mM) with the purine cocktail did successfully restore growth of the mutant, and therefore confirmed suitability of the *purD* gene truncation for repair and complementation.

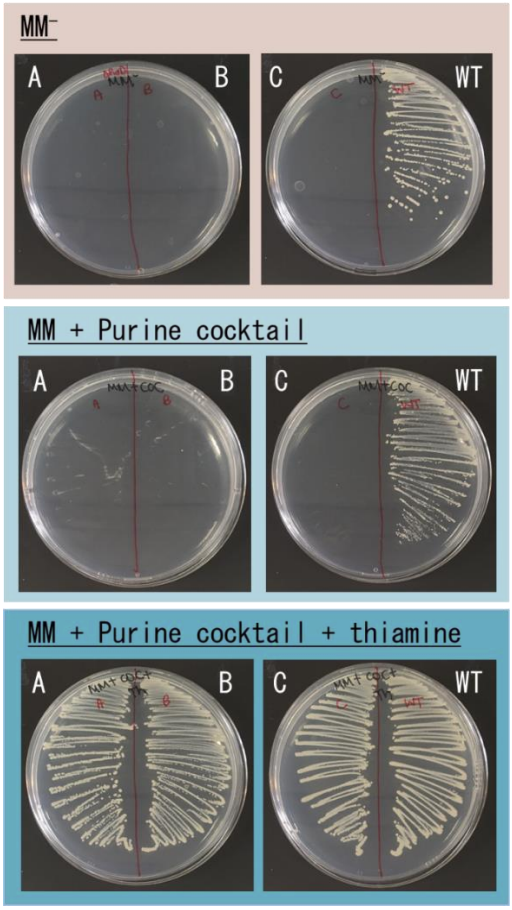


Figure 4.19 Growth of the *C. butyricum ΔpurD* strain on minimal media (MM). Biological triplicates A-C of the knockout and wildtype (WT) strains. On minimal media with no supplementation (MM-) *ΔpurD* was unable to grow as it is unable to produce purines endogenously. On MM with purine cocktail (adenine, guanine, hypoxanthine at a final concentration of 1 mM) supplementation, *ΔpurD* did not grow. Addition of thiamine (0.05 mM) restored growth of *ΔpurD*.

ΔhisI

Prior to assessing growth of the *ΔhisI* strain, development of the minimal media was required due to the presence of histidine in the cas amino acids (BactoTM Casamino Acids, Thermo Fisher) component. Alternative formulations, shown in Table 4.5, based on the essential amino acid requirements of other clostridia were tested, in addition to additional nitrogen, to identify variations that did not contain histidine but allowed growth. Wildtype *C. butyricum* was plated onto the five formulations and growth measured in the form of CFU/mL (Figure 4.20).

Table 4.5 Composition of alternative minimal media requirements tested for *C. butyricum.*

Media	Description	Amino acid content
AA-	Minimal media without Cas amino acids	None
6 AA	Minimal media with 6 amino acids (Karasawa, Ikoma, Yamakawa, & Nakamura, 1995)	Cysteine, isoleucine, leucine, proline, tryptophan, valine
9 AA	Minimal media with 9 amino acids (Seddon & Borriello, 1989)	Aspartate, glutamate, glycine, valine, proline, cysteine, phenylalanine, leucine, lysine
AA- + N	Minimal media without Cas amino acids, with additional $(NH_4)_2SO_4$	None. Double concentration of $(NH_4)_2SO_4$
MM	Minimal media with Cas amino acids	All 20 amino acids

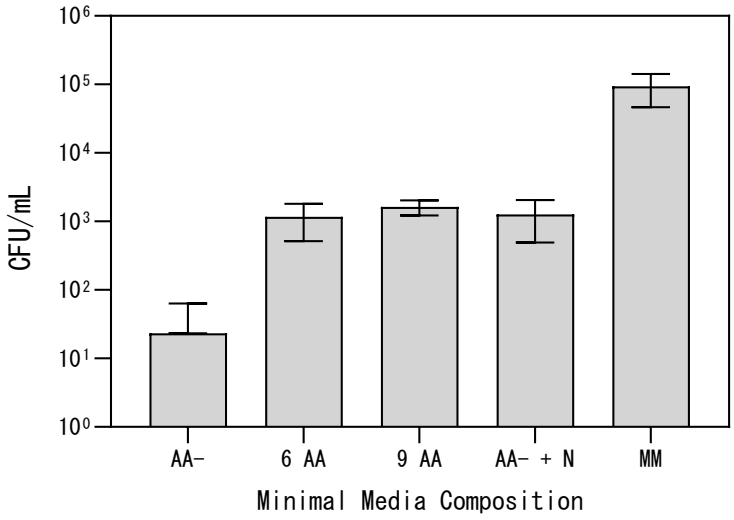


Figure 4.20 Alternative minimal media formulations for growth of *C. butyricum*. Growth of *C. butyricum* WT was assessed with different media formulations as follows: **AA-** no cas amino acids. **6 AA** Cysteine, isoleucine, leucine, proline, tryptophan, valine. **9 AA** Aspartate, glutamate, glycine, valine, proline, cysteine, phenylalanine, leucine, lysine. **AA- + N** no cas amino acids with additional nitrogen, double concentration of $(NH_4)_2SO_4$. **MM** original minimal media formula.

As shown in Figure 4.20, minimal media without amino acids did not allow growth of wildtype *C. butyricum*, as expected. The addition of 6 amino acids, 9 amino acids or additional nitrogen in the form of $(NH_4)_2SO_4$ all increased growth to a similar degree. Although the CFU was lower than complete minimal media, growth was

sufficient to allow screening for auxotrophy. The formulation with additional nitrogen only was chosen due to the ease of use compared to adding individual amino acids.

Following minimal media identification, the *ΔhisI* strain was plated onto the chosen modified minimal media and was unable to grow, shown in Figure 4.21. Growth was then restored with addition of histidine (20 µg/mL). This demonstrates successful gene inactivation and suitability for further use.

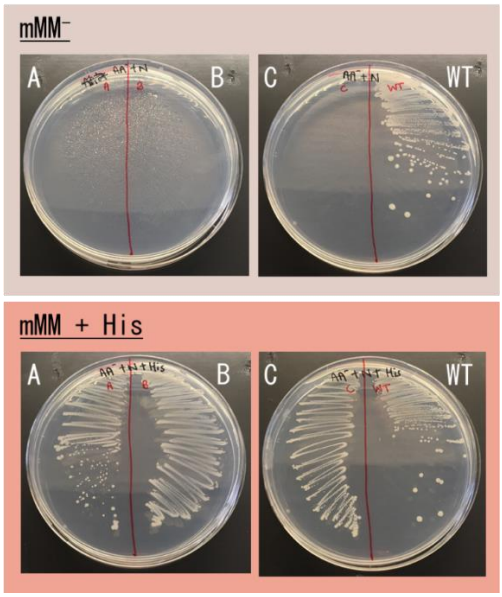


Figure 4.21 Growth of the *C. butyricum ΔhisI* strain on modified minimal media (mMM) Biological triplicates A-C of the knockout and wildtype (WT) strain. On minimal media with no supplementation (mMM-) *ΔhisI* was unable to grow as it is unable to produce histidine endogenously. Growth was achieved on mMM with histidine supplementation (20 µg/mL).

4.3.3 Repair of auxotrophic mutants

Repair vector construction

Once successful truncation and auxotrophy had been demonstrated for each mutant strain, suitability for gene repair and subsequent cargo insertion was assessed. Vectors were designed to repair the truncations by ACE, with the truncated gene fragment positioned between appropriate homology arms for integration back into the genome. The repair fragments were generated using primers in Table 4.6, digested with SbfI

and AscI, and ligated into pMTL82151 digested with the same enzymes. Insertion was confirmed by colony PCR (ColE1_tra_F2 and pBP1_R1) and Sanger sequencing, to generate the final vectors: pMTL82151_PErepair, pMTL82151_PDrepair and pMTL82151_HIrepair. Final vectors were then transformed into *E. coli* CA434.

Table 4.6 Fragments for ACE repair vectors in *C. butyricum* mutant strains. Primers used to generate fragments by PCR to insert into pMTL82151. Forward primers contain a SbfI restriction site, reverse primers contain an AscI site. To generate the final vectors pMTL82151_PErepair, pMTL82151_PDrepair and pMTL82151_HIrepair by digestion and ligation.

Locus	Primers	Template
pyrE	BUT_PyrErepair_LF1	*C. butyricum* DNA
	BUT_PyrErepair_RR1	
purD	BUT_PurDrepair_LF1	*C. butyricum* DNA
	BUT_PurDrepair_RR1	
hisI	BUT_HisIrepair_LF1	*C. butyricum* DNA
	BUT_HisIrepair_RR1	

Repair and return to prototrophy

Repair vectors were conjugated into each mutant strain, and transconjugants selected for with thiamphenicol. The double crossover and subsequent gene repair was selected for by plating of transconjugants onto minimal media without appropriate supplementation; only colonies that had repaired back to wildtype would be able to synthesise the missing amino acid or vitamin and grow. Gene repair was confirmed by colony PCR screening and Sanger sequencing, the plasmid was cured, and the restoration to wildtype was confirmed for all three loci by return to prototrophy on minimal media (Figure 4.22).

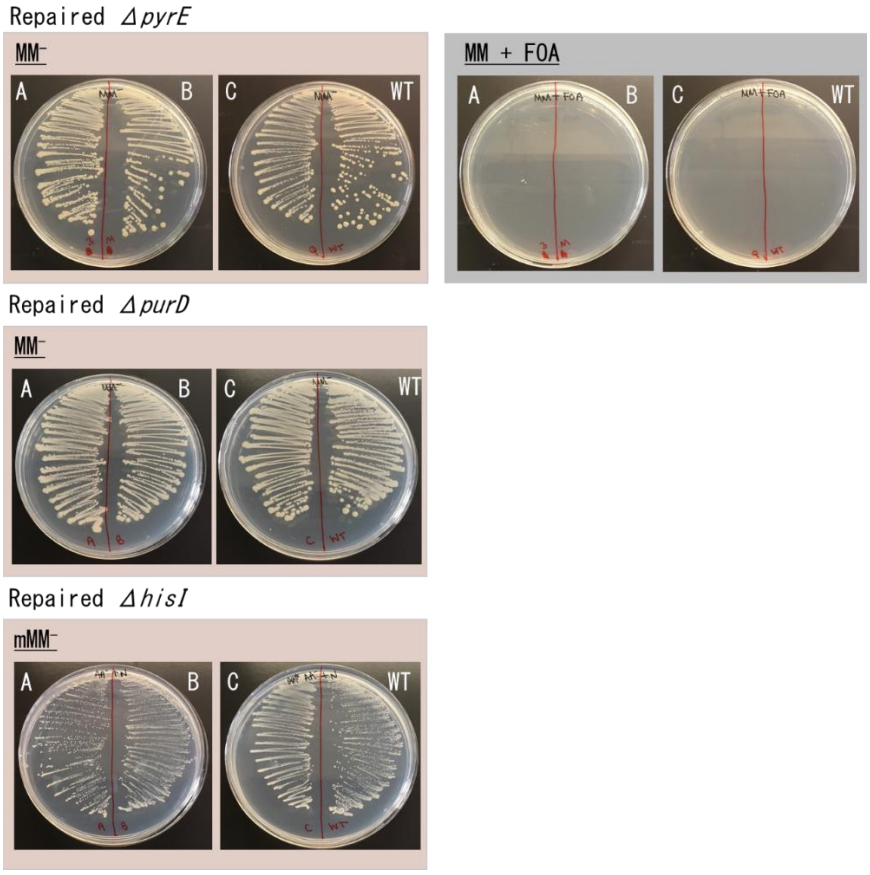


Figure 4.22 Repaired truncations of *pyrE*, *purD* and *hisI* in *C. butyricum*. Sequence confirmed truncation repairs plated onto minimal media (MM) without appropriate supplementation. Growth is restored at all three loci, as expected from a repaired gene. Additional characterisation with FOA for the *pyrE* knockout demonstrates a lack of growth for the repaired strain, also as expected. Biological triplicates A-C shown, and wildtype (WT).

Complementation vector construction

Following confirmation of the suitability of each locus for use in ACE applications, ACE complementation vectors were constructed. Homology arms used in the repair vectors were generated by PCR (Table 4.7) and inserted into the ACE backbone pMTL-YN2C, either side of the MCS. The ~1200 bp LHA was digested with SbfI and NotI and inserted first, with confirmation by colony PCR (ColE1_tra_F2 and M13_F). The ~300 bp RHA + truncated fragment was then digested with BmtI and AscI, inserted into the digested backbone, and insertion confirmed by colony PCR (MCS_F1 and pBP1_R1) and Sanger sequencing. This yielded the final ACE complementation

vector backbones pMTL-YN2C_PE, pMTL-YN2C_PD and pMTL-YN2C_HI (for *pyrE*, *purD* and *hisI* respectively), ready for insertion of cargo into the MCS. The truncated fragment from each gene was included in the RHA to allow repair to prototrophy with the second crossover, of the shorter homology arm.

Table 4.7 Fragments for ACE complementation vectors in *C. butyricum* mutant strains. Primers used to generate homology arm fragments by PCR to insert into pMTL-YN2C. Primers encode the following restriction sites: LF1= SbfI, LR1= NotI, RF1 = BmtI, RR1= AscI.

Target	Fragment	Primers	Template	Final vector
pyrE	LHA (1191 bp)	BUT_PyrErepair_LF1 BUT_PyrEcomp_LR1	*C. butyricum* DNA	pMTL-YN2C_PE
	RHA (298 bp)	BUT_PyrEcomp_RF1 BUT_PyrErepair_RR1	*C. butyricum* DNA	
purD	LHA (1199 bp)	BUT_PurDrepair_LF1 BUT_PurDcomp_LR1	*C. butyricum* DNA	pMTL-YN2C_PD
	RHA (299 bp)	BUT_PurDcomp_RF1 BUT_PurDrepair_RR1	*C. butyricum* DNA	
hisI	LHA (326 bp)	BUT_HisIrepair_LF1 BUT_HisIcomp_LR1	*C. butyricum* DNA	pMTL-YN2C_HI
	RHA (1200 bp)	BUT_HisIcomp_RF1 BUT_HisIrepair_RR1	*C. butyricum* DNA	

4.3.4 A Triple Auxotroph strain

Phenotypic characterisation

As truncation and repair were successfully demonstrated at all three loci, as well as the expected auxotrophy and prototrophy, all were deemed suitable to use in the final strain. The RiboCas vectors were conjugated, induced and lost to sequentially truncate each gene in the order of *pyrE*, *purD*, *hisI*, generating a double then a triple knockout strain. This was carried out to produce biological triplicates, producing a single, double and triple knockout strain originating from the biological triplicates of the *pyrE*

knockout, A-C. Taken together these strains provide opportunities for gene insertion and expression at one or more loci in *C. butyricum*, including orthogonal gene expression.

The final *ΔpyrEΔpurDΔhisI C. butyricum* strain was also assessed for auxotrophy (Figure 4.23). Biological triplicates A-C for the triple knockout strain were plated onto minimal media without supplements, with two of the three restorative supplements, or all three of the supplements (uracil, histidine and purines). As expected, without supplements or with only two supplements the triple knockouts were unable to grow, compared to the WT which could. There was some streaking on the media when triple knockouts were grown with histidine and purines, but this was marks within the agar possibly from the purine cocktail, and did not produce real colonies. This therefore renders the triple knockout strain suitable for use when concomitantly restoring one locus and inserting cargo; addition of the other two supplements will not interfere with growth and therefore allows selection with restoration to prototrophy for the targeted locus. When all three relevant supplements were added growth of the triple knockouts was restored.

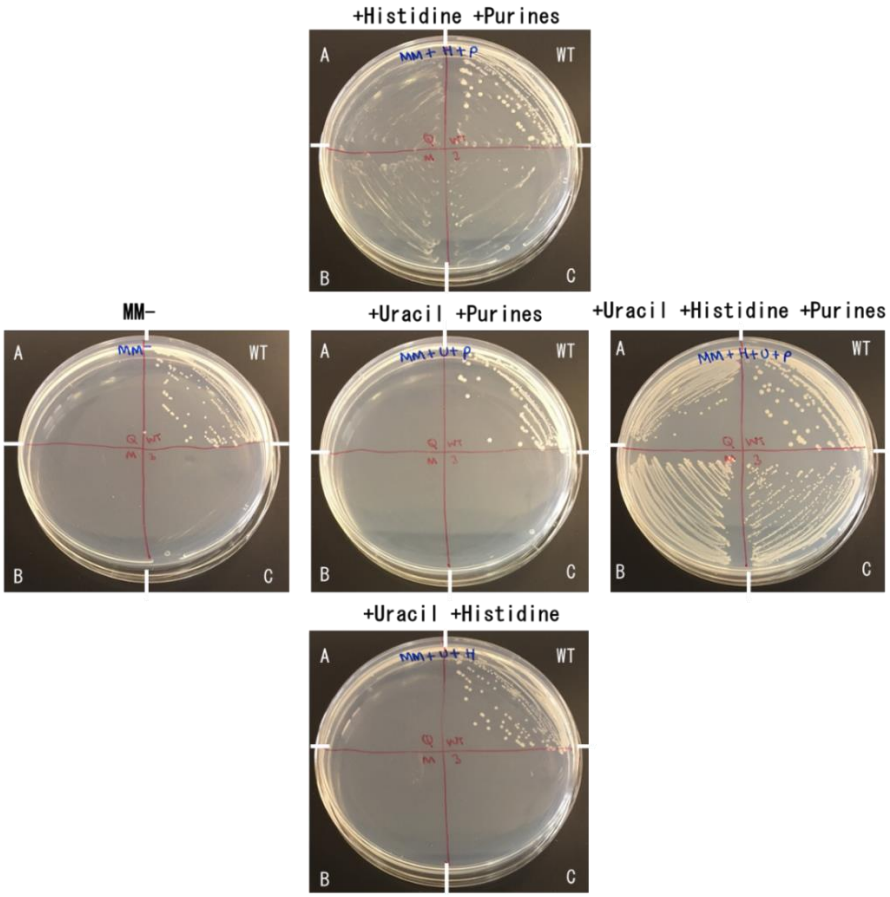


Figure 4.23 Growth of *C. butyricum* triple knockouts on minimal media with and without supplementation. Three biological triplicates (A-C) of *ΔpyrEΔpurDΔhisI C. butyricum* strains vs wildtype (WT) *C. butyricum*. Growth on minimal media without any supplements (MM-), with two of the three auxotrophic supplements, and with all three supplements.

A growth curve was also carried out for the auxotrophic knockouts in rich RCM liquid media, to ensure that growth was not altered compared to WT and that it would be sufficient to culture at a similar rate in RCM. All single knockout strains and the triple knockout reached the same OD_{600} as the WT, and the growth profiles were very similar over 24 h apart from *ΔpyrE* which lagged behind slightly until 6 h (Figure 4.24).

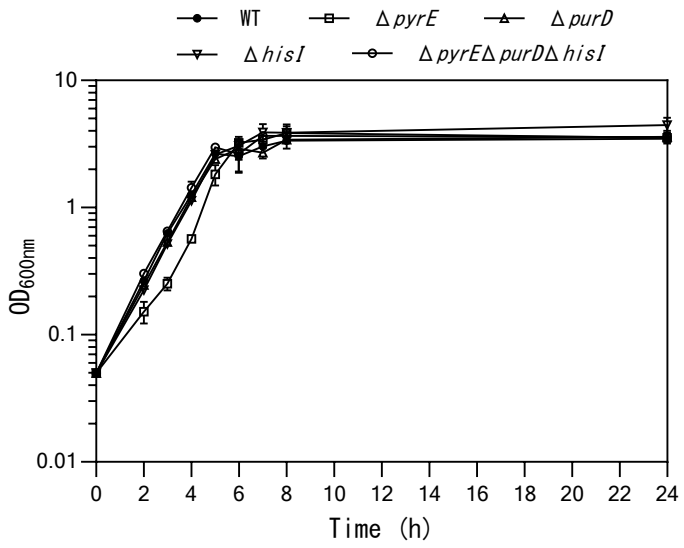


Figure 4.24 Growth of *C. butyricum* auxotrophic strains. Growth curve of *C. butyricum* strains in liquid RCM medium measured by OD600. Strains studied as follows: wildtype (WT), single auxotrophs *ΔpyrE*, *ΔpurD* and *ΔhisI*, triple auxotroph *ΔpyrEΔpurDΔhisI*. Data represent mean values of independent cultures of three biological replicates ± SD.

Whole genome sequencing

To determine whether any genomic variants had occurred during the process of generating truncations, genomic DNA samples were prepared by phenol chloroform extraction (2.4.15) for the three biological triplicates of the *ΔpyrE* and *ΔpyrEΔpurDΔhisI* strains, and sent for whole genome resequencing using the Illumina MiSeq platform (DeepSeq, Nottingham, UK)(2.5.2). Using CLC Genomics Workbench, Illumina reads were aligned to the *C. butyricum* NCTC 7423 reference genome (GenBank assembly GCA_014131795.1), and genomic variants were identified according to the parameters outlined in 2.6.5. The WT strain was also sent to allow identification of variants originating in the laboratory collection stock, shown in Table 4.8. Additional variants identified in the newly generated truncated strains are shown in Table 4.9.

Table 4.8 Resequencing analysis of *C. butyricum* auxotrophic strains: Variants identified in the WT stock. SNPs and other genomic variants identified in the WT stock of *C. butyricum* NCTC 7423 and their identification in *ΔpyrE* replicates A-C and *ΔpyrEΔpurDΔhisI* replicates A-C. Gene and function identified according to annotations of the reference (Ref.) genome (GCA_014131795.1) (IG=intergenic). For variants, A=sequence at locus, C=coverage of reads, F=frequency in reads. Amino acid changes arising from variants are listed (AA). Blank cells denote no change from reference sequence.

				WT			***ΔpyrE* strain A**			***ΔpyrE* strain B**			***ΔpyrE* strain C**			
Gene	**Function**	**Position**	**Ref.**	**A**	**C**	**F**	**A**	**C**	**F**	**A**	**C**	**F**	**A**	**C**	**F**	**AA**
IG		385316^385317	-	**A**	47	95.74	**A**	45	93.33	**A**	32	100	**A**	37	97.3	IG
FF104_02960	ABC transporter	711817	**C**	**A**	45	100	**A**	37	100	**A**	44	97.73	**A**	50	100	Ser343Tyr
FF104_05000	cation symporter	1184025	**A**	-	78	98.72	-	51	100	-	40	97.5	-	66	100	Lys2FS
FF104_06405	insulinase family protein	1467699	**G**	**T**	34	76.47				**T**	19	100				Asp373Tyr
FF104_08935	Subtilisin/kexin/sedolisin	2053199	**C**	**A**	61	96.72	**A**	38	100	**A**	40	100	**A**	49	100	Thr386Lys
FF104_09950	triacylglycerol lipase	2256328	**G**	**T**	66	100	**T**	37	100	**T**	32	100	**T**	53	98.11	Ser368Stop
IG		2257309	**A**	**C**	33	100	**C**	35	100	**C**	29	100	**C**	37	100	IG
FF104_13335	GGDEF domain protein	2941636	**C**	**A**	41	100	**A**	55	100	**A**	26	100	**A**	45	100	Lys267Asn
FF104_14500	chemotaxis protein CheC	3199662	**G**	**A**	74	100	**A**	54	100	**A**	46	100	**A**	74	100	Ala66Val
FF104_15580	cytidine deaminase	3439952	**C**	**A**	79	100	**A**	60	100	**A**	35	97.14	**A**	56	100	Arg4Ile
FF104_17685	HD-GYP domain protein	3865029	**A**	-	46	97.83	-	46	100	-	34	100	-	47	100	Leu58FS
FF104_19060	PAS domain S-box protein	268402	**C**	**A**	41	100	**A**	43	100	**A**	21	100	**A**	27	100	Ser285Tyr

				***ΔpyrEΔpurDΔhisI* strain A**			***ΔpyrEΔpurDΔhisI* strain B**			***ΔpyrEΔpurDΔhisI* strain C**			
Gene	**Function**	**Position**	**Ref.**	**A**	**C**	**F**	**A**	**C**	**F**	**A**	**C**	**F**	**AA**
IG		385316^385317	-	**A**	35	91.43	**A**	51	94.12	**A**	46	95.65	IG
FF104_02960	ABC transporter	711817	**C**	**A**	41	100	**A**	45	100	**A**	44	100	Ser343Tyr
FF104_05000	cation symporter	1184025	**A**	-	65	98.46	-	52	100	-	63	100	Lys2FS
FF104_06405	insulinase family protein	1467699	**G**				**T**	43	100				Asp373Tyr
FF104_08935	Subtilisin/kexin/sedolisin	2053199	**C**	**A**	43	100	**A**	38	100	**A**	54	100	Thr386Lys
FF104_09950	triacylglycerol lipase	2256328	**G**	**T**	33	100	**T**	38	100	**T**	54	100	Ser368Stop
IG		2257309	**A**	**C**	30	100	**C**	34	100	**C**	24	100	IG
FF104_13335	GGDEF domain protein	2941636	**C**	**A**	49	97.96	**A**	41	100	**A**	40	100	Lys267Asn
FF104_14500	chemotaxis protein CheC	3199662	**G**	**A**	41	100	**A**	59	100	**A**	60	100	Ala66Val
FF104_15580	cytidine deaminase	3439952	**C**	**A**	55	100	**A**	54	100	**A**	64	100	Arg4Ile
FF104_17685	HD-GYP domain protein	3865029	**A**	-	45	100	-	61	100	-	55	100	Leu58FS
FF104_19060	PAS domain S-box protein	268402	**C**	**A**	30	100	**A**	36	100	**A**	41	100	Ser285Tyr

Table 4.9 Resequencing analysis of *C. butyricum* auxotrophic strains: Variants unique from the WT stock. SNPs and other genomic variants, unique to those identified in the WT strain, identified in the *C. butyricum* NCTC 7423 *ΔpyrE* replicates A-C and *ΔpyrEΔpurDΔhisI* replicates A-C. Gene and function identified according to annotations of the reference (Ref.) genome (GCA_014131795.1) (IG=intergenic). For variants, A=sequence at locus, C=coverage of reads, F=frequency in reads. Amino acid changes arising from variants are listed (AA). Blank cells denote no change from reference sequence.

				***ΔpyrE* strain A**			***ΔpyrE* strain B**			***ΔpyrE* strain C**			
Gene	**Function**	**Position**	**Ref.**	**A**	**C**	**F**	**A**	**C**	**F**	**A**	**C**	**F**	**AA**
IG		235358	**G**										IG
FF104_02935	alpha-galactosidase	701791	**T**	**G**	40	100							Asp84Glu
FF104_03780	copper homeostasis	918693	**G**	**T**	53	100				**T**	64	100	Met1Ile
IG		1573194	**G**				**T**	21	100				IG
FF104_07435	DUF3427 domain protein	1716501	**G**										Glu369Stop
FF104_07890	LysM	1826176	**C**										Asp52Tyr
FF104_11735	methyltransferase PglX	2633866	**C**										Glu906Asp
IG		2674616	**G**	**T**	53	100							IG
FF104_12495	HD-GYP domain protein	2792655	**A**							**T**	53	100	Ile176Lys
FF104_14125	P-type ATPase	3122115	**C**				**A**	31	100				Arg58Met
FF104_16255	PRD domain protein	3566261	**C**				**A**	40	100				Glu865Stop
FF104_20205	CPBP metalloprotease	516784	**C**				**A**	32	100				Val243Phe
FF104_20360	glycerol dehydrogenase	555607	**C**				**A**	26	100				Gly271Val

				***ΔpyrEΔpurDΔhisI* strain A**			***ΔpyrEΔpurDΔhisI* strain B**			***ΔpyrEΔpurDΔhisI* strain C**			
Gene	**Function**	**Position**	**Ref.**	**A**	**C**	**F**	**A**	**C**	**F**	**A**	**C**	**F**	**AA**
IG		235358	**G**							**T**	43	97.67	IG
FF104_02935	alpha-galactosidase	701791	**T**	**G**	27	100							Asp84Glu
FF104_03780	copper homeostasis	918693	**G**	**T**	41	100				**T**	42	100	Met1Ile
IG		1573194	**G**				**T**	38	97.37				IG
FF104_07435	DUF3427 domain protein	1716501	**G**							**T**	52	96.15	Glu369Stop
FF104_07890	LysM	1826176	**C**							**A**	27	100	Asp52Tyr
FF104_11735	methyltransferase PglX	2633866	**C**				**A**	46	100				Glu906Asp
IG		2674616	**G**	**T**	53	98.11							IG
FF104_12495	HD-GYP domain protein	2792655	**A**							**T**	50	100	Ile176Lys
FF104_14125	P-type ATPase	3122115	**C**				**A**	41	100				Arg58Met
FF104_16255	PRD domain protein	3566261	**C**				**A**	47	100				Glu865Stop
FF104_20205	CPBP metalloprotease	516784	**C**				**A**	46	100				Val243Phe
FF104_20360	glycerol dehydrogenase	555607	**C**				**A**	51	98.04				Gly271Val

A total of 12 variants were identified in the WT stock (Table 4.8) with SNPs, insertions and deletions, and variants were not clustered in one region but throughout the whole genome and one large plasmid. Of the 12 variants, 10 were within coding regions and consist of 7 missense, 2 frameshift and 1 nonsense mutations. All but one of the variants were consistent across the 6 knockout strains analysed, with the SNP in FF104_06405 having a 76% frequency in WT reads, but 100% frequency in *ΔpyrE* strain B and the daughter *ΔpyrEΔpurDΔhisI* strain. This could suggest that there was a heterogenous genotype in the original culture, resulting in either the presence or absence of the variant in the three independently derived *ΔpyrE* strains. The variants consistently identified affect a range of proteins, including transporter, proteases and deaminases.

There were a further 13 variants identified across the 6 KO strains in addition to those identified in the WT stock (Table 4.9). These were as follows in the single *ΔpyrE* strains: A= 3, B= 5 and C= 2. In the subsequent rounds of mutagenesis with RiboCas to generate the triple knockouts, strain A did not accumulate any additional variants, but strain B gained 1 and strain C gained 3. Of the total variants 10 were in coding regions and these consisted of 8 missense, 1 frameshift and 1 nonsense mutation. Again these were located throughout the chromosome and large plasmid. A variant of particular interest is one in the *ΔpyrEΔpurDΔhisI* strain B; a Glu906Asp in *pglx*. The missense mutation in *pglx* previously may have been predicted to improve conjugation efficiency of this strain due to its predicted involvement in RM systems, however the *Δpglx* strain generated in 4.2.4 suggested it would not. Moreover, experimentally no difference in efficiency was observed compared to strains B and C.

Overall a number of variants were identified in the final auxotrophic strains, with a total number of 14, 19 and 16 variants for the *ΔpyrEΔpurDΔhisI* A, B and C strains respectively compared to the reference sequence. Whilst the function of important proteins may be affected due to missense, frameshift and nonsense mutations, evidently none of them are essential. Moreover, as these strains are being developed to eventually use as chassis strains for gene expression in the gut, these variants are not as impactful as if loss or gain of function mutations were being studied. However, the variants may require characterisation before their eventual use as a delivery tool to assess possible impacts on safety or efficacy of the WT as a probiotic.

4.4 Characterisation of promoters for heterologous gene expression

4.4.1 Lactose inducible tcdR/P_{tcdB} system

Following generation of the *ΔpyrEΔpurDΔhisI* strain in preparation for cargo insertion by ACE, the promoter for expression of exogenous genes in the *C. butyricum* chromosome required characterisation. The triple knockout strains were developed with the lactose inducible *tcdR*/P_{tcdB} system in mind, allowing insertion of the lactose inducible *bgaR* at one locus, upstream of *tcdR* under control of P_{bgaL} which is induced by BgaR. Exogenous genes under the control of P_{tcdB} can be inserted at the other two loci, and expression is activated by TcdR.

It was chosen to assess the expression system using a FAST reporter assay, and the cloning strategy to achieve this is outlined in Table 4.10. Issues with cloning at each stage necessitated various assembly techniques and plasmid designs to be attempted, and the strategy to evolve alongside this. Firstly, plasmids were designed to allow subsequent assessment of the two promoter components of the system; the inducible P_{bgaL} along with its transcriptional activator *bgaR*, and P_{tcdB} which would require characterisation alongside TcdR production. To allow this a two plasmid system was then designed, to provide the lactose inducible *tcdR* component to be expressed on a second plasmid. As the cloning for this was not successful it was then attempted to express the whole system on a single plasmid with a reporter, but this was not successful either. As the final aim is to insert lactose inducible *tcdR* into the genome, this was attempted for the last cloning strategy, which could then be assessed alongside the P_{tcdB} reporter plasmid.

Table 4.10 Cloning strategies to allow assessment of the lactose inducible *tcdR*/P$_{tcdB}$ expression system using FAST. Strategies taken for cloning of components of the lactose inducible *tcdR*/P$_{tcdB}$ for insertion with FAST into appropriate plasmids to allow a reporter assay in *C. butyricum*. Strategies include assessment on plasmids and insertion into the genome.

A) INDIVIDUAL PROMOTERS: pMTL87151 + *bgaR*/P$_{bgaL}$ + FAST and pMTL87151 + P$_{tcdB}$ + FAST

Method	Description	Fragment	Production	Template	Primers	Fragment made? (Y/N)	Assembly method	Overall outcome?
1.	PCR *bgaR*/P$_{bgaL}$ from HZ6 and ligate with FAST reporter.	*bgaR*/P$_{bgaL}$	PCR then dig with NotI and NdeI	pMTL-HZ6	lacFAST_lac_Not_F1 lacFAST_lac_Nde_R1	N	Ligation into pMTL87151 digested with X and X	Could not carry out ligation without fragment.
		FAST	PCR then dig with NdeI and XhoI	pMTL8315_ptcdB_FAST	lacFAST_FAST_Nde_F1 lacFAST_FAST_XhoI_F1	Y		
2.	Swap the backbone of pMTL8315_ptcdB_ FAST to pMTL87151.	P$_{tcdB}$/FAST	Dig with SbfI and AscI	pMTL8315_ptcdB_FAST	N/A	Y	Ligation into pMTL87151 digested with SbfI and AscI	**Final plasmid: pMTL87151_ ptcdB_FAST**

B) TWO PLASMID SYSTEM: pMTL82151 + *bgaR*/P$_{bgaL}$.*tcdR* and pMTL87151_ ptcdB_FAST

Method	Description	Fragment	Production	Template	Primers	Fragment made? (Y/N)	Assembly method	Overall outcome?

1.	Digest out *bgaR*/P$_{bgaL}$.*tcdR* and ligate into pMTL82151	*bgaR*/P$_{bgaL}$.*tcdR*	Dig with NotI and XhoI	pMTL-HZ13	N/A	Y	Ligation into pMTL82151 digested with NotI and XhoI	No correct insertions from *E. coli* colony PCR screening.
2.	PCR *bgaR*/P$_{bgaL}$.*tcdR* with new restriction sites, digest and ligate into pMTL82151	*bgaR*/P$_{bgaL}$.*tcdR*	PCR then dig with SalI and AatII	pMTL-HZ13	82_lactcdR_Sal_F1 82_lactcdR_AatII_R1	N	Ligation into pMTL82151 digested with SalI and AatII	Could not carry out ligation without fragment.
3.	PCR *bgaR*/P$_{bgaL}$.*tcdR* and insert into pMTL82151 with hifi assembly	*bgaR*/P$_{bgaL}$.*tcdR*	PCR	pMTL-HZ13	82_lactcdrHF_F1 82_lactcdrHF_R1	N	Hifi assembly into pMTL82151 digested with X and X	Could not carry out hifi PCR without fragment.

C) PLASMID-BASED FULL SYSTEM: pMTL87151 + lactose inducible tcdR/PtcdB/FAST

Method	Description	Fragment	Production	Template	Primers	Fragment made? (Y/N)	Assembly method	Overall outcome?
1.	Replace existing *catP* reporter with FAST. PCR *bgaR*/P$_{bgaL}$.*tcdR*/P*tcdB* from HZ6 and ligate with FAST.	*bgaR*/P$_{bgaL}$.*tcdR*/P*tcdB*	PCR then dig with NotI and BsaI	pMTL-HZ6	87_lactcdrFAST_lac_Not_F1 87_lactcdrFAST_lac_Bsa_R1	Y	Three-way ligation into pMTL87151 digested with NotI and XhoI	No colonies from *E. coli* trans-formation
		FAST	PCR then dig with BsaI and XhoI	pMTL8315_ptcdB_FAST	87_lactcdrFAST_FAST_Bsa_F1 87_lactcdrFAST_FAST_Xho_R1	Y		

D) PLASMID AND CHROMOSOMAL SYSTEM: chromosomal *bgaR*/P$_{bgaL}$.*tcdR* + pMTL87151_ ptcdB_FAST

Method	Description	Fragment	Production	Template	Primers	Fragment made? (Y/N)	Assembly method	Overall outcome?
1.	Digest out *bgaR*/P$_{bgaL}$.*tcdR* and ligate into pMTL-YN2C_PE	*bgaR*/P$_{bgaL}$.*tcdR*	Dig with NotI and XhoI	pMTL-HZ13	N/A	Y	Ligation into pMTL-YN2C-PE digested with NotI and XhoI	No correct insertions from *E. coli* colony PCR screening.
2.	PCR *bgaR*/P$_{bgaL}$.*tcdR* and insert into pMTL-82151 with *pyrE* HAs with hifi assembly	LHA (same as pMTL-YN2C_PE)	PCR	*C. butyricum* DNA	PE_lactcdrHF_LHA_F1 PE_lactcdrHF_LHA_F1	Y	Hifi assembly with pMTL-YN2C-PE digested with NotI and XhoI	Could not carry out hifi PCR without fragment.
		bgaR/P$_{bgaL}$.*tcdR*	PCR	pMTL-HZ13	PE_lactcdrHF_lac_F1 PE_lactcdrHF_lac_F1	N		
		RHA (same as pMTL-YN2C_PE)	PCR	*C. butyricum* DNA	PE_lactcdrHF_RHA_F1 PE_lactcdrHF_RHA_F1	Y		
3.	PCR *bgaR*/P$_{bgaL}$.*tcdR* and insert into pMTL-82151 with *purD* HAs with hifi assembly	LHA (same as pMTL-YN2C_PD)	PCR	*C. butyricum* DNA	PD_lactcdrHF_LHA_F1 PD_lactcdrHF_LHA_F1	Y	Hifi assembly with pMTL-YN2C-PD digested with NotI and XhoI	Could not carry out hifi PCR without fragment.
		bgaR/P$_{bgaL}$.*tcdR*	PCR	pMTL-HZ13	PD_lactcdrHF_lac_F1 PD_lactcdrHF_lac_F1	N		

		RHA (same as pMTL-YN2C_PD)	PCR	*C. butyricum* DNA	PD_lactcdrHF_RHA_F1 PD_lactcdrHF_RHA_F1	Y		
4.	PCR *bgaR*/P$_{bgaL}$.*tcdR* and insert into pMTL-82151 with *hisI* HAs with hifi assembly	LHA (same as pMTL-YN2C_HI)	PCR	*C. butyricum* DNA	HI_lactcdrHF_LHA_F1 HI_lactcdrHF_LHA_F1	Y	Hifi assembly with pMTL-YN2C-PD digested with NotI and XhoI	Could not carry out hifi PCR without fragment.
		bgaR/P$_{bgaL}$.*tcdR*	PCR	pMTL-HZ13	HI_lactcdrHF_lac_F1 HI_lactcdrHF_lac_F1	N		
		RHA (same as pMTL-YN2C_HI)	PCR	*C. butyricum* DNA	HI_lactcdrHF_RHA_F1 HI_lactcdrHF_RHA_F1	Y		
5.	Insert homology arms for *pyrE* insertion into pMTL-HZ13. Generate by PCR, dig and ligate each HA in turn. Swap backbone to pMTL82151 (for inclusion of *traJ*)	LHA (same as pMTL-YN2C_PE)	PCR then dig with SbfI and NotI	*C. butyricum* DNA	BUT_PyrErepair_LF1 BUT_PyrEcomp_LR1	Y	Ligation into pMTL-HZ13 digested with SbfI and AscI.	Generated **pMTL_HZ13_LHA**
		RHA (same as pMTL-YN2C_PE)	PCR then dig with XhoI and AscI	*C. butyricum* DNA	HZ13new_RHA_XhoI_F1 BUT_PyrErepair_RR1	Y	Ligation into pMTL_HZ13_ LHA digested with XhoI and AscI.	Generated **pMTL_HZ13_LHA_RHA**
		LHA-lac/tcdR-RHA	Dig with SbfI and AscI	pMTL_HZ13_ LHA_RHA	N/A	Y	Ligation into pMTL82151 digested with SbfI and AscI.	Generated **pMTL82151_lac tcdR**

The final plasmid pMTL82151_lactcdR, containing HAs for insertion of *bgaR*/P_{bgaL}*.tcdR* at *pyrE*, was confirmed with colony PCR (ColE1_tra_F2 and pBP1_R1) and Sanger sequencing. It was then transformed into *E. coli* CA434 and conjugated into *C. butyricum ΔpyrEΔpurDΔhisI*, with transconjugants selected for with thiamphenicol. Transconjugants were restreaked onto thiamphenicol then streaked twice onto minimal media supplemented with purines and histidine, to allow selection of a double crossover at *pyrE*. Colony PCR screening was carried out with flanking primers (BUT_pyrE_F1 and BUT_pyrE_R3), but the bands produced were either the size of the *ΔpyrE* truncation, or there was no band. This was unexpected as only colonies where *pyrE* has been repaired should be able to grow. A lack of band suggests that there could be a single crossover, so the PCR was repeated with one flanking and one internal primer each side (BUT_pyrE_R3 and lacFAST_lac_Nde_R1; BUT_pyrE_F1 and tcdR_scr_F1) but there was no band produced. A range of PCR conditions was attempted and yielded the same results.

Despite growth on minimal media, it was not possible to determine whether there was insertion of *bgaR*/P_{bgaL}*.tcdR* due to inconclusive PCR results. It is possible that the insertion has not occurred, but this is unlikely as ACE proved to be successful with the repair vector utilising the same HAs. It is also possible that the insertion has occurred but the colony PCRs were not successful, and given the difficulties experienced during the cloning stage this is more likely.

As it was not possible to confirm with a reporter assay that the lactose inducible/*tcdR* system, or its individual components, is functional in *C. butyricum*, it was decided to explore alternative expression systems. Moreover, the difficulties in cloning it would impede further use, and chromosomal insertion of *bgaR*/P_{bgaL}*.tcdR* was unable to be achieved.

Construction of FAST plasmids

Promoters commonly used in other clostridial tools were chosen to assess their expression in *C. butyricum* for subsequent expression of exogenous genes. Three promoters were chosen to cover a range of promoter regulation, and to allow initial assessment of FAST as a reporter gene in *C. butyricum*: the putatively constitutive $P_{araE,}$ the theophylline inducible P_{fdxE} , and the tetracycline inducible P_{tet} system. The

first two promoters are outlined in Table 3.13. The P_{tet} system consists of a pair of divergent promoters (P_{tetR} and P_{tet}), each with an overlapping *tet* operator sequence. The transcription factor TetR negatively regulates both promoters, and its expression is driven by P_{tetR}. Expression of the gene of interest is driven by P_{tet}. Repression by TetR is relieved on addition of tetracycline, allowing gene expression and also increased expression of TetR, providing a negative feedback loop for tight regulation (Fagan & Fairweather, 2011). The chosen promoters were assembled upstream of FAST into pMTL87151 (Table 4.11). Insertion was confirmed by colony PCR (ColE1_tra_F2 and p19_R1) and Sanger Sequencing. Final plasmids were then transformed into *E. coli* CA434.

Table 4.11 Promoters to assay for use in gene expression in *C. butyricum*. Promoters commonly used in clostridial tools chosen to assess in *C. butyricum* by a FAST reporter assay. Construction method used in construction of reporter plasmid for each promoter. Promoter and FAST sequence inserted into pMTL87151.

Promoter	Construction method	Primers	Final plasmid
Promoter-less	Gram positive replicon exchange by digestion of pMTL84151_FAST_neg (FscI and AscI). Insertion of p19 replicon by digestion of pMTL87151 (FscI and AscI) and ligation.	N/A	pMTL87151_FAST_neg
P_{araE}	Digestion of P_{araE}/FAST fragment from pMTL84151_FAST_araE (NotI and AscI). Ligation into pMTL87151 (NotI and AscI).	N/A	pMTL87151_FAST_araE
P_{fdxE}	Digestion of P_{fdxE}/FAST fragment from pMTL84151_FAST_PfdxE (NotI and AscI). Ligation into pMTL87151 (NotI and AscI).	N/A	pMTL87151_FAST_fdxE
P_{tet}	PCR of P_{tet} from pMTL-CW21, digestion (NotI and NdeI) and ligation into pMTL84151_FAST_araE (NotI and NdeI).	CW21_Tet_F1_Not CW21_tet_R1_Nde	pMTL87151_FAST_tet

All four final plasmids were conjugated into *C. butyricum*.

FAST assay

The three *C. butyricum* reporter strains, plus the promoter-less control, were grown in RCM with thiamphenicol selection and samples taken at 5 h for a FAST reporter assay. The assay was carried out with technical triplicates for a preliminary assessment of FAST. An example of these results is shown in Figure 4.25.

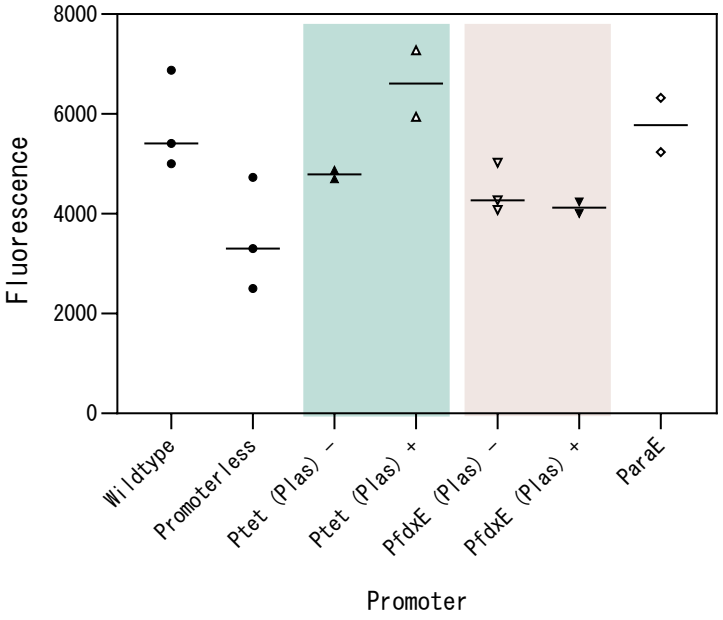


Figure 4.25 FAST reporter assay for Clostridia promoters in *C. butyricum*. *C. butyricum* harbouring the following reporter plasmids was cultured for 5 hours in RCM liquid medium + thiamphenicol: pMTL87151_FAST_neg, pMTL87151_FAST_tet, pMTL87151_FAST_fdxE and pMTL87151_FAST_araE. Anhydrous tetracycline (0.06 µg/mL) or theophylline (5 mM) inducers added at 2 hours, where indicated (- = without inducer, + = with inducer). Samples were washed and the TF-Amber ligand added. Fluorescence was measured immediately at an excitation wavelength of 499 nm and emission at 558 nm using a microplate reader. Readings were blanked to blank RCM liquid media. Data points represent technical triplicates.

The fluorescence values produced in the FAST assay for the technical triplicates was not consistent with some promoters, and even the WT, demonstrating variability. High variability was also observed in further repeats of this assay (data not shown). Moreover, the high level of fluorescence observed in the WT suggests possible background interference and that readings are not necessarily reflective of promoter activity, particularly highlighted by the reduction in fluorescence for the promoter-less control compared to the WT. Whilst there was no positive control for the assay, as there are no published active promoters in this strain, induction of P_{fdxE} was expected to be observed as this has been confirmed as active in *C. butyricum* by functioning in the RiboCas system to successfully drive *cas9* activity. However, there

was no significant difference in the fluorescence values with or without the theophylline inducer (unpaired two-tailed Student's t-test; p=0.4466). These factors together suggest that the FAST reporter assay is not suitable for use in *C. butyricum*, despite successful use in other clostridia (Poulalier-Delavelle et al., 2023; Streett et al., 2019). Because of this, the CAT activity assay was explored as an alternative.

4.4.2 catP promoter library

The CAT activity assay is used to evaluate promoter activity by coupling the activity of a downstream gene, *cat*, to the production of a substance with a strong absorbance that can be measured, 5-nitro-2-thiobenzoic acid (TNB). The bacterial *cat* gene encodes chloramphenicol acetyltransferase which catalyses the acetylation of chloramphenicol by the co-factor acetyl-CoA and renders it inactive, conferring chloramphenicol resistance. This reaction produces free CoA-SH and this can be quantified by the inclusion of 5,5'-Dithiobis (2-nitrobenzoic acid, DTNB); DTNB reacts with the thiol group and produces the yellow compound TNB which can be measured by absorbance at 412 nm (Kozakai, Shimofusa, Nomura, & Suzuki, 2021; Shaw, 1975; Silverstein, 1975).

Construction of catP plasmids

Reporter plasmids were chosen or generated for clostridial promoters (Table 4.12), consisting of the promoter upstream of the *catP* gene. The backbone pMTL8225 was chosen to allow plasmid selection by erythromycin.

Table 4.12 Promoters to assay for use in gene expression in *C. butyricum*. Promoters successfully used in clostridial tools chosen to assess in *C. butyricum* by a CAT reporter assay. pMTL8225 is the backbone for all reporter plasmids.

Promoter	Description	Plasmid	Source
Promoter-less	pMTL8225_MCS with the *catP* reporter gene inserted into the MCS.	pMTL82254	Laboratory collection
P_{araE}	Promoter of the *Clostridium acetobutylicum araE* gene. Exchanged the p19 rep from 87254_ParaE (laboratory collection) for pBP1 by digestion (FseI and AscI) and ligation.	pMTL82251_CAT_araE	This study
P_{fdx}	Promoter of the *Clostridium sporogenes* ferredoxin gene.	pMTL-IC101	Laboratory collection

P_{fdxE}	P_{fdx} under the control of the theophylline responsive riboswitch E.	pMTL-IC111-E	Laboratory collection
LAC	LAC system from *C. perfringens* strain 13. Includes the transcriptional activator, *bgaR*, and the corresponding inducible promoter, P_{bgaL}.	pMTL-HZ1	Laboratory collection
RiboLac	Riboswitch-E fused to the core of the *bgaR* promoter, P_{bgaR}, downstream of the TSS. Rest of system remains as in HZ1. Provides translational control over *bgaR* and transcriptional control over the target gene (*catP*).	pMTL-ICE1	Laboratory collection

CAT assay

All six final CAT reporter plasmids (Table 4.12) were conjugated into *C. butyricum* and transconjugants were selected for by erythromycin. Plasmid presence was confirmed by re-streaking on selection.

The CAT activity reporter assay was then carried out (Figure 4.26) according to the protocol outlined in 2.3.17. Reporter strains were cultured in RCM liquid media with erythromycin selection. The inducers theophylline (5 mM), lactose (10 mM) and theophylline with lactose were added for P_{fdxE}, LAC and RiboLac respectively at 2 h. Samples were taken at 5 h and CAT activity measured. CAT activity was normalised to protein concentration.

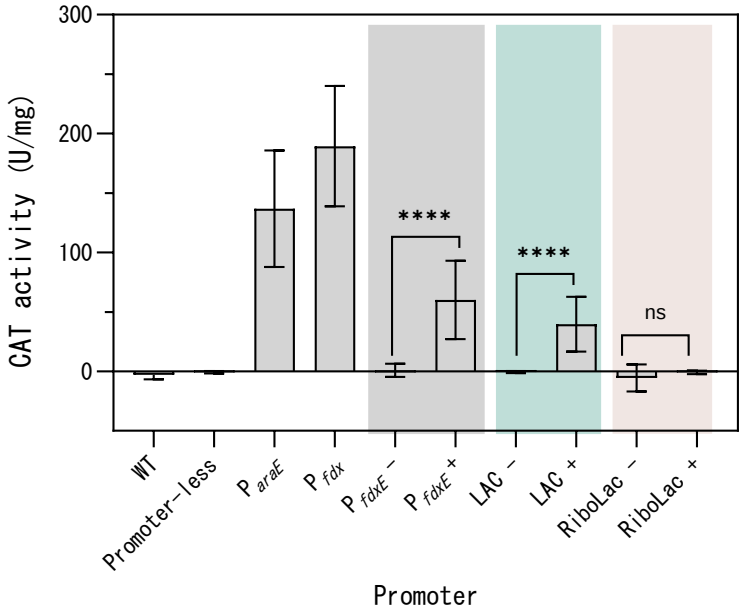


Figure 4.26 CAT reporter assay for clostridial promoters in *C. butyricum*. *C. butyricum* harbouring the following reporter plasmids was cultured for 5 h in RCM liquid medium + erythromycin: pMTL82254, pMTL82151_CAT_araE, pMTL-IC101, pMTL-IC111-E, HZ1 and pMTL-ICE1. Theophylline (5 mM), lactose (10 mM) or both were inducers added at 2 h, where indicated (- = without inducer, + = with inducer), for P_{fdxE}, LAC and RiboLac respectively. CAT activity was measured at 412 nm and standardised to protein concentration. Readings were blanked to blank RCM liquid media. Statistical analysis was carried out to compare non-induced and induced conditions (unpaired two-tailed Student's t-test); p-values are indicated as non-significant (ns), or $p<0.0001$ (****). Data represent mean values of three independent cultures in technical triplicate ± SD.

Both P_{araE} and P_{fdx} were confirmed to be constitutively active, producing the highest CAT activity of the 5 promoters assayed. P_{fdxE} and the LAC system showed good levels of inducibility, as expected, with limited background activity without inducer. Activity with inducer was significant for both promoters ($p<0.0001$), with P_{fdxE} having a slightly higher level of activity of induction than LAC, although only approximately a third of the activity of its constitutive counterpart P_{fdx}. There was no activity of RiboLac observed, even on induction with theophylline and lactose ($p=0.2421$). This is unexpected given that functionality is shown of both 'Ribo' and 'Lac' individual components under test conditions (P_{fdxE} and LAC). It could be due to a low level of CAT activity produced that is below the lower limit of quantification of the assay; the

assay would require further characterisation to determine its sensitivities in *C. butyricum*.

The CAT assay gave more reproducible data than that of the FAST assay, and demonstrated activity of the 'positive control' P_{fdxE}, so it was therefore deemed suitable for further use in *C. butyricum*. It should be noted, however, that there was some variation observed in protein concentrations, particularly across assay repeats. This is perhaps due to difficult lysis of *C. butyricum*, which has been observed anecdotally in other areas of this work. Attempts were made to improve the lysis step, for example in using different lysis buffers and different lysis conditions, however, no improvement was seen. This highlights the importance of the BCA assay and protein normalisation, and caution should perhaps be taken to compare promoter activity across assays, and limit comparisons to within an assay.

Overall this promoter assay highlights two constitutive and two inducible promoters to use for exogenous gene expression in *C. butyricum*: LAC, P_{fdxE}, P_{araE} and P_{fdx}.

4.4.3 Promoter characterisation and suitability

The CAT activity assay was carried out under one set of conditions to allow initial identification of functional promoters. Following this, further characterisation of the inducible promoters was explored to expand their utilisation.

PfdxE

Firstly theophylline toxicity to *C. butyricum* was determined with an MIC assay (Figure 4.27A), and then growth of the strain harbouring the P_{fdxE} reporter plasmid (pMTL-IC111-E) was assessed at different theophylline concentrations (Figure 4.27B).

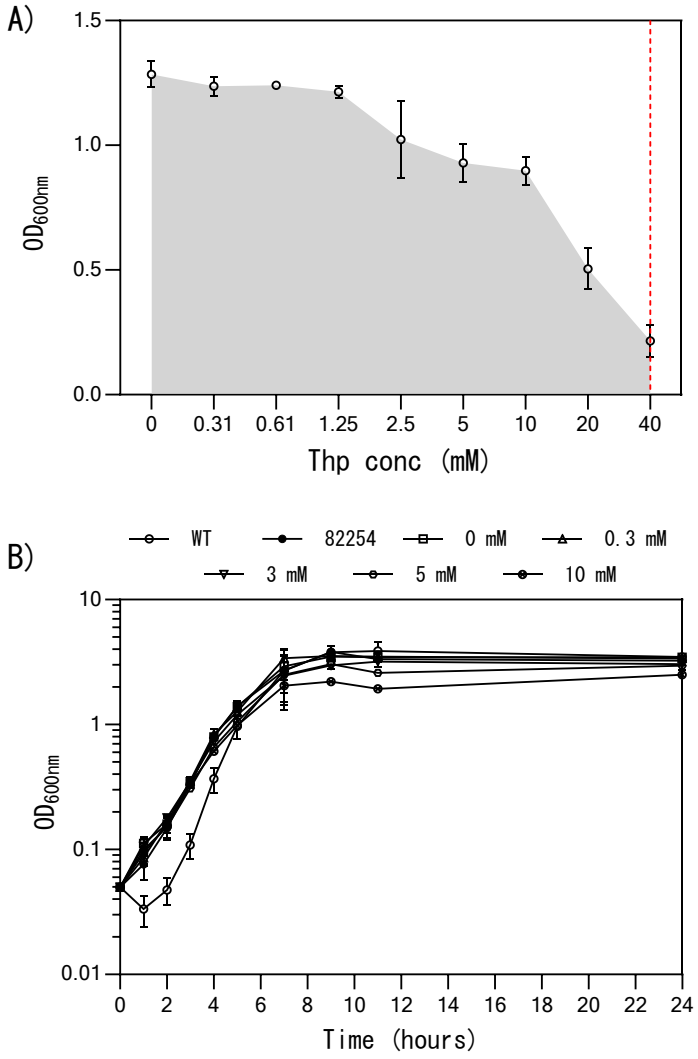


Figure 4.27 Determining suitability of theophylline for use with *C. butyricum*. A) Minimum inhibitory concentrations of theophylline. *C. butyricum* WT strain cultured with a range of theophylline concentrations. Growth measured after 24 h by OD_{600}. B) Growth of *C. butyricum* P_{fdxE} reporter plasmid in RCM + erythromycin selection and varying concentrations of theophylline inducer. Growth measured by OD_{600}. Promoter-less control pMTL82254. Data represent mean values of three independent cultures ± SD.

A dose response was observed at higher concentrations of theophylline, with a reduction in growth from approximately 1.25 mM and an estimated MIC of 40 mM (Figure 4.27A). This allows use of the 5 mM working concentration of theophylline without seriously impacting growth of *C. butyricum*. Moreover, growth of pMTL-IC111-E strain was not affected by theophylline concentration, with all growth curves

following the same profile for 0 mM to 10 mM. It was therefore deemed suitable to assess CAT activity at these concentrations, as samples taken at 5 h will be at the same growth phase regardless of theophylline concentration. The CAT assay was conducted as before for *C. butyricum* harbouring pMTL-IC111-E, with different concentrations of theophylline added at 2 h (Figure 4.28).

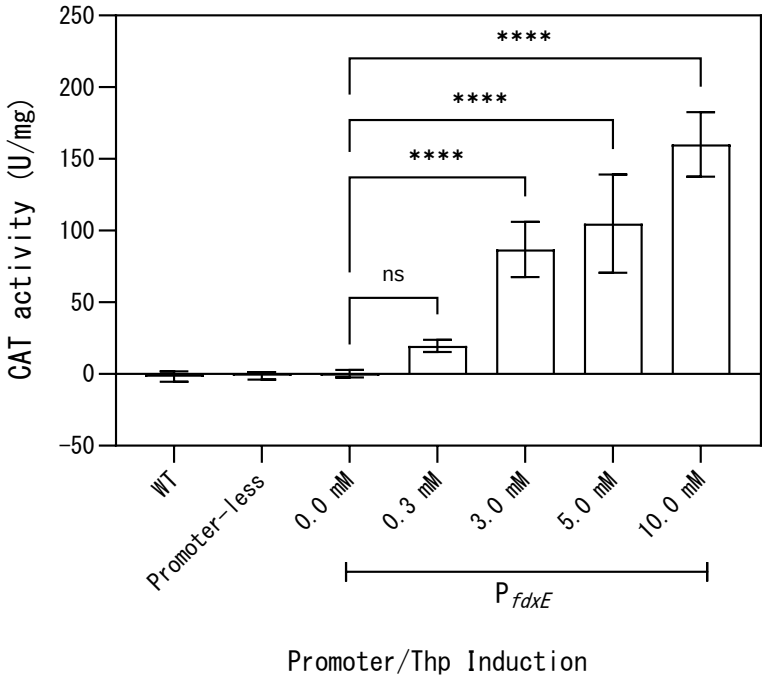


Figure 4.28 P_{fdxE} dose response to theophylline induction. *C. butyricum* harbouring the pMTL-IC111-E (P_{fdxE}) reporter plasmid was cultured in RCM liquid medium + erythromycin. Theophylline at varying concentrations was added at 2 h and samples taken at 5 h. Promoter-less= pMTL82254. Statistical analysis was carried out using one-way ANOVA with Dunnett's test for multiple comparisons against the non-induced (0 mM); p-values are indicated as non-significant (ns), or p<0.0001 (****). Values are blanked to RCM liquid medium. Data represent mean values of three independent cultures in technical triplicate ± SD.

A dose response was observed for increasing theophylline concentrations from 3 mM, with the CAT activity of 3, 5 and 10 mM significantly increased from the non-induced (one-way ANOVA with Dunnett's test for multiple comparisons against 0 mM; p<0.0001). With mean values of 87, 105 and 160 U/mg respectively, increasing the theophylline did increase promoter activity, although not in a linear fashion. This does,

however, afford the beneficial possibility to control levels of gene expression when under P_{fdxE}. Unfortunately CAT activity was not measured for P_{fdx}, which would have allowed a comparison to activity at higher levels of theophylline induction; in the previous assay activity by 5 mM was approximately two thirds less than P_{fdx} (Figure 4.26).

LAC

The dynamic response of the LAC system was also assessed with a CAT activity assay, under varying levels of lactose induction (Figure 4.29).

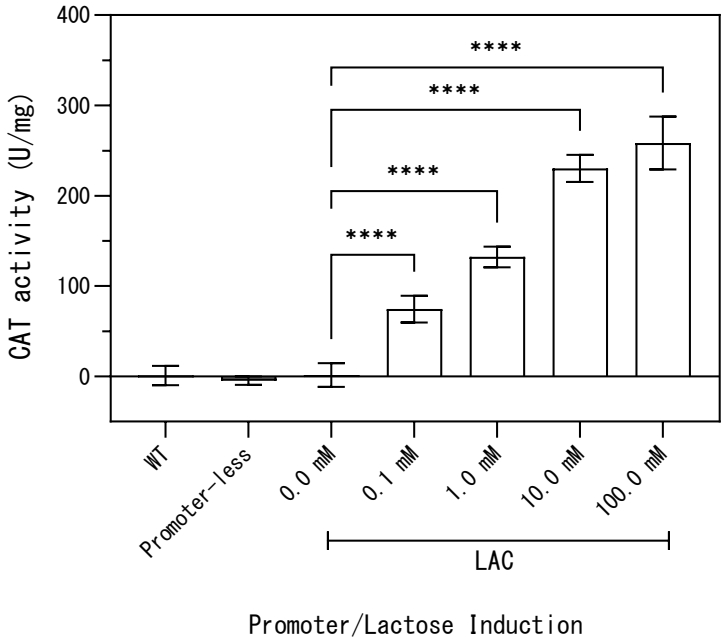


Figure 4.29 LAC dose response to lactose induction. *C. butyricum* harbouring the pMTL-HZ1 (LAC) reporter plasmid was cultured in RCM liquid medium + erythromycin. Lactose at varying concentrations was added at 2 h and samples taken at 5 h. Promoter-less= pMTL82254. Statistical analysis was carried out using one-way ANOVA with Dunnett's test for multiple comparisons against the non-induced (0 mM); p-values are indicated as $p<0.0001$ (****). Values are blanked to RCM liquid medium. Data represent mean values of three independent cultures in technical triplicate ± SD.

Similarly to P_{fdxE} the LAC expression system also exhibited a dose response to the lactose inducer, from as low as 1 mM lactose to 100 mM. CAT activity increased with

each 10-fold increase of inducer, but this tailed off at 100 mM; this was the smallest increase. This confirmed effective use of the 10 mM working concentration of the lactose inducer, and also demonstrated a dynamic range of the promoter which could be used in future applications.

Overall the work to characterise promoters has identified a suite of promoters available for future expression in *C. butyricum*, spanning a range of expression strengths, dynamic ranges, and inducibilities. This will be harnessed with the triple auxotroph strain developed in 4.3 for insertion and expression of *C. scindens bai* genes, allowing options to suit the desired expression needs.

4.5 Key Outcomes

The aim of this chapter was to characterise and establish *C. butyricum* as a chassis strain for gene insertion and heterologous expression. Following the establishment of DNA transfer by conjugation, attempts were made to maximise conjugation efficiency, by assaying various Gram-positive replicons and use of the sExpress *E. coli* donor to overcome RM systems. An RM system knockout was explored, although this surprisingly did not improve efficiency, as well as the loss of the native pCB102 plasmid, which did improve transfer of pMTL83151. The RiboCas system was then demonstrated in *C. butyricum* for the first time, being used to generate three auxotrophic gene truncations, at *pyrE, purD* and *hisI*, all of which demonstrated the expected auxotrophic phenotype on minimal media which was rescued with appropriate supplementation. ACE technology was then exemplified and used to repair the truncations, leading to a return to prototrophy, demonstrating the suitability of the knockout strains for cargo insertion by ACE. RiboCas was then used to create a double and triple auxotroph strain of *C. butyricum* for insertions at multiple loci. Attempts were then made to generate plasmids to characterise the lactose inducible *tcdR*/P_{tcdB} system in *C. butyricum* to allow expression across multiple loci, however it was not possible to achieve this. A reporter assay was therefore used to establish the expression of other clostridial promoters in *C. butyricum*, demonstrating the utility of a range of promoters including the constitutive P_{araE} and P_{fdx}, and the inducible P_{fdxE} and LAC systems, for future expression.

Chapter 5 Utilising *Clostridium butyricum* to generate a model of 7α-dehydroxylation from *Clostridium scindens*

5.1 Introduction

5.1.1 7α-dehydroxylation and colonisation resistance by C. scindens

C. scindens was first isolated in 1973 from the faeces of a patient with colon cancer and characterised by Holdeman and Moore (1973). It was soon implicated in bile acid conversion when it was shown to have an inducible bile acid 7α-dehydroxylase activity (White, Lipsky, Fricke, & Hylemon, 1980). Its process of 7α-dehydroxylation has been recognised as a multi-step pathway (Coleman, White, Egestad, Sjövall, & Hylemon, 1987), and this has led to the extensive characterisation of the *C. scindens* VPI 12708 *bai* operon structure and function (discussed in 1.2.2). The importance of 7α-dehydroxylation became more prominent several years later following the identification that microbial environments containing 7α-dehydroxylating bacteria, and high levels of secondary BAs, were more resistant to CDI (Rajani Thanissery et al., 2017; Theriot et al., 2016, 2014). This was suggested to be due to the toxic effects of secondary BAs to *C. difficile* vegetative growth (Theriot et al., 2014), and their inhibition of spore germination (Giel et al., 2010; Rajani Thanissery et al., 2017).

Until recently the protective role of *C. scindens* against *C. difficile* was only speculative. This proposed role was based on its high 7α-dehydroxylation activity, >10 times higher than other clostridial species (Doerner et al., 1997). It was not until the pivotal work of Buffie *et al.* (2015) that *C. scindens* was identified as a probable single species mediator of colonisation resistance against *C. difficile*, through its bile acid modulating activity, and since then several other research groups have characterised its protective effects (discussed in 1.2.2).

C. scindens is an organism of great interest, for both its use as the model organism for the 7α-dehydroxylation pathway and its importance in CDI resistance. However, direct studies of the *bai* operon have been prevented due to its genetic intractability.

5.1.2 Barriers against genetic manipulation of C. scindens

Whilst genetic modification of clostridia has been achieved in many species, prevention of DNA transfer as a result of host restriction modification (RM) systems is a common barrier. These systems identify foreign DNA through the recognition of

foreign methylation patterns, where expected modification motifs are absent, or modifications are present in non-native locations. Once identified, foreign DNA is degraded, and this can impede genetic modification efforts. Therefore, understanding of the RM systems present in host bacteria is required to predict these barriers and allow potential circumvention, for example through *in vitro* or *in vivo* methylation of exogenous DNA in appropriate patterns (S A Kuehne, Rood, & Lyras, 2019; Woods et al., 2019).

RM systems provide defences for a wide variety of unicellular organisms and are comprised of two main enzymes: a methyltransferase (Mase) and a restriction endonuclease (REase). The Mase allows discrimination between host and foreign DNA by transferring methyl groups to cytosine (to the C5 carbon or N4 amino group) or adenine (N6 amino group) residues, always to the same specific DNA sequence within the genome of the host. The REase identifies and catalyses endolytic cleavage of phosphodiester bonds in the foreign DNA (Vasu & Nagaraja, 2013). RM systems can be classified into four categories, Type I-IV, based on their cleavage position, target sequence and subunit composition in addition to cofactors and substrate specificity. Types I-III target DNA lacking host methylation patterns, and Type IV target foreign modification patterns (Vasu & Nagaraja, 2013; Woods et al., 2019). RM systems can be extremely diverse with nearly 4,000 enzymes known and their high frequency of occurrence demonstrates their success as a defence; approximately 90% of sequenced bacterial genomes contain one RM system, with 80% containing multiple systems (Roberts, Vincze, Posfai, & Macelis, 2015; Vasu & Nagaraja, 2013).

The inability to transfer DNA in *C. scindens* is most likely a consequence of the presence of RM systems. Following assembly and annotation of the genome of *C. scindens* ATCC 35704, two type I, twelve type II and one type IV RM systems were identified (Dempster, 2017). DNA transfer into the organism subsequently proved impossible, despite the implementation of a myriad of rational strategies to circumvent their effects, including conjugation and electroporation parameters, conjugal donors, and employing methyltransferase recombineering. In all transfer could not be demonstrated across eight different *C. scindens* strains (Dempster, 2017).

An alternative approach to study the 7α-dehydroxylation of *C. scindens* has been carried out by Funabashi et al. (2020). The use of a plasmid-based system for

expression of the *C. scindens bai* operon genes in *C. sporogenes* as a host organism avoids the need for DNA transfer. Following attempts at various promoter and *bai* gene fragment combinations the authors cloned the *bai* operon across three pMTL shuttle vectors: *baiB–baiF* in pMTL83153, *baiG* in pMTL83353 and *baiH–baiI* in pMTL83253. These were sequentially conjugated into *C. sporogenes*, and colonisation of germ-free mice demonstrated production of DCA, albeit at a substantially lower level than *C. scindens*. This was achieved despite all three plasmids utilising the same Gram-positive replicon and in the absence of antibiotic selection. Further shuttle vectors were also developed to delete individual genes, giving the authors a range of *bai* gene combinations in which to study BA intermediates and further elucidate the 7α-dehydroxylation pathway (Funabashi et al., 2020).

Whilst this work provides an important starting point for study of the *C. scindens bai* genes, the plasmid-based expression system limits the applications of the model in both *in vitro* testing and as a final probiotic therapy, due to the requirement for antibiotic supplementation to ensure plasmid maintenance. This chapter will therefore explore the insertion and expression of the *bai* genes in the genome of *C. butyricum*, capitalising on the development of ACE technologies and characterisation of promoter activity outlined in Chapter 4. In addition to the availability of genetic tools, the transfer of the pathway into *C. butyricum* presents further advantages over the *C. sporogenes* host in its recorded benefits on the gut microbiota, including colonisation resistance against *C. difficile* (Hagihara et al., 2021, 2020, 2018; Hayashi et al., 2021).

5.1.3 Objectives

-Insert the 8 genes responsible for 7α-dehydroxylation in *C. scindens*, from the major *bai* operon, into the *C. butyricum* genome using the genetic tools established in Chapter 4.

-Insert the 5 *bai* genes suggested to be essential for 7α-dehydroxylation from *C. scindens* into the *C. butyricum* genome

-Assess the heterologous expression of the inserted *bai* genes.

-Evaluate 7α-dehydroxylation in the *C. butyricum bai* strains and explore bile acid mediated resistance against *C. difficile.*

5.2 Insertion of the *bai* operon at *pyrE*

The simplest method of achieving 7α-dehydroxylation in *C. butyricum* is to insert the entire *C. scindens* bai operon into the genome, under its own promoter, to create a model organism. Due to the already extensive characterisation of *C. scindens* VPI 12708, the *bai* genes from *C. scindens* ATCC 35704 were chosen. The *bai* systems across the two strains share similar structures but have some differences (Figure 5.1), so study of *C. scindens* ATCC 35704 may expand current knowledge. It was first decided to insert the whole major *bai* operon downstream of *pyrE*, a well characterised locus, to limit gene disruption and avoid polar effects of gene insertion. Inserting the operon under its own promoter would allow biologically relevant levels of 7α-dehydroxylation to be studied.

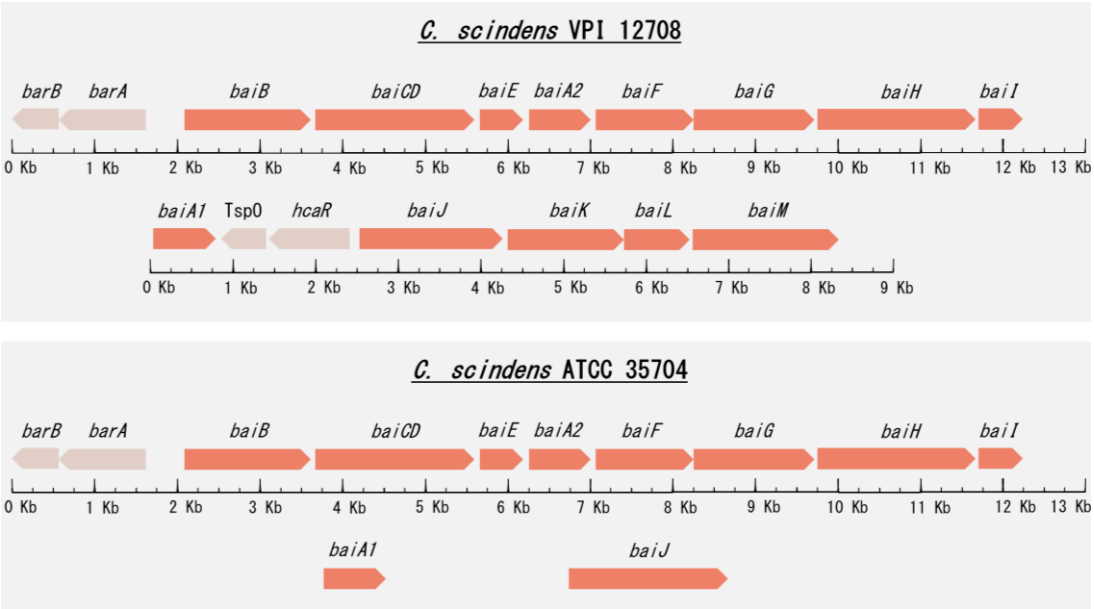


Figure 5.1 The *bai* operon of *C. scindens.* Comparison of the *bai* genes encoded by *C. scindens* VPI 12708 and ATCC 35704. Representation of the major *bai* operon *bai*BCDEA2FGHI proposed to be responsible for the 7α-dehydroxylation of CA to DCA, and the secondary *bai* genes proposed to be involved in bile acid metabolism.

5.2.1 Strategy 1: Whole operon insertion by RiboCas

The RiboCas system was chosen to insert the *C. scindens bai* operon into the *C. butyricum* genome, as it had previously been used successfully in this study. However, using the system as given was deemed unsuitable; the large size of pRECas_p19_MCS (9.2 Kb) limits the possible size of cargo to be inserted, especially considering the inclusion of homology arms, as it is approaching the approximate plasmid size that would likely reduce successful conjugation (~13 kB). This allows for cargo insertions of only approximately 2 Kb and would therefore require six sequential RiboCas plasmids for insertion of the whole *bai* operon.

Due to these size restrictions it was decided to split the RiboCas system into two plasmids: a shuttle vector containing the editing template (homology arms and cargo), and a Cas vector, containing *cas9* and guide RNA under their respective promoters of the RiboCas system (Figure 5.2). The pMTL vector backbones were chosen with Gram-positive replicons that yielded high transfer efficiency: pMTL82251 for the shuttle vector (pBP1 and *ermB*); pMTL87151 for the Cas vector (p19 and *catP*). It was decided to split the *bai* operon in half across two shuttle vectors, inserting the first downstream of *pyrE* and the second directly adjacent, to reproduce the native operon. To allow insertion of the next operon fragment, a bookmark was included to allow targeting of Cas9.

As outlined below three strategies arose in order to achieve *bai* operon cloning (strategies 1A-C). The cloning procedure was consistent for the shuttle and Cas9 plasmids throughout. For the shuttle plasmids pMTL82251 was digested with NotI and XhoI and the three fragments required (LHA, *bai* genes and RHA) were generated by PCR then inserted by hifi assembly. To keep plasmid size to a minimum the homology arms were chosen to be approximately 300 bp, and a bookmark was inserted by inclusion on primers. For the Cas vector, the sgRNA was generated by a primer dimer PCR, digested with SalI and AscI, and inserted by ligation into pRECas_p19_MCS digested with the same restriction enzymes. The primers used for cloning are listed in Table 5.2, Table 5.4 and Table 5.6, alongside the outline of each strategy.

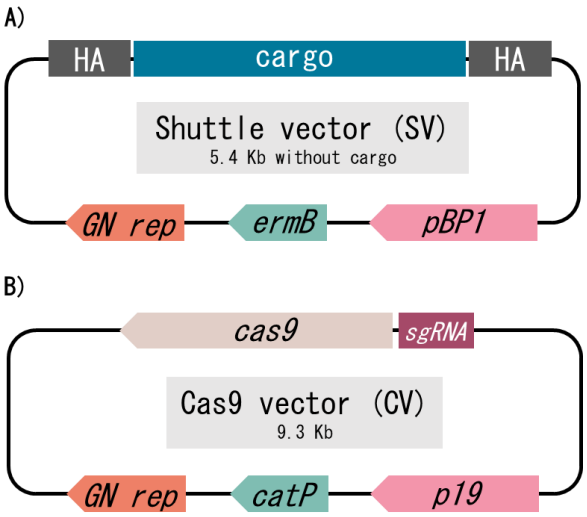


Figure 5.2 Two plasmid RiboCas system for cargo insertion. A) Shuttle vector (SV). pMTL82251 backbone with desired cargo and homology arms for insertion. B) Cas9 vector (CV). pMTL87151 backbone with *cas9* and sgRNA. Promoters and restriction sites for the editing template (cargo + HAs), *cas9*, and sgRNA insertion same as in pRECas1_MCS. GN rep = Gram-negative replicon, ColE1 + *traJ*.

Strategy 1A

It was first decided to attempt to split the *bai* operon across two shuttle plasmids by dividing it equally into two (Table 5.1 and Table 5.2). This would reduce the number of conjugations required into *C. butyricum* and reduce the overall manipulations required, keeping the risk of SNPs and other genetic variants to a minimum.

Table 5.1 Strategy 1A for cloning of the *C. scindens bai* operon in *E. coli*. Shuttle vectors (SV) containing the bai operon, and Cas9 vectors (CV) for selection of genome insertion.

Shuttle vector (SV)	Insert genes	Insert length (bp)	Bookmark
1.0	*barB, barA, baiB, baiCD, baiE*	6103	BM8
2.0	*baiA2, baiF, baiG, baiH, baiI*	6150	None

Cas9 vector (CV)	Target	Guide SEED sequence	Position	Strand	PAM	On-target score	Off-target score
1.0	Upstream of *pyrE*	ATAATAAACTAGAAGAATAA	2786095	-	GGG	44.4	96.3
2.0	BM8	TCCGGAGCTCCGATAAAAAA	-	+/-	TGG	-	-

Table 5.2 Primers used to clone plasmids for strategy 1A

Vector	Fragment	Size (bp)	Primers	Template
pSV1.0	LHA	306	SV1_LHA_F1 SV1_LHA_R1	*C. butyricum* DNA
	Insert	6103	SV1_BaiOp_F1 SV1_BaiOp_BM8_R1	*C. scindens* DNA
	RHA	305	SV1_RHA_BM8_F1 SV1_RHA_R1	*C. butyricum* DNA
pSV2.0	LHA	306	SV2_LHA_F1 SV2_LHA_R1	*C. butyricum* DNA
	Insert	6150	SV2_BaiOp_F1 SV2_BaiOp_R1	*C. scindens* DNA
	RHA	305	SV2_RHA_F1 SV2_RHA_R1	*C. butyricum* DNA
pCV1.0	sgRNA	152	CV1_guide_F1 CV1_guide_R1	Primer dimer
pCV2.0	sgRNA	152	CV2_BM8_F1 CV2_BM8_R1	Primer dimer

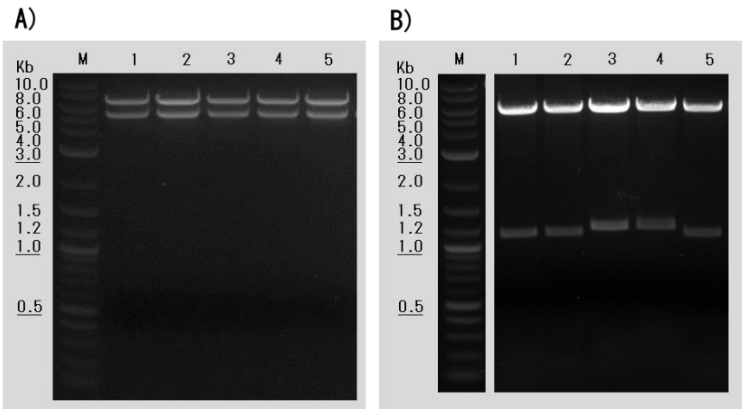


Figure 5.3 Confirmatory digests for creation of pSV1.0 and pSV2.0 Agarose gel electrophoresis visualisation of digests with NotI and XhoI to assess for *C. scindens bai* gene insertion into pMTL82251. A) Lanes 1-5 = SV1.0 transformants, all lanes demonstrate expected insert (6.7 Kb) and backbone (5.4 Kb). B) Lanes 1-5 = SV2.0 transformants, all lanes demonstrate unexpected band sizes for backbone and insert (correct sizes = 5.4 Kb and 6.7 Kb respectively). M = DNA Marker (2-log ladder; NEB).

Following transformation of the cloning reactions into *E. coli*, a confirmatory digest was carried out to assess the insert of pSV1.0 and pSV2.0 (NotI and XhoI) (Figure 5.3). The correct size was estimated for pSV1.0 (6.7 Kb), and the correct insert was confirmed by Sanger sequencing. However, the sizes of the two digested bands were incorrect for pSV2.0 (expected sizes 5.4 Kb and 6.7 Kb for backbone and insert, respectively), with a larger band of approximately 6 Kb and a smaller band ranging from 1.2 Kb – 1.4 Kb. Given that insertion of both homology arms alone would only produce a fragment of 600 bp, this suggests that recombination may have occurred which could be due to toxicity in *E. coli* of one or more genes in this half of the operon. Funabashi et al. (2020) also experienced effects of possible toxicity on their attempts at cloning the *C. scindens bai* operon in *E. coli*, and this was overcome by splitting the operon into three segments. Therefore, a second suite of shuttle vectors was designed to attempt to overcome toxicity.

Successful cloning of pCV1.0 and pCV2.0 was achieved, confirmed by colony PCR (sgRNA_F4 and p19_R1) followed by Sanger sequencing.

Strategy 1B

Funabashi et al. (2020) achieved successful cloning of the operon into *E. coli* by splitting it as follows: *baiB – baiF*, *baiG*, and *baiH – baiI*, and suggest that *baiG* confers toxicity. For Strategy 1B (Table 5.3) pSV1.0 was unchanged as complete cloning of the first half of the operon was already achieved, as was pCV1.0 and pCV2.0 as the sgRNA target sequence remained the same, *pyrE* and BM8 respectively (Table 5.4). However, the second half of the operon was split into two, interrupting *baiG* and preventing the entire gene being present in either plasmid. This hopefully allows cloning and should confirm whether *baiG* is responsible for the toxicity experienced in strategy 1A.

Table 5.3 Strategy 1B for cloning of the *C. scindens bai* operon in *E. coli*.

Shuttle vector (SV)	Insert genes	Insert length (bp)	Bookmark
1.0	*barB, barA, baiB, baiCD, baiE*	6103	BM8
2.1	*baiA2, baiF, baiG (first half)*	2846	BM4
3.0	*baiG (second half), baiH, baiI*	3304	None

Cas9 vector (CV)	Target	Guide SEED sequence	Position	Strand	PAM	On-target score	Off-target score
1.0	Upstream of *pyrE*	ATAATAAACTAGAAGAATAA	2786095	-	GGG	44.4	96.3
2.0	BM8	TCCGGAGCTCCGATAAAAAA	-	+/-	TGG	-	-
3.0	BM4	AGGGTTGTGGGTTGTACGGA	-	+/-	AGG	-	-

Table 5.4 Primers used to clone plasmids for strategy 1B

Vector	Fragment	Size (bp)	Primers	Template
SV2.1	LHA	300	SV2.1_LHA_F1 SV2.1_LHA_R1	*C. butyricum* DNA
	Insert	2846	SV2.1_BaiOp_F1 SV2.1_BaiOp_BM4_R1	*C. scindens* DNA
	RHA	305	SV2.1_RHA_BM4_F1 SV2.1_RHA_R1	*C. butyricum* DNA
SV3.0	LHA	300	SV3_LHA_F1 SV3_LHA_R1	*C. butyricum* DNA

	Insert	3304	SV3_BaiOp_F1 SV3_BaiOp_R1	*C. scindens* DNA
	RHA	305	SV3_RHA_F1 SV3_RHA_R1	*C. butyricum* DNA
CV3.0	sgRNA	152	CV3_BM4_F1 CV3_BM4_R1	Primer dimer

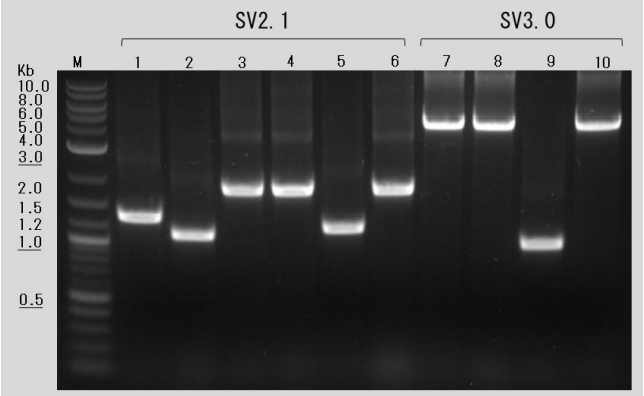


Figure 5.4 Colony PCR for creation of pSV2.1 and pSV3.0 Agarose gel electrophoresis visualisation of colony PCRs with flanking primers (pBP1_R1 and ColE1-tra_F2) to assess for *C. scindens bai* gene insertion into pMTL82251. Lanes 1 – 6 = SV2.1 transformants, all with incorrect insert (correct size = 4.0 Kb). Lanes 7 – 10 = SV3.0 transformants, with lanes 7, 8 and 10 demonstrating expected insert (correct size = 4.4 Kb). M = DNA Marker (2-log ladder; NEB).

Following transformation of the cloning reactions into *E. coli*, a colony PCR (pBP1 and ColE1-tra_F2) was carried out to assess the insert of pSV2.1 and pSV3.0 (Figure 5.4). The correct size was demonstrated for pSV3.0 (4.4 Kb) in three of four transformants, and the insert was confirmed by Sanger sequencing. However, the insert was incorrect for pSV2.1 (expected size 6.8 Kb), with a variety of sizes ranging 1.0 Kb – 1.8 Kb. This again suggests possible recombination and/or toxicity, narrowing down the likely responsible sequence to the genes present in pSV2.1: *baiA2*, *baiF* and *baiG* (first half).

Successful cloning of pCV3.0 was achieved, confirmed by colony PCR (sgRNA_F4 and p19_R1) followed by Sanger sequencing.

Strategy 1C

An alternative method of splitting the second half of the operon was attempted, and in a similar manner to Strategy 1B, *baiF* was interrupted to ascertain whether this gene was causing toxicity (Table 5.5).

Table 5.5 Strategy 1C for cloning of the *C. scindens bai* operon in *E. coli*.

Shuttle vector (SV)	Insert genes	Insert length (bp)	Bookmark
1.0	*barB, barA, baiB, baiCD, baiE*	6103	BM8
2.2	*baiA2*, *baiF* (first half)	1390	BM4
3.1	*baiF* (second half), *baiG, baiH, baiI*	4760	None

Cas9 vector (CV)	Target	Guide SEED sequence	Position	Strand	PAM	On-target score	Off-target score
1.0	Upstream of *pyrE*	ATAATAAACTAGAAGAATAA	2786095	-	GGG	44.4	96.3
2.0	BM8	TCCGGAGCTCCGATAAAAAA	-	+/-	TGG	-	-
3.0	BM4	AGGGTTGTGGGTTGTACGGA	-	+/-	AGG	-	-

Table 5.6 Primers used to clone plasmids for strategy 1C

Vector	Fragment	Size (bp)	Primers	Template
SV2.2	LHA	306	SV3.1_LHA_F1 SV3.1_LHA_R1	*C. butyricum* DNA
	Insert	1390	SV3.1_BaiOp_F1 SV3.1_BaiOp_R1	*C. scindens* DNA
	RHA	305	SV3.1_RHA_F1 SV3.1_RHA_R1	*C. butyricum* DNA
SV3.1	LHA	300	SV3.1_LHA_F1 SV3.1_LHA_R1	*C. butyricum* DNA
	Insert	4760	SV3.1_BaiOp_F1 SV3.1_BaiOp_R1	*C. scindens* DNA
	RHA	305	SV3.1_RHA_F1 SV3.1_RHA_R1	*C. butyricum* DNA

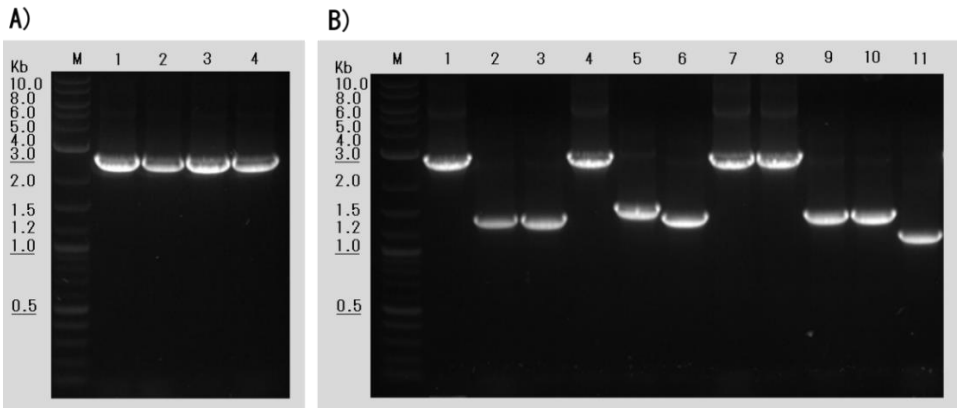


Figure 5.5 Colony PCR for creation of pSV2.2 and pSV3.1 Agarose gel electrophoresis visualisation of colony PCRs with flanking primers (pBP1_R1 and ColE1-tra_F2) to assess for *C. scindens bai* gene insertion into pMTL82251. A) Lanes 1-4 = SV2.2 transformants, all lanes demonstrate expected insert (2.6 Kb). B) Lanes 1-11 = SV3.1 transformants, all lanes demonstrate incorrect insert (correct size = 5.8 Kb). M = DNA Marker (2-log ladder; NEB).

Following transformation of the cloning reactions into *E. coli*, a colony PCR (pBP1 and ColE1-tra_F2) was carried out to assess the insert of pSV2.1 and pSV3.0 (Figure 5.5). The correct size was demonstrated for pSV2.2 (2.6 Kb) in all four transformants, and the insert was confirmed by Sanger sequencing. However, the insert was incorrect for pSV3.1 (expected size 5.8 Kb), with a variety of sizes ranging 1.2 Kb – 2.8 Kb. Combining this with results from pSV2.1 and pSV3.0 suggests that the cloning issue arises with either the second half of *baiF*, the first half of *baiG*, or a combination which has resulted from inclusion of the upstream LHA with homology to preceding *bai* operon genes.

Successful cloning of pCV3.1 was achieved, confirmed by colony PCR (sgRNA_F4 and p19_R1) followed by Sanger sequencing.

Testing two-plasmid RiboCas

As Strategy 1A had yielded both a correct Shuttle and Cas9 vector, it was decided to utilise these in assessing the functionality of a two-plasmid RiboCas system in *C. butyricum*. The final pSV1.0 and pCV1.0 plasmids were transformed into *E. coli* CA434 and conjugated into *C. butyricum* sequentially. Transconjugants for pSV1.0

were selected for with erythromycin and re-streaked to confirm plasmid presence. This strain was then used as the conjugal acceptor for pCV1.0, with transconjugants selected for with erythromycin and thiamphenicol and confirmed by re-streaking.

The *C. butyricum* strain harbouring both pSV1.0 and pCV1.0 plasmids was plated onto selection and theophylline (5 mM) to induce the RiboCas system. This yielded limited growth and colonies selected for screening by colony PCR (BUT_pyrE_F1 and BUT_pyrE_R2) demonstrated a WT genotype, with no successful insertions. The same results occurred following further overnight culture and plating with selection to extend the time available for homologous recombination.

As there was WT growth on theophylline, albeit limited, this suggests that the RiboCas is not functioning or is inefficient. Weak growth could be due to the burden of maintaining two large plasmids, especially one containing *cas9*, and is therefore not an indication of RiboCas selection. Given that a successful insertion for a P_{tet}/FAST cargo was later achieved with the one-plasmid RiboCas using the same homology arms and sgRNA (data not shown), failure in this instance could be due to either the *bai* gene cargo, size of cargo, or the two-plasmid vs one-plasmid system.

Given that it was not possible to generate a complete set of vectors for insertion of the *bai* operon by a two-plasmid RiboCas system, it was deemed inefficient to explore and troubleshoot the system further. Alternative methods would instead be investigated.

5.2.2 Strategy 2: Whole operon insertion by ACE and RiboCas

As successful cloning of the first half of the *bai* operon had already been achieved, it was decided to re-attempt to insert it through the use of ACE; ACE technology in *C. butyricum* was exemplified in Chapter 4 and allows insertions of larger cargo compared to RiboCas. As the results from strategies 1A-C suggested toxicity is arising from *baiF* or *baiG*, it was decided to insert *barB-baiA2* at *pyrE* with a bookmark and attempt to insert the remaining portion of the operon following troubleshooting to overcome toxicity.

ACE insertion

An ACE complementation vector for *barB-baiA2* insertion was generated using hifi assembly. The *bai* fragment was amplified from *C. scindens* DNA (RYPE_Sci_barB-baiA2_F1 and RYPE_Sci_barB-baiA2_R1) and inserted into pMTL-YN2C_PE digested with NotI and XhoI. Insertion was confirmed by colony PCR (ColE1_tra_F2 and pBP1_R1) and Sanger Sequencing. The final pMTL-YN2CPE_bBbA2 plasmid was then transformed into *E. coli* CA434.

The plasmid pMTL-YN2CPE_bBbA2 was conjugated into *C. butyricum ΔpyrE* with transconjugants selected for with thiamphenicol. Following confirmation by restreaking, transconjugants were then plated twice onto minimal media, and colonies screened for cargo insertion by colony PCR with flanking primers (BUT_pyrE_F1 and BUT_pyrE_R3) (Figure 5.6).

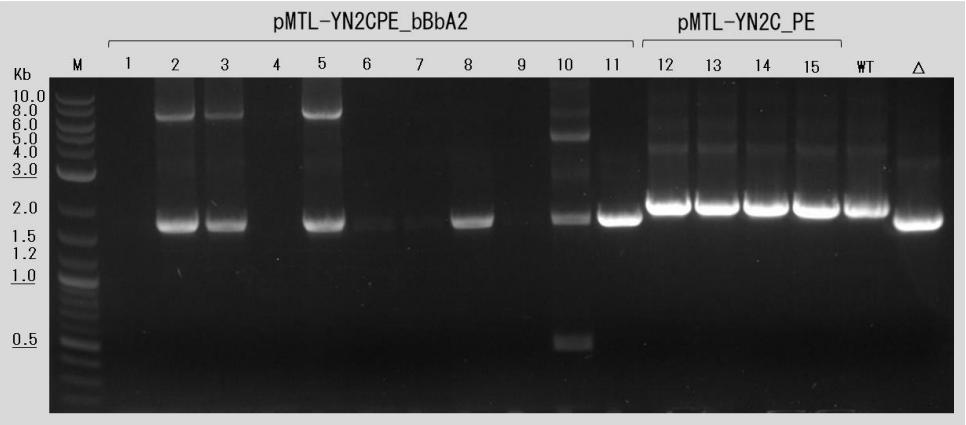


Figure 5.6 Colony PCR screening for *bai* operon insertion in *C. butyricum*. Agarose gel electrophoresis visualisation of colony PCRs with flanking primers (BUT_pyrE_F1 and BUT_pyrE_R3) to assess for insertion of *C. scindens barB-baiA2* into *C. butyricum* at *pyrE* by ACE. Lanes 1-11 = pMTL-YN2CPE_bBbA2, Lanes 12-15 = pMTL-YN2C_PE (control). Numbered lanes = transconjugant colonies plated onto minimal media. WT = wildtype *C. butyricum*, Δ = *C. butyricum ΔpyrE*, M = DNA Marker (2-log ladder; NEB). Expected fragment sizes (bp) Insert = 9254, WT = 2046, KO = 1809.

Colony PCR screening for pMTL-YN2CPE_bBbA2 identified 3 out of 11 colonies as a possible mixture of insertional mutants and the original acceptor strain *ΔpyrE*. The

remaining colonies screened either presented a *ΔpyrE* band only, or no band. The presence of a *ΔpyrE* band is unexpected as this genotype is not able to grow on minimal media, as shown in 4.3.2. The pMTL-YN2C_PE plasmid was also conjugated and screened as a positive control, and this demonstrated repair to WT in all colonies screened. The colonies without a band could either have been single crosses or a failed colony PCR reaction, which could be due to a lower rate of success because of the large insert requiring amplification.

Although it appears that the mixed colonies contain insertional mutants, the band produced is smaller than expected, at ~8 Kb vs 9.2 Kb. This is unlikely due to performance of the polymerase used (Q5 High-Fidelity) as this can produce amplicons of up to 20 Kb according to the manufacturer. Sanger sequencing confirmed repair of *pyrE* and insertion of both the 5' and 3' end of the *bai* operon fragment but was unfortunately inconclusive with regards to insert size due to repeated read failure. Moreover, regardless of this uncertainty the mixed colonies were further problematic as it was not possible to isolate the insertional mutant; further restreaks onto minimal media still demonstrated mixed bands on colony PCR screening.

Overcoming bai *operon toxicity*

To further investigate the toxicity issues arising in the second half of the *bai* operon, it was decided to clone *baiF* and *baiG* individually. Both genes were amplified (87_SCI_baiF_F1 and 87_SCI_baiF_R1 for *baiF*; 87_SCI_BaiG_F1 and 87_SCI_BaiG_R1 for *baiG*), digested with NotI and XhoI and ligated into pMTL87151 digested with the same restriction enzymes. Insertion was screened for by colony PCR (ColE1_tra_F2 and p19_R1). For *baiF* there was one transformant, this one colony exhibited the correct band size, and Sanger sequencing confirmed a fully correct *baiF* without SNPs. For *baiG* there were 8 transformants, with 3 exhibiting the correct band size, and Sanger Sequencing confirmed all 3 were the correct sequence with no SNPs. This suggests that toxicity may arise when the two genes are adjacent in the operon. A geneblock was therefore synthesised (2.4.10) for *baiA2-baiG*. This was amplified with primers for digestion and ligation into pMTL87151 (Geneblock_baiA2_F1 and Geneblock_baiG_R1; NotI and XhoI). Whilst the fragment was successfully amplified by PCR, there were no *E. coli* transformants.

The cloning results suggest that it is not possible to clone *baiF* and *baiG* adjacently in *E. coli* and therefore it will not be possible to insert them together into the genome; even if the individual genes are cloned separately, the upstream homology arm required for adjacent insertion by RiboCas will always result in at least some portion of the two genes encoded together in the editing template. Considering this with the inability to insert *barB-baiA2* by ACE at *pyrE*, it was decided to consider the alternative option of *bai* operon insertion and expression across multiple loci instead of at *pyrE* alone; this can be achieved using the triple auxotroph *C. butyricum* strain generated in Chapter 4. This should allow separation of *baiF* and *baiG*, and the use of ACE complementation vectors with fewer genes, hopefully resulting in higher conjugation efficiencies and improved insertion compared to pMTL-YN2CPE_bBbA2.

5.3 Insertion of the *bai* operon at multiple loci

5.3.1 Insertion of the whole operon

To allow insertion of the whole operon across three loci, and considering the cloning results of 5.2.1 and 5.2.2, it was proposed to divide the *bai* operon into the three sections outlined in Figure 5.7. Preliminary cloning was carried out to assess the ability to insert the chosen *bai* gene combinations into pMTL87151. The three inserts were generated by PCR (Sc_BO_1_F1 and Sc_BO_1_R1; Sc_BO_2_F1 and Sc_BO_2_R1; Sc_BO_3_F1 and Sc_BO_3_R1) and inserted into pMTL87151 (AatII and XhoI) by hifi assembly. Insertion was confirmed by colony PCR screening (ColE1_tra_F2 and p19_R1) and Sanger sequencing, and all three *bai* gene fragments were inserted successfully with no SNPs.

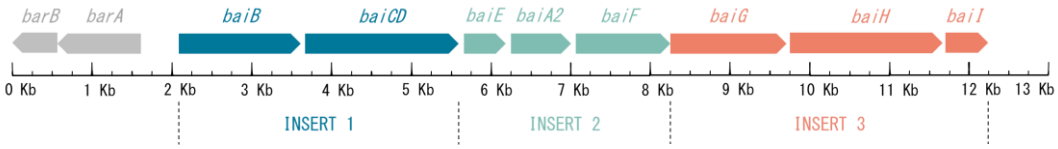


Figure 5.7 Division of the *bai* operon for chromosomal insertion. Proposed scheme of division for the *C. scindens bai* genes for insertion into the *C. butyricum* genome by ACE: 1) *baiB* and *baiCD*. 2) *baiE*, *baiA2* and *baiF*. 3) *baiG*, *baiH* and *baiI*. The promoter region upstream of *baiB* and its regulatory genes *barA* and *barB* will not be required due to use of an alternative promoter.

To allow insertion and expression into the genome at three loci it was attempted to insert the chosen operon fragments into the ACE complementation vectors (pMTL-YN2C_PE, pMTL-YN2C_PD and pMTL-YN2C_HI), under the control of a promoter. Promoter characterisation in *C. butyricum* in 4.4.3 identified two inducible promoters, the LAC promoter system and P_{fdxE}, and two constitutive promoters, P_{fdx} and P_{araE}, and it was decided to attempt multiple combinations of promoter and *bai* gene insertion due to the difficulties in cloning previously; the range of promoter strengths (Figure 4.26) should improve the chances of identifying at least one suitable promoter for each operon fragment. As it was hypothesised that strong *bai* gene expression could contribute to toxicity it was decided to try an alternative to $P_{fdx,}$ which was the strongest promoter tested. The P_{fdxOID} promoter was chosen to provide a further avenue of exploration to reduce toxicity in *E. coli*, by attenuating a promoter instead of replacing it. The P_{fdxOID} promoter contains a *lac* operator sequence at its transcriptional start site which results in repression of transcription by a $lacI^{Q}$ containing *E. coli* host, such as *E. coli* XL1-blue (sequence provided by Kovács et al. (2013)). This method was used successfully for the P_{thl} expression of cellulosomal operons in *C. acetobutylicum* (Kovács et al., 2013)

The cloning strategies attempted are outlined in Table 5.7. The promoter and *bai* operon fragments were first generated by PCR then insertion attempted into a digested complementation vector by hifi assembly. Due to difficulties with colony PCR, final plasmids were assessed by nanopore sequencing, and the outcomes for each promoter/*bai* gene combination are listed in the table.

Unfortunately, despite the multiple combinations attempted, a final plasmid was only generated for P_{fdxOID} + *baiGHI* (pMTL-YN2CPD_fOID_ins3) and P_{fdxE} + *baiGHI* (pMTL-YN2CHI_fdxE_ins3). Although a plasmid was also generated for expression of fragment 1 during cloning of the essential *bai* genes in 5.3.2, P_{araE} + *baiBCD* (pMTL-YN2CPD_aeBE3), the third and final fragment *baiEA2F* could not be cloned under any promoter. Although it was believed that attempting multiple cloning combinations would increase the chance of successful cloning, ultimately it was not possible to produce a suite of vectors for expression of all genes of the *bai* operon. It was therefore decided to focus efforts on the second theme of this chapter and study the genes hypothesised to be essential for 7α-dehydroxylation.

Table 5.7 Cloning of the *C. scindens bai* operon for insertion and expression across three loci. Insertion of promoter (LAC, $P_{fdxOID,}$ P_{araE} and P_{fdxE}) and *bai* gene fragments (1: *baiB, baiCD*; 2: *baiE, baiA2* and *baiF*: 3: *baiG, baiH* and *baiI*) into *C. butyricum* complementation vectors (pMTL-YN2C_PE, pMTL-YN2C_PD and pMTL-YN2C_HI) for insertion at *pyrE, purD* and *hisI* by ACE. Fragments generated by PCR by primers outlined and inserted into digested complementation vector by hifi assembly. Final vector analysed by nanopore sequencing.

Promoter	Insert	Promoter primers	Promoter template	Insert primers	Insert template	Backbone (Digest)	Outcome/ Final plasmid
LAC	1	HI_lac_ins1_prom_F2	pMTL-HZ1	HI_lac_ins1_bai_F1	*C. scindens* DNA	pMTL-YN2C_HI (NotI and XhoI)	No colonies from *E. coli* transformation
		HI_lac_ins1_prom_R1		HI_lac_ins1_bai_R1			
LAC	2	PE_lac_ins2_prom_F1	pMTL-HZ1	PE_lac_ins2_bai_F1	*C. scindens* DNA	pMTL-YN2C_PE (NotI and XhoI)	Deletion (~800 Kb) in LHA and no insertion of LAC. Possible major recombination.
		PE_lac_ins2_prom_R1		PE_lac_ins2_bai_R1			
LAC	3	PD_lac_ins3_prom_F1	pMTL-HZ1	PD_lac_ins3_bai_F1	*C. scindens* DNA	pMTL-YN2C_PD (NotI and XhoI)	No colonies from *E. coli* transformation
		PD_lac_ins3_prom_R1		PD_lac_ins3_bai_R1			
P_{fdxOID}	1	HI_fOID_ins1_prom_F1	pJ201_lacIQ_fdxOID	HI_fOID_ins1_bai_F1	*C. scindens* DNA	pMTL-YN2C_HI (SalI and XhoI)	Deletion (~90 bp) in promoter
		HI_fOID_ins1_prom_R1		HI_fOID_ins1_bai_R1			
P_{fdxOID}	2	PE_fOID_ins2_prom_F1	pJ201_lacIQ_fdxOID	PE_fOID_ins2_bai_F1	*C. scindens* DNA	pMTL-YN2C_PE (SalI and XhoI)	Insertion (~1.2 Kb) in LHA
		PE_fOID_ins2_prom_R1		PE_fOID_ins2_bai_R1			
P_{fdxOID}	3	PD_fOID_ins3_prom_F1	pJ201_lacIQ_fdxOID	PD_fOID_ins3_bai_F1	*C. scindens* DNA	pMTL-YN2C_PD (SalI and XhoI)	**pMTL-YN2CPD_fOID_ins3**
		PD_fOID_ins3_prom_R1		PD_fOID_ins3_bai_R1			
P_{araE}	2	PE_araE_ins2_prom_F1	pMTL82251_CAT_araE	PE_araE_ins2_bai_F1	*C. scindens* DNA	pMTL-YN2C_PE (NotI and XhoI)	Major recombination resulting in 3.8 Kb plasmid.
		PE_araE_ins2_prom_R1		PE_araE_ins2_bai_R1			
P_{fdxE}	2	PE_fdxE_ins2_prom_F1	pRECas1_MCS	PE_fdxE_ins2_bai_F1	*C. scindens* DNA	pMTL-YN2C_PE (NotI and XhoI)	Major recombination resulting in 3.8 Kb plasmid.
		PE_fdxE_ins2_prom_R1		PE_fdxE_ins2_bai_R1			
P_{araE}	3	HI_araE_ins3_prom_F1	pMTL82251_CAT_araE	HI_araE_ins3_bai_F1	*C. scindens* DNA	pMTL-YN2C_HI (NotI and XhoI)	SNP in *baiH*.
		HI_araE_ins3_prom_R1		HI_araE_ins3_bai_R1			
P_{fdxE}	3	HI_fdxE_ins3_prom_F1	pRECas1_MCS	HI_fdxE_ins3_bai_F1	*C. scindens* DNA	pMTL-YN2C_HI (NotI and XhoI)	**pMTL-YN2CHI_fdxE_ins3**
		HI_fdxE_ins3_prom_R1		HI_fdxE_ins3_bai_R1			

5.3.2 Insertion of the essential bai genes

Cloning of essential bai genes

The *C. scindens* major *bai* operon contains 8 *bai* genes that are proposed to be responsible for the 7α-dehydroxylation of primary bile acids and production of secondary bile acids, via a multi-enzyme conversion process. The roles of the 8 major *bai* enzymes have been elucidated, but the complete 7α-dehydroxylation pathway continues to be contested, with suggested enzyme redundancy, identification of additional *bai* genes, and unidentified enzymes in the pathway for secondary allo-bile acid production (Funabashi et al., 2020; Lee et al., 2022). In addition to this, it has been suggested that there is a pathway for 7α-dehydroxylation that does not require all major *bai* genes, and that there is a minimum set of enzymes capable of the conversion: *baiB, baiCD, baiE* and *baiJ,* plus the transporter *baiG* (Rizlan, B., personal communications, 2020). These essential genes are suggested to encode, respectively, a bile acid CoA ligase, a 3-dehydro-Δ^4-7α-oxidoreductase, a 7-dehydratase and a 3-oxo-Δ^4-5α-reductase (Lee et al., 2022).

To allow investigation into the set of putatively essential *bai* genes of *C. scindens,* it was proposed to generate 5 separate strains to allow a step-wise expression of each gene in the proposed pathway, building by one gene at a time until the final strain with all essential genes. This should allow bile acid conversions to be studied by mass spectrometry at each step of the proposed essential pathway. To do this cloning was undertaken to insert the genes into the ACE complementation vectors (pMTL-YN2C_PE, pMTL-YN2C_PD and pMTL-YN2C_HI), outlined in Table 5.8, with promoter and *bai* fragments inserted into digested complementation vectors by hifi assembly. The inducible P_{fdxE} promoter characterised previously was chosen to express the *baiG* transporter, to afford eventual control over 7α-dehydroxylation by attenuation of bile acid entry into the cell. The inducible LAC promoter system and the constitutive P_{araE} promoter were chosen as candidate promoters for cloning and expression of the other essential genes, with *baiBCDE* expressed together and *baiJ* expressed separately to maintain this similarity to *C. scindens*. The proposed insertion and resulting strains are outlined in Figure 5.8.

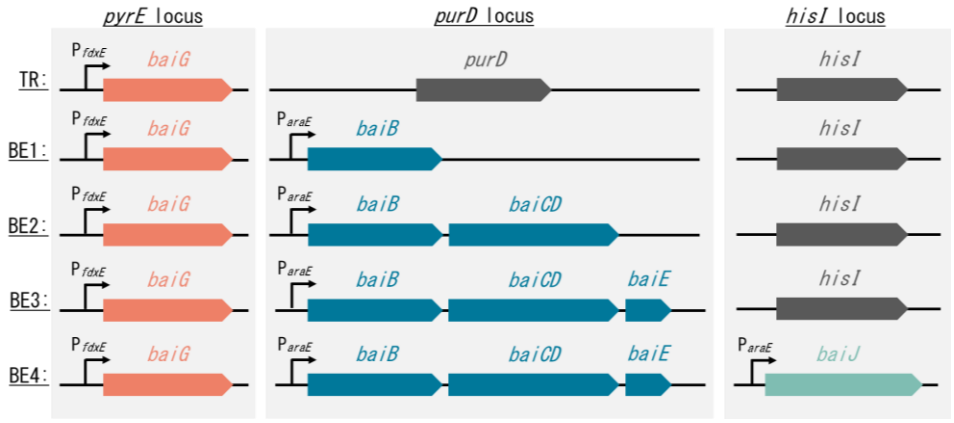


Figure 5.8 Proposed insertion of essential *bai* genes into *C. butyricum*. The 5 putatively essential *bai* genes for 7α-dehydroxylation in *C. scindens* inserted across 1-3 loci (*pyrE, purD* and *hisI*), in the *C. butyricum ΔpyrE, ΔpyrEΔpurD and ΔpyrEΔpurDΔhisI* strains. Insertions to be achieved by ACE technology. Five strains will allow the essential pathway to be studied in a step-wise manner: Transporter (TR) strain and four *bai* essential (BE) strains (1-4).

The P_{fdxE} + *baiG* complementation vector was successfully generated (pMTL-YN2CPE_baiG) for insertion at *pyrE*. For the remaining *bai* essential plasmids only BE1 and BE2 were generated under the LAC promoter system. Fortunately, the full set of BE plasmids was generated with expression driven by P_{araE}, producing the final vectors: pMTL-YN2CPD_aeBE1, pMTL-YN2CPD_aeBE2, pMTL-YN2CPD_aeBE3 and pMTL-YN2CHI_aeBE4. These were confirmed by nanopore sequencing and subsequently transformed into *E. coli* CA434.

Insertion of essential bai *genes*

To allow insertion of the essential *bai* genes into the *C. butyricum* genome, ACE was carried out using the *ΔpyrE, ΔpyrEΔpurD and ΔpyrEΔpurDΔhisI* strains generated in Chapter 4. The complementation vector was conjugated into the acceptor strain, selected for with thiamphenicol, and then minimal media without the respective supplement was used to select for return to prototrophy and cargo insertion. Colony PCR was then used to screen for *bai* gene insertion with flanking primers (*pyrE*: BUT_pyrE_F1 and BUT_pyrE_R3; *purD*: BUT_purD_F1 and BUT_purD_R1; *hisI*: BUT_hisI_F1 and BUT_hisI_R1), and insertion confirmed with Sanger sequencing. The plasmid was then cured and insertion reconfirmed with colony PCR.

Table 5.8 Cloning of the *C. scindens* essential *bai* genes for insertion and expression in *C. butyricum*. Insertion of promoter (P_{fdxE}, LAC and P_{araE}) and essential *bai* gene fragments (TR: *baiG*; 1: *baiB*; 2: *baiB* and *baiCD*; 3: *baiB, baiCD* and *baiE*; 4: *baiJ*) into *C. butyricum* complementation vectors (pMTL-YN2C_PE, pMTL-YN2C_PD and pMTL-YN2C_HI) for insertion at *pyrE, purD* and *hisI* by ACE. Fragments generated by PCR by primers outlined, and inserted into digested complementation vector by hifi assembly. Final vector analysed by nanopore sequencing.

Promoter	Insert	Promoter primers	Promoter template	Insert primers	Insert template	Backbone (Digest)	Outcome/ Final plasmid
P_{fdxE}	TR	PE_fdxE_G_prom_F1	pRECas1_MCS	PE_fdxE_G_bai_F1	*C. scindens* DNA	pMTL-YN2C_PE	**pMTL-YN2CPE_baiG**
		PE_fdxE_G_prom_R1		PE_fdxE_G_bai_R1		(NotI and XhoI)	
LAC	BE1	PD_LAC_BE_prom_F1	pMTL-HZ1	PD_LAC_BE1_bai_F1	*C. scindens* DNA	pMTL-YN2C_PD	**pMTL-YN2CPD_lacBE1**
		PD_LAC_BE_prom_R1		PD_LAC_BE1_bai_R1		(NotI and XhoI)	
LAC	BE2	PD_LAC_BE_prom_F1	pMTL-HZ1	PD_LAC_BE_bai_F1	*C. scindens* DNA	pMTL-YN2C_PD	**pMTL-YN2CPD_lacBE2**
		PD_LAC_BE_prom_R1		PD_LAC_BE2_bai_R1		(NotI and XhoI)	
LAC	BE3	PD_LAC_BE_prom_F1	pMTL-HZ1	PD_LAC_BE_bai_F1	*C. scindens* DNA	pMTL-YN2C_PD	No colonies from *E. coli* transformation
		PD_LAC_BE_prom_R1		PD_LAC_BE3_bai_R1		(NotI and XhoI)	
LAC	BE4	HI_LAC_BE4_prom_F1	pMTL-HZ1	HI_LAC_BE4_bai_F1	*C. scindens* DNA	pMTL-YN2C_HI	Insertion of BE4 fragment only
		HI_LAC_BE4_prom_R1		HI_LAC_BE4_bai_R1		(NotI and XhoI)	
P_{araE}	BE1	PD_araE_BE_prom_F1	pMTL82251_CAT_araE	PD_araE_BE1_bai_F1	*C. scindens* DNA	pMTL-YN2C_PD	**pMTL-YN2CPD_aeBE1**
		PD_araE_BE_prom_R1		PD_araE_BE1_bai_R1		(NotI and XhoI)	
P_{araE}	BE2	PD_araE_BE_prom_F1	pMTL82251_CAT_araE	PD_araE_BE_bai_F1	*C. scindens* DNA	pMTL-YN2C_PD	**pMTL-YN2CPD_aeBE2**
		PD_araE_BE_prom_R1		PD_araE_BE2_bai_F1		(NotI and XhoI)	
P_{araE}	BE3	PD_araE_BE_prom_F1	pMTL82251_CAT_araE	PD_araE_BE_bai_F1	*C. scindens* DNA	pMTL-YN2C_PD	**pMTL-YN2CPD_aeBE3**
		PD_araE_BE_prom_R1		PD_araE_BE3_bai_F1		(NotI and XhoI)	
P_{araE}	BE4	HI_araE_BE4_prom_F1	pMTL82251_CAT_araE	HI_araE_BE4_bai_F1	*C. scindens* DNA	pMTL-YN2C_HI	**pMTL-YN2CHI_aeBE4**
		HI_araE_BE4_prom_R1		HI_araE_BE4_bai_R1		(NotI and XhoI)	

Carrying out cargo insertion by ACE had mixed success. The strains TR, BE1, and BE3 were generated. However, attempts to generate BE2 were not successful; despite conjugation and growth on minimal media, colony PCR screening produced bands larger than the expected insert size, suggesting recombination. This continued to occur despite repeated attempts and confirmation of the correct plasmid sequence in *E. coli* CA434. This was unexpected, particularly as the BE3 strain, containing the same two *bai* genes plus an additional gene, was able to be generated. Moreover, generation of BE4 strain by ACE was also unsuccessful; pMTL-YN2CHI_aeBE4 was conjugated into the BE3 strain, but there was no growth on minimal media to select for the double crossover. To attempt to overcome this, a RiboCas vector was generated for insertion of the cargo by CRISPR.

Firstly, pRECasp19_HIins, a RiboCas vector for insertion downstream of *hisI* was generated. Homology arms were generated by PCR and inserted into pRECas_p19_MCS digested with SalI and AscI using hifi assembly (Table 5.9). An AatII site was included by primer design between the LHA and RHA to allow digestion of the backbone and insertion of an MCS (RE1_HIB_MCS_F1 and RE1_HIB_MCS_R1) by hifi assembly. The final plasmid pRECasp19_HIins was confirmed by colony PCR (sgRNA_F4 and p19_R1) and Sanger sequencing. The P_{araE} + *baiJ* insert was then generated by PCR from pMTL-YN2CHI_aeBE4 (Hiins_aeJ_F1 and Hiins_aeJ_F1) and inserted by hifi assembly into pRECasp19_HIins digested with SalI. The final plasmid, pRECas_HIins_BE4, was then confirmed by colony PCR and Sanger sequencing as above and transformed into *E. coli* CA434.

Table 5.9 Fragments used to generate pRECasp19_HIins. Primers used to generate the RiboCas pRECasp19_HIins with application specific elements for an insertion downstream of *hisI* in *C. butyricum*. Assembled by hifi PCR. LHA = Left homology arm. RHA = Right homology arm.

Fragment	Primers	Template
sgRNA	RE1_HIA_guide_F1	pCHRB_CD
(SEED = TGTACATCAATAAATCTTAA)	RE1_HIA_guide_R1	
LHA (506 bp)	RE1_HIA_LHA_F1	*C. butyricum* DNA
	RE1_HIA_LHA_R1	
RHA (500 bp)	RE1_HIA_RHA_F1	*C. butyricum* DNA

The pRECas_HIins_BE4 plasmid was conjugated into the *C. butyricum* BE3 strain (generated from the *ΔpyrEΔpurD* parent strain), selected for by thiamphenicol and the RiboCas system induced by plating onto theophylline. Colony PCR was carried out to screen for cargo insertion (BUT_HIscr_F1 and BUT_HIscr_F1), and successful insertion was confirmed. The plasmid was then cured and insertion reconfirmed to yield the final *C. butyricum* BE4 strain. Sanger sequencing confirmed correct insertion.

Although not all five *C. butyricum bai* essential strains could be generated, the four that could be (TR, BE1, BE3 and BE4) will allow this proposed pathway to be studied, particularly strain BE4 with all genes present.

5.3.3 Suitability of C. butyricum *for work with bile acids*

As 7α-dehydroxylation is not a native process in *C. butyricum* its ability to withstand growth with bile acids was assessed, to ensure its suitability as a model organism for 7α-dehydroxylation. The *C. butyricum* WT strain was grown with varying concentrations of cholic acid and deoxycholic in an MIC assay (Figure 5.9). Initial attempts to use RCM liquid media to reconstitute the bile acids posed problems, with the media becoming gelatinous and solid, which severely limited concentrations available for the assay. Bile acid salts are soluble in water but can aggregate and form micelles with facial amphiphilicity and they have been shown to form gels, influenced by a range of factors such as pH and availability of multivalent cations (Di Gregorio, Cautela, & Galantini, 2021). Because of this, 2YTg broth was assessed given it has a lower complexity than RCM but still allows *C. butyricum* growth, albeit at a slower rate (data not shown). There was a dose response for both CA and DCA, with MIC values estimated at 8 mM and 0.5 mM, respectively. From the graphs it was decided that a working concentration of 1 mM CA would be appropriate. For further work involving *C. butyricum* and BAs it would be useful to characterise growth rates under different BA concentrations, but this requires an anaerobic plate reader.

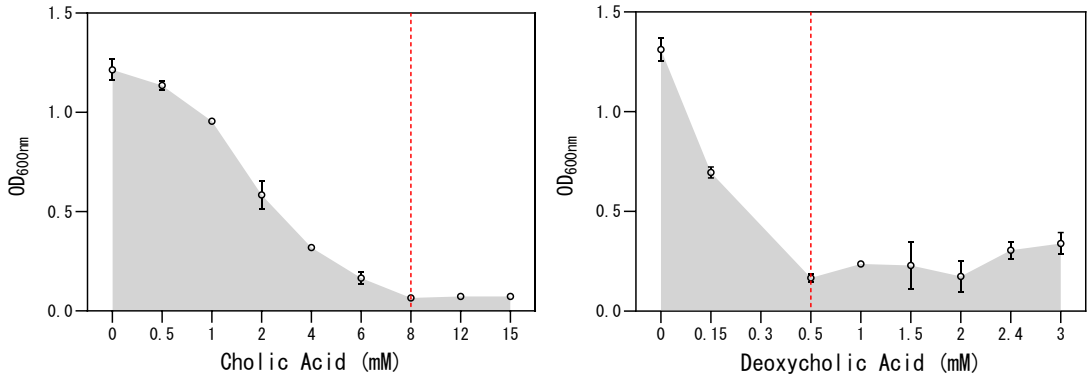


Figure 5.9 Minimum inhibitory concentrations of bile acids in *C. butyricum*. *C. butyricum* WT strain was cultured with a range of concentrations of cholic acid and deoxycholic acid in 2YTg liquid media. Starting from a 1 in 10 dilution of overnight culture, growth was measured after 24 hours by OD_{600}. Data represent mean values of three independent cultures ± SD.

5.4 Characterisation of the essential *bai* gene model in *C. butyricum*

5.4.1 Characterisation of bile acid transport

Although the *C. butyricum* BE4 strain provides an opportunity to assess the minimal requirements for 7α-dehydroxylation of CA, the prerequisite requirement for this to occur is the transport of CA into the cell; if the *baiG* transporter is non-functional the expression and function of the other *bai* genes in the model is irrelevant. Moreover, although the MIC of CA and DCA was established for the WT strain, the toxicity of BAs may alter when able to enter the cell. It was decided to assess both of these variables in the TR strain, firstly with a growth curve (Figure 5.10). The TR strain was grown in the presence of 5 mM theophylline (from 0 h) to allow expression of *baiG* and production of the transporter protein, and then CA (5 µM) was added at 5 h for transport into the cell. The TR strain without theophylline demonstrated a similar growth profile to the WT control, and the WT +Thp +CA control. The TR strain with theophylline induction demonstrated a slightly slower growth rate, which possibly demonstrates some burden from *baiG* expression. However, growth matched the control after 8 h and reached the same OD_{600} at stationary phase, even with addition of CA. The adverse effect of possible transporter expression and CA transport on *C. butyricum* growth was overall low, and is therefore suitable for study in the BE strains.

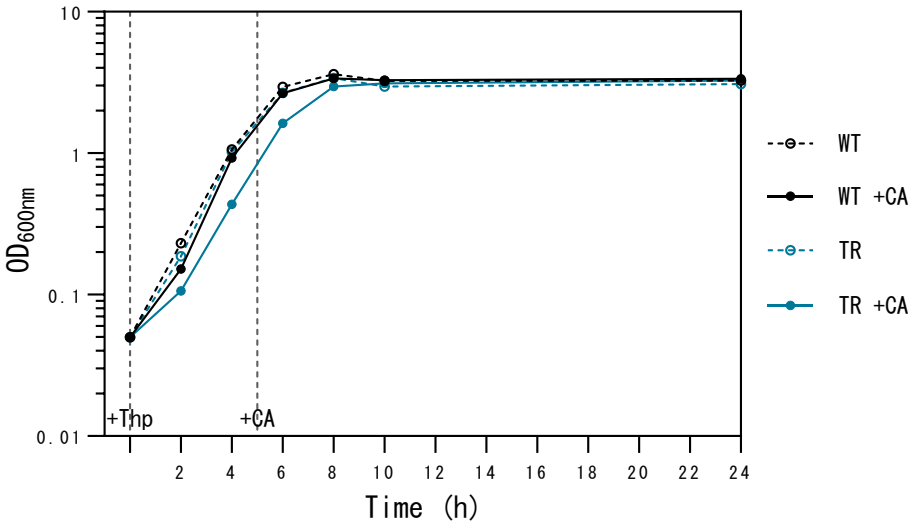


Figure 5.10 Growth curve of *C. butyricum* TR strain with cholic acid (CA). *C. butyricum* wildtype (WT) and TR strains grown in RCM media. Theophylline (Thp) inducer (5 mM) added to the +CA cultures at 0h, CA (5 µM) added at 5 h. Growth measured by OD_{600}. Data represent mean values of three independent cultures ± SD.

A preliminary assay was also undertaken to assess if CA transport into the cell could be measured. Previous work characterising *baiG* has been undertaken in *E. coli* (Darrell H. Mallonee & Hylemon, 1996a) where it was shown to transport the unconjugated primary bile acids CA and CDCA, but not the secondary bile acids LCA and DCA. There has been no study of *baiG* activity in a Gram-positive organism, despite Funabashi et al. (2020) expressing the *C. scindens baiG* gene on a shuttle plasmid in *C. sporogenes*; CA transport was only confirmed by DCA conversion, it was not studied as a standalone process. The *C. butyricum* TR strain, if expressing a functional BaiG protein, could therefore provide an opportunity to further study of CA transport. Confirming transport would also confirm the presence of a functional BaiG in the TR strain.

With limited work on BaiG, there is not a well-established transport assay. It was decided to first assess whether CA transport into the cell could be measured by a reduction in CA concentration in the media, before any further assay development. WT and RT strains were grown in RCM with 5 mM theophylline, and CA (5 µM) added after 5 h. Samples were taken 1.5 h later and the supernatant assessed for CA

concentration to assess for a reduction, and therefore entry into the cell (according to protocol 2.3.18; Figure 5.11). Mass spectrometry was carried out by Dr. David Tooth (SBRC, University of Nottingham).

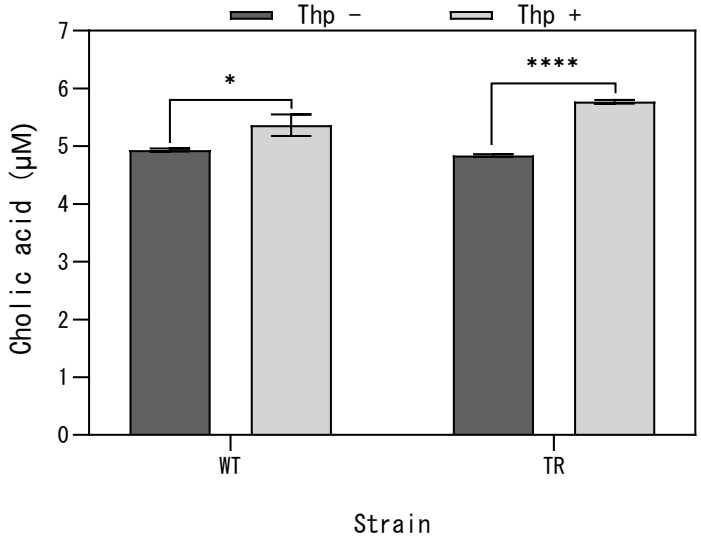


Figure 5.11 Cholic acid (CA) transport in the *C. butyricum* TR strain. *C. butyricum* wildtype (WT) and TR strains grown in RCM media with or without theophylline (Thp) inducer (5 mM). CA (5 μM) added at 5 h and concentration measured after 1.5 h. Statistical analysis was carried out to compare non-induced and induced conditions (unpaired two-tailed Student's t-test); p-values are indicated as $p<0.05$ (*), or $p<0.0001$ (****). Data represent mean values of three independent cultures ± SD.

The results obtained are unexpected for both the WT and the TR strain. Contrary to the hypothesised decrease in CA concentration for the TR strain, there is an increase in CA (unpaired two-tailed Student's t-test, $p<0.0001$) when comparing the uninduced and induced conditions. Moreover, one would expect the CA concentration to be consistent in the WT strain when comparing uninduced and induced conditions as there is no transporter present. However, they are statistically different (unpaired two-tailed Student's t-test, $p=0.0165$).

The results for both strains tested question the validity of this preliminary assay. It is difficult to explain why there would be an increase in CA in the media for either strain, but particularly for the WT, resulting in an inability to draw conclusions on CA transport with confidence. More insight into two factors is required to interpret these

inconclusive results: BaiG directionality and BA transport in *C. butyricum*. Whilst studies have shown BaiG transports unconjugated bile acids into *C. scindens* (Darrell H. Mallonee & Hylemon, 1996a), it is unclear whether the protein is also involved in export, with its absence shown to be associated with CA accumulation in cells (Thanassi, Cheng, & Nikaido, 1997). BaiG-independent export must also be considered, with bile acid efflux permeases shown to export bile acids in *E. coli* for example (Rosenberg, Bertenthal, Nilles, Bertrand, & Nikaido, 2003), and this has not previously been characterised in *C. butyricum*.

Overall these results do not show CA transport in the TR strain but import of CA by BaiG into the cell cannot be dismissed. In addition to the lack of knowledge into BaiG function, the assay requires further validation to ensure transport can be confidently assessed. This would require extensive characterisation, particularly of sample timing. The rate of CA transport by BaiG has not been well characterised, and this is crucial in choosing a sample timing to observe a CA reduction. Whilst the rate of CA to DCA conversion in other studies can suggest the rate of transport, for example 50% after 24 h in *C. scindens* (Marion et al., 2019), none have studied transport independently of 7α-dehydroxylation. Considering the unknown variables of transport directionality and timing, it is unclear whether a reduction in CA concentration is an appropriate marker for BaiG transport. The complexity of validating this assay was deemed outside the scope and capabilities of this project, and it was decided to use assays for 7α-dehydroxylation as an assessment of BaiG function.

5.4.2 Expression of bai *genes*

RT-PCR

Following insertion of the essential *bai* genes into the *C. butyricum* genome it was decided to validate their expression by RT-PCR. Primers were firstly designed using the PrimerQuest Tool by Integrated DNA Technologies (https://eu.idtdna.com/PrimerQuest/Home/Index) to generate primers with equal annealing temperatures and amplicon sizes for the *bai* gene targets: *baiG, baiB, baiCD, baiE* and *baiJ*. The *recA* gene, encoding recombinase A, was chosen as the housekeeping gene, having previously been used in qPCR studies for *C. butyricum* (So, Oh, & Shin, 2022). Primers were first validated using *C. scindens* and *C.*

butyricum WT DNA extractions as the template for the *bai* genes and *recA* respectively, and all amplicons were produced as expected (gel not shown).

To assess for gene expression, RNA extractions were carried out for the *C. butyricum* WT strain, and the BE4 strain (with and without 5 mM theophylline), according to 2.4.17. To assess for lack of DNA contamination a PCR of the RNA preparation was carried out using q5 DNA Polymerase and primers for *recA* (cD_recA_F1 and cD_recA_R1) (Figure 5.12). As no amplicon was produced, and was only observed in the DNA controls, the RNA extractions were deemed suitable and used as a template to generate cDNA according to 2.4.18. RT-PCR was then carried out using the cDNA templates to assay for expression of *bai* genes (Figure 5.13).

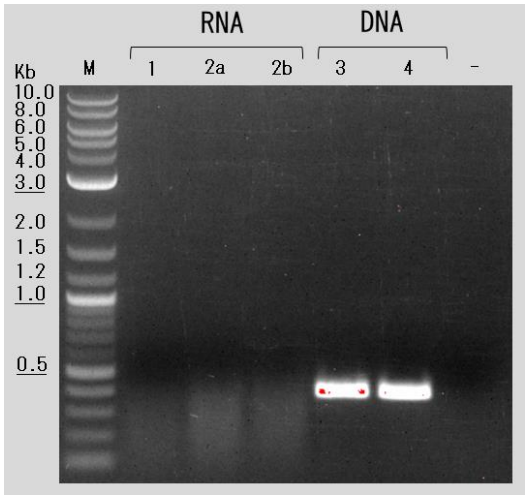


Figure 5.12 RNA extraction screening. Screening for DNA contamination with primers for *recA* (cD_recA_F1 and cD_recA_R1). Numbered lanes = RNA (1-2) or DNA (3-4) preparations of *C. butyricum* strains of interest. 1= wildtype RNA, 2a= BE strain without theophylline RNA, 2b= BE strain with theophylline RNA, 3= wildtype DNA, 4= BE strain DNA, - = negative dH_2O control. M = DNA Marker (2-log ladder; NEB). Expected fragment size (bp) = 395.

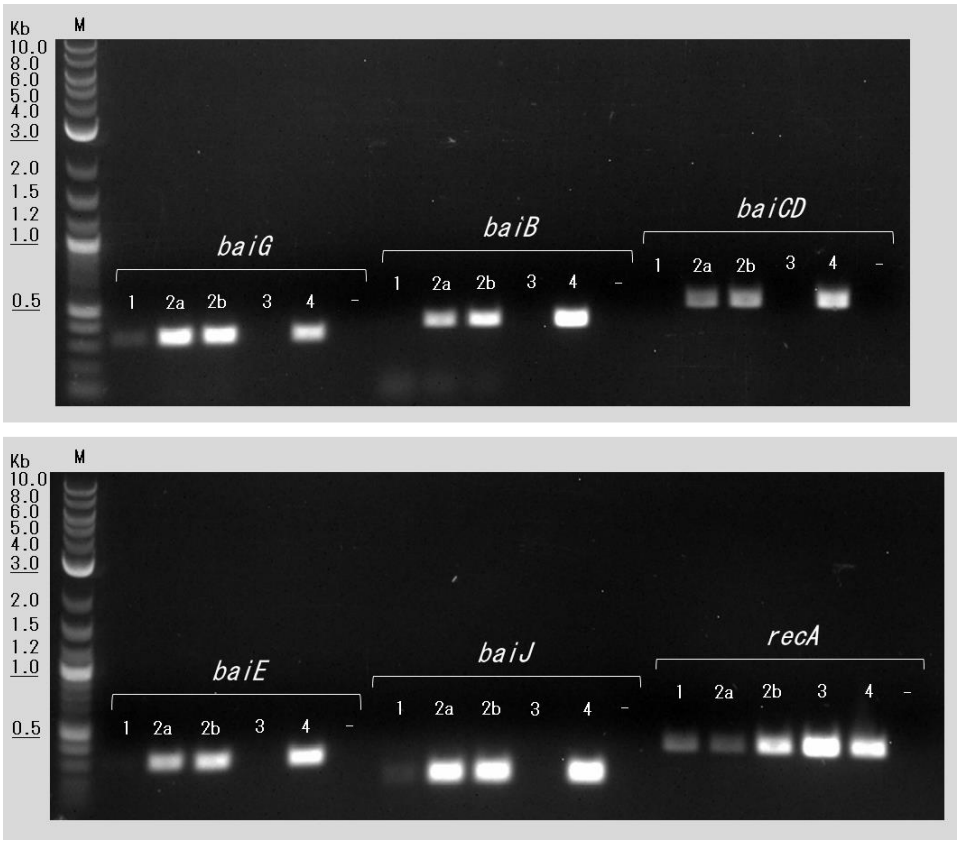


Figure 5.13 Heterologous expression of *C. scindens bai* genes in *C. butyricum*. PCR screening of cDNA preparations to determine gene expression. Numbered lanes = cDNA (1-2) or DNA (3-4) preparations of *C. butyricum* strains of interest. 1= wildtype cDNA, 2a= BE strain without theophylline cDNA, 2b= BE strain with theophylline cDNA, 3= wildtype DNA, 4= BE strain DNA, - = negative dH_2O control. M = DNA Marker (2-log ladder; NEB). Expected fragment sizes: *baiG* (cD_baiG_F1 and cD_baiG_R1) = 337 bp, *baiB* (cD_baiB_F1 and cD_baiB_R1) = 399 bp, *baiCD* (cD_baiCD_F1 and cD_baiCD_R1) = 521 bp, *baiE* (cD_baiE_F1 and cD_baiE_R1) = 315 bp, *recA* (housekeeping; cD_recA_F1 and cD_recA_R1) = 395 bp.

As expected the gel demonstrated expression of all five *bai* genes in the BE4 strain; there was a band present for all primer pairs using the BE4 + theophylline cDNA template (2b). Using the BE4 without theophylline cDNA template (2a) one would expect expression of the constitutive *bai* genes (B, CD, E and J), but no expression of *baiG* under the theophylline inducible $P_{fdxE.}$ There is, however, expression of *baiG* in 2a, which suggests there is some background expression from the promoter without induction.

With regards to the control screenings, the presence of the *recA* amplicon validates all cDNA and DNA preparations. There is no amplicon for *baiB*, *baiCD* or *baiE* in the WT cDNA preparations as expected, but there is a faint band for *baiG* and *baiJ* which is unexplainable as these genes are not present in the WT strain. This is unlikely due to false positives from the primer pairs as the absence of an amplicon for all *bai* genes in the WT DNA preparation (3) but presence in the BE4 DNA preparation (4) excludes this as a possibility. The faintness of these bands compared to the consistently bright bands for *bai* gene amplicons in 2a and 2b does not reduce confidence in the conclusion of successful *bai* gene expression.

To gain a further insight into the levels of expression of each of the *bai* genes qRT-PCR could be carried out to quantify expression. If carried out alongside *C. scindens*, this could allow a comparison to native expression. This would allow assessment of the impact that varying the promoter and operon structure has on expression levels, which could alter rates of 7α-dehydroxylation due to enzyme ratios.

Assessment of 7α-dehydroxylation

Following confirmation of *bai* gene expression, Bai protein function needed to be verified. The most well-established method to assess this is to carry out bile acid analysis and monitor the conversion of CA to DCA by mass spectrometry. This will also allow the study of intermediates generated by the different *bai* essential gene combinations encoded by strains BE1-4, and assess if the four genes are sufficient for 7α-dehydroxylation. This entails complex processes not available using the tools of the in-house team. The BE strains were therefore sent to our collaborators at the École Polytechnique Fédérale de Lausanne (EPFL) for bile acid analysis, and the work is ongoing. As the data was not available for this dissertation, it was hoped that functional assays assessing DCA production could provide an insight into 7α-dehydroxylation in the BE4 strain.

5.4.3 Preliminary assays for C. difficile *challenge assays*

The overall aim of *bai* gene heterologous expression in *C. butyricum* is to provide a vehicle to impart the hypothesised 7α-dehydroxylation-mediated resistance of *C. scindens* against *C. difficile*. Assessing the CA-mediated ability of the BE4 strain to

provide resistance could confirm both functional expression of the Bai proteins, and whether the putative essential genes are sufficient alone to 7α-dehydroxylate CA.

To establish whether the presence of the putative essential *bai* genes had altered the tolerance of *C. butyricum* to BAs, an MIC assay was carried out (Figure 5.14). There was a dose response for both CA and DCA, with MIC values estimated at 4 mM and 0.3 mM compared to 8 mM and 0.5 mM in the WT respectively. From the results 0.5 mM was chosen as an appropriate concentration of CA for *C. butyricum* BE4.

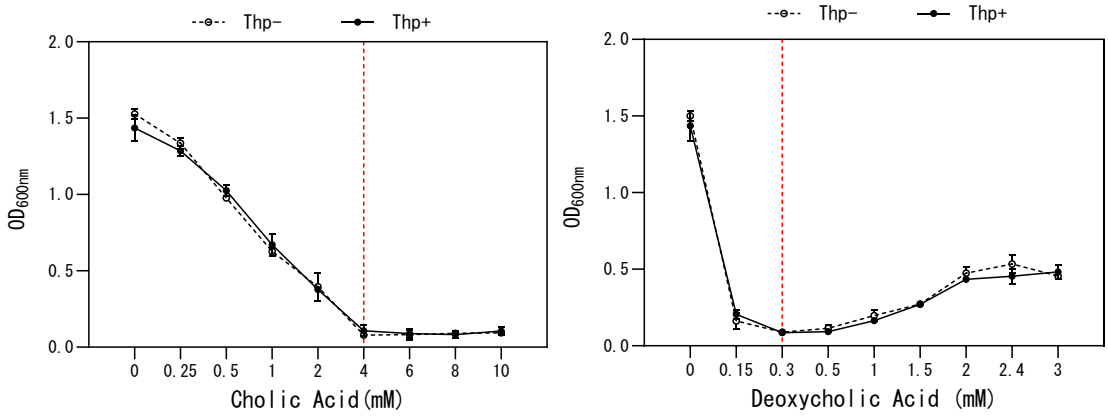


Figure 5.14 Minimum inhibitory concentrations of bile acids in *C. butyricum* BE4. *C. butyricum* BE4 was cultured with a range of concentrations of cholic acid and deoxycholic acid in 2YTg liquid media, with and without the theophylline inducer (5 mM). Starting from a 1 in 10 dilution of overnight culture, growth was measured after 24 h by OD_{600}. Data represent mean values of three independent cultures ± SD.

In preparation for supernatant assays the growth of the WT and BE4 strains was assessed with theophylline (5 mM) and CA (0.5 mM) (Figure 5.15). For the exponential phase there was a similar rate of growth for the WT control and BE4 across all three conditions, but the WT with Thp and CA had a slower rate of growth. Both strains reached stationary phase at approximately 8 h under all conditions, with the WT reaching a higher OD_{600} than BE4, of approximately 3.1 vs 2.2 respectively. The graph indicates that 6 h is an appropriate sampling time to maintain a similar growth phase and OD_{600} between the two strains across the three conditions.

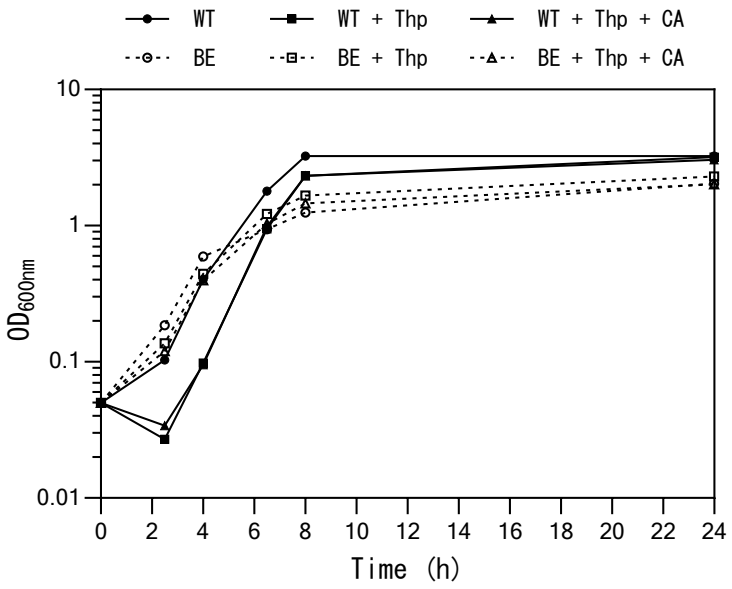


Figure 5.15 Growth curve of *C. butyricum* BE4. Growth of *C. butyricum* wildtype (WT) and BE4 (BE) strain in liquid RCM media + theophylline (5 mM) and cholic acid (CA; 0.5 mM). Growth measured by OD_{600}. Data represent mean values of three independent cultures ± SD.

As the *C. butyricum* strain requires the addition of theophylline for *baiG* expression, the toxicity of theophylline to *C. difficile* R20291 was assessed to ensure growth would not be affected by theophylline present in *C. butyricum* supernatants. An MIC assay was carried out (Figure 5.16), with the MIC estimated at 20 mM, deeming the working concentration of 5 mM suitable.

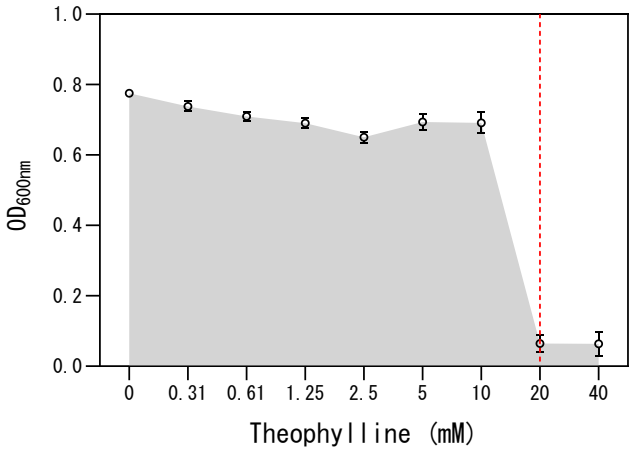


Figure 5.16 Minimum inhibitory concentrations of theophylline in *C. difficile* R20291. *C. difficile* was cultured with a range of concentrations of cholic acid and deoxycholic acid in BHIS liquid media. Starting from a 1 in 10 dilution of overnight culture, growth was measured after 24 h by OD_{600}. Data represent mean values of three independent cultures ± SD.

5.4.4 Supernatant Assays with C. difficile

C. scindens has been shown to offer colonisation resistance against *C. difficile* in a bile acid-dependent manner (Buffie et al., 2015). As it is hypothesised that the essential *C. scindens bai* genes encoded by *C. butyricum* BE4 are sufficient to enable 7α-dehydroxylation, the strain's ability to inhibit germination of *C. difficile* spores was tested.

Spore outgrowth

C. difficile spore outgrowth was measured to assess the impact of the *C. butyricum* BE4 supernatants on spore germination (Figure 5.17). With the strain incubated with CA, observations of reduced spore germination would indicate DCA production in *C. butyricum*, and therefore 7α-dehydroxylation by the essential *bai* genes. Supernatants were taken for growth of *C. butyricum* BE4 with and without CA (0.5 mM) and theophylline, and were prepared for the assay according to 2.3.11.The spore outgrowth assay was carried out according to 2.3.12.

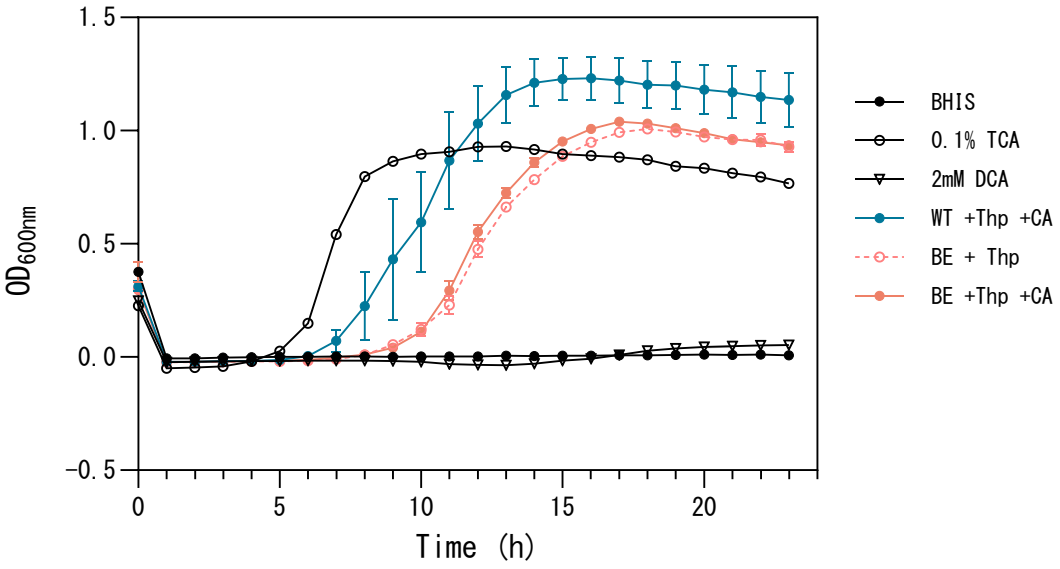


Figure 5.17 Spore outgrowth of *C. difficile* R20291. *C. difficile* spores were incubated with prepared supernatants from *C. butyricum* WT and BE4 strains incubated with theophylline (5 mM), and with and without cholic acid (0.5 mM). The germinant TCA (0.1%) was also incubated with all spores. Growth was measured by OD_{600} in a plate reader every hour for 23 h. Data represent mean values of spores incubated with three independent supernatant preparations, each in technical triplicate ± SD.

Outgrowth of spores incubated with the supernatant from the BE4 strain demonstrated a slower growth rate than the WT and reached a lower OD_{600} at stationary phase. Whilst this could suggest a reduction in germination this is unlikely to be resultant of 7α-dehydroxylation by the BE4 strain as spore outgrowth followed a similar profile with and without the inclusion of CA. This could suggest another mechanism of inhibition by the BE4 strain.

With regards to the controls, no spore growth was seen without the germinant TCA, or in the presence of DCA, as expected. However, the positive control of spores + 0.1% TCA presented a different outgrowth profile to that of the supernatants tested; outgrowth occurred earlier, and a lower OD_{600} was reached at stationary phase. This is unlikely to be an endogenous effect from the *C. butyricum* supernatants, as the *C. butyricum* MIYAIRI588 strain that protects against CDI was shown to do so via alternate mechanisms and does not directly inhibit the germination or growth of *C. difficile* (Hayashi et al., 2021). Moreover, if there was an endogenous effect in strain

7423 it would be more likely to inhibit and not improve spore outgrowth. It could therefore be as a result of the introduction of RCM media following *C. butyricum* incubation. Unfortunately, the data for *C. butyricum* WT +Thp indicated possible contamination so this data set was not available for comparison. Further work should therefore be carried out to characterise the impact of *C. butyricum* WT supernatant on *C. difficile* spore outgrowth, before conclusions can be drawn from this assay. Regardless, the results show that there is not a bile-acid dependent effect on germination with the BE4 strain, suggesting that 7α-dehydroxylation is not impacting germination. However, it cannot be said whether this is due to non-functioning Bai protein expression, that the BE4 proteins are functional but not sufficient for 7α-dehydroxylation, or whether 7α-dehydroxylation is occurring but not at sufficient levels for inhibition of germination.

Germination initiation

To further assess the BE4 strain for *C. difficile* resistance a germination initiation assay (2.3.13) was carried out using the same supernatant samples (Figure 5.18). Germination initiation occurred in all supernatant conditions, as indicated by an initial decrease in OD_{600}, and followed the same or similar profile to that of the positive control. This demonstrates that there is no endogenous mechanism preventing germination initiation in *C. butyricum* WT, nor is there a CA-dependent mechanism in the BE4 strain. This is unsurprising given that the DCA control also does not prevent germination initiation.

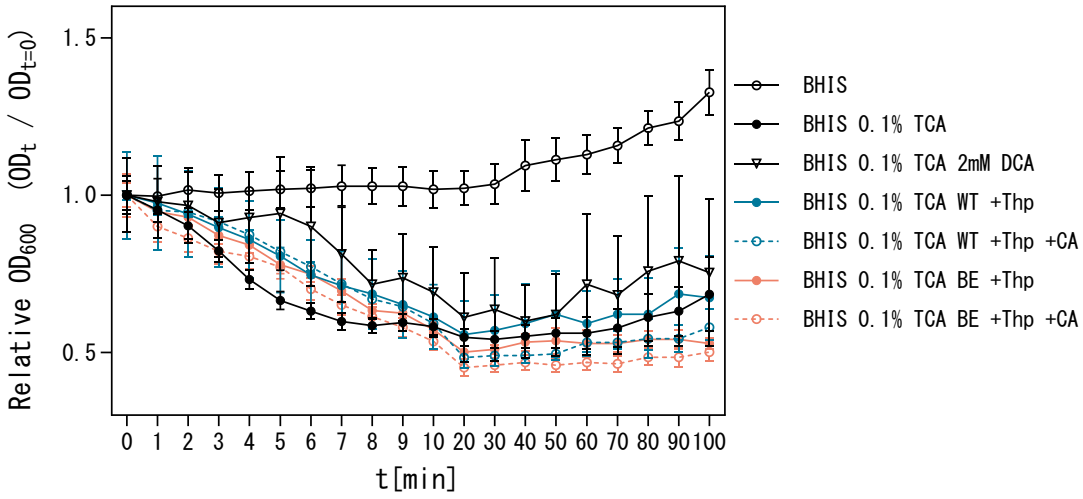


Figure 5.18 Germination initiation of *C. difficile* R20291. *C. difficile* spores were incubated with prepared supernatants from *C. butyricum* WT and BE4 strains incubated with theophylline (5 mM), and with and without cholic acid (CA 0.5 mM). The germinant TCA (0.1%) was also incubated with all spores. Growth was measured by OD_{600} in a plate reader to capture early decreases. Data represent mean values of spores incubated with three independent supernatant preparations, each in technical triplicate ± SD.

Overall the supernatant assays do not conclusively show whether the *C. butyricum* BE4 strain can 7α-dehydroxylate CA and inhibit *C. difficile* germination. Developing the spore outgrowth assay with *C. butyricum* WT, as discussed, would improve insight into this and future work could carry out co-cultures of the BE4 strain with *C. difficile* spores and/or vegetative cells. However, further competition assays should only be carried out after confirmation of 7α-dehydroxylation in this strain by mass spectrometry; this is the biggest limitation to the interpretation of these results. If confirmed, the *C. butyricum bai* essential strains could subsequently improve the understanding of the 7α-dehydroxylation pathway, and assessment and optimisation of *C. difficile* inhibition in the BE4 strain could continue.

5.5 Key Outcomes

The *C. butyricum* strains developed in Chapter 4 were harnessed for insertion and expression of the *bai* genes of *C. scindens* in the hope of generating a model of 7α-dehydroxylation. Initially it was attempted to insert all 8 genes of the major *bai* operon

into the *C. butyricum* genome, utilising both individual and triple auxotrophic strains to attempt various strategies of insertion, using both ACE and RiboCas technologies. Despite multiple attempts this was unfortunately unsuccessful, likely due to toxicity of individual *bai* genes, and efforts were therefore focused on the cloning and insertion of the *bai* genes proposed to be essential for 7α-dehydroxylation in *C. scindens*. Cloning attempts with various promoter and *bai* gene combinations resulted in a full suite of ACE vectors for insertion of the essential genes in a step wise manner across the triple knockout strain: *baiG, baiB, baiCD, baiE* and *baiJ*. The transporter *baiG* strain was assessed for CA transport into the cell, but unfortunately the results were inconclusive, and the assay requires further development. RT-PCR confirmed expression of all inserted essential *bai* genes, but assessment of bile acid metabolism and 7α-dehydroxylation is awaiting mass spectrometry data. The *bai* essential strain underwent initial assessment for inhibition of *C. difficile* spore germination, by incubation with CA and supernatant challenge of spores. The assay suggested that there was not a bile-acid dependent effect on germination, however further characterisation is required before it can be concluded whether this is due to conditions of the assay, or absence of 7α-dehydroxylation.

Chapter 6 General Discussion

6.1 P. hiranonis

6.1.1 Genetic tools

P. hiranonis as a 7α-dehydroxylating species of interest has previously been studied with regards to its bile acid metabolism, however there is no published work developing genetic tools or mutant strains. Whilst this study did achieve DNA transfer, subsequent attempts to apply existing clostridial tools to engineering in *P. hiranonis* were not successful, and this is likely explained by non-functioning promoters. However, this was surprising given the apparently universal use of these tools in other clostridia. The reclassification of *Clostridium hiranonis* to *Peptacetobacter hiranonis* in 2020 (Chen et al., 2020), partway through this report, could have suggested that the lineage of *P. hiranonis* was too divergent for an overlap of functional tools. However, *P. hiranonis* was reclassified from the family Clostridiaceae to Peptostreptococcaceae, the same family as *C. difficile*, a species in which RiboCas has been shown to function (Cañadas et al., 2019). Moreover, this report has demonstrated that RiboCas does function in Clostridiaceae as it was used successfully in *C. butyricum*. This suggests that, based on its classification, one would expect RiboCas to function in *P. hiranonis*, and therefore there may be a species-specific factor that is responsible for the inability to use universal clostridial tools.

The use of native *P. hiranonis* promoters in the development of genetic tools did not prove to be the solution, further suggesting that there is an unknown mechanism preventing engineering of the genome. However, whilst unsuccessful in this instance, identification of native promoters and the development of inducible promoters for *P. hiranonis* does offer a small toolkit for plasmid-based expression systems if required in the future. Moreover, the confirmation of the bile acid inducibility of the *barAB*/*baiB* promoter region also expands the repertoire of inducible promoters available, although only for non-bile acid or gut associated studies. Future work could investigate the use of this system in the engineering or gene expression of other organisms, with particular interest to its functionality in other clostridia.

6.1.2 Investigating colonisation resistance

The difficulty in studying *P. hiranonis* with regards to its bile acid-mediated role in CDI resistance is that it is less characterised than *C. scindens*, which as a single species had been directly implicated. With less data characterising *P. hiranonis* wildtype resistance, in *in vitro* and animal models, this makes it more difficult to study the impact of removal of 7α-dehydroxylating activity in the *ΔbaiCD* strain generated in Chapter 3. For optimal insights to be gained from the *ΔbaiCD* strain, future work should focus on developing a suite of *in vitro* assays and data to demonstrate bile acid-mediated mechanisms of *C. difficile* inhibition by *P. hiranonis*.

The spore outgrowth assay in Chapter 3 did allow such a comparison, as it demonstrated inhibition of *C. difficile* spore germination by the supernatant of the *P. hiranonis* WT when incubated with CA, but not in the absence of CA. Moreover, this effect was eliminated in the *ΔbaiCD* strain demonstrating the requirement for 7α-dehydroxylation. However, as the WT supernatant only delayed the onset of germination and did not abolish it, resulting in final spore outgrowth no different from the positive control, this suggests that *C. difficile* colonisation would not be prevented. This demonstrates the importance of the choice of the indicator of *C. difficile* inhibition to be measured, and should be considered in future assays.

A further consideration for 7α-dehydroxylation assays is the concentration of CA chosen; a balance is required between maximising the production of observable bile acid-mediated effects, and the maintenance of physiologically relevant concentrations. For example, whilst the *P. hiranonis* wildtype when incubated with 1 mM CA demonstrated a DCA-mediated inhibition of *C. difficile* germination, the chosen CA concentration may have resulted in a rate of 7α-dehydroxylation that is not reflective of that in the colon. Unfortunately, the colonic concentration of cholic acid is unclear in the literature, but it has been suggested that 1 mM is the maximum concentration reached, after a high fat meal (Zeng, Umar, Rust, Lazarova, & Bordonaro, 2019). Moreover, previous work in the better characterised *C. scindens* does not reflect a universal CA concentration for incubation, with examples including 1 µM (Funabashi et al., 2020), 100 µM (Marion et al., 2019) and 2.5 mM (Reed et al., 2020). This highlights the importance of expanding the germination assay, and future assays, to study incubations of *P. hiranonis* with a range of CA concentrations, to ensure any

CDI resistance mechanisms observed are not a result of artificial 7α-dehydroxylation levels. Quantification of the DCA concentration present in the supernatant would also assist with this. If varying CA concentrations are studied, the sensitivity and lower limits of quantification for the chosen *in vitro* assays would need to be considered.

The main objective of the *ΔbaiCD* strain would be to study the bile acid-mediated colonisation resistance against *C. difficile* in a mouse model compared to the WT, however studies of *P. hiranonis* in a mouse model are limited. *P. hiranonis* DSM 13275 has been shown to colonise mice, however no DCA, and therefore 7α-dehydroxylation, was observed in the mono-associated gnotobiotic mice (Narushima, Itoh, Takamine, & Uchida, 1999b). Unfortunately, there have been no recent studies to confirm this finding. The colonisation of *P. hiranonis* DSM 13275 with other bile acid-deconjugating bacteria has been shown to produce DCA (Narushima, Itoh, Takamine, & Uchida, 1999a), and as part of a consortium (Ridlon et al., 2020), and this could allow a comparison to the *ΔbaiCD* strain in future studies. The *P. hiranonis* JCM 10542 strain has been shown to result in DCA production and protection against CDI on mono-association of germ-free mice (Aguirre et al., 2021), so could be a strain of focus for future genetic tool development. Although not optimal, given the possible difficulties in generating further gene knockouts, the dog model could be considered to study the *P. hiranonis* DSM 13275 *ΔbaiCD* strain as colonisation of dogs by *P. hiranonis* DSM 13275 has been demonstrated, and is associated with CDI protection (Thanissery et al., 2020).

Whilst it may not be possible to study CDI protection by the *ΔbaiCD* strain in an animal model, it will still be beneficial in studying 7α-dehydroxylation *in vitro*, as evidenced in this dissertation. Had the development of genetic tools for *P. hiranonis* been successful, further *bai* gene knockouts could have been generated to elucidate enzyme activities in the 7α-dehydroxylation pathway by studying bile acid intermediates of these strains. In contrast the *ΔbaiCD* strain simply allows a binary on/off study of 7α-dehydroxylation which, whilst not as enlightening, may be beneficial as a control species in studies of 7α-dehydroxylating bacteria in general.

6.2 C. butyricum

6.2.1 Genetic manipulation and its use as a delivery tool

C. butyricum proffers many opportunities for exploitation due to the variety of niches it inhabits, no more so than its existence as gut commensal; the promising development of probiotic strains could open many doors to its utilisation. However, there is limited work in the genetic manipulation of the bacterium due to a lack of tools available. This dissertation has developed strains of *C. butyricum* 7423 primed for insertion of application specific modules through the generation of auxotrophic mutants using RiboCas. It has demonstrated successful use of RiboCas in the strain, generating single, double and triple knockout strains, which not only demonstrate auxotrophy at *purD* and *hisI*, but can also be utilised for cargo insertion via ACE.

The first successful use of RiboCas in *C. butyricum* is also only the second example of the use of CRISPR technology in *C. butyricum*. Until recently only three mutations had been generated in different strains of *C. butyricum*: *hbd* and *adhE* in *C. butyricum* W5 to enhance hydrogen production (Cai, Jin, Monis, & Saint, 2013; Cai, Jin, Saint, & Monis, 2011), and *dltD* in *C. butyricum* CB1002 involved in cell septation (Wydau-Dematteis et al., 2015). In all instances a low-throughput intron-based mutagenesis was used. The use of a lactose-inducible heterologous Type II CRISPR-Cas9 system was attempted by Zhou et al. (2021) in probiotic strains for deletion of *spo0A* and *adhE* which was successful. However, the authors argued that the low conjugation efficiency and inconsistent editing efficiency prevented its further use, and proceeded to develop genome editing using the endogenous Type 1-B CRISPR-Cas (Zhou et al., 2021). In contrast, use of the RiboCas system with the p19 replicon in this study allowed consistent conjugation into *C. butyricum* NCTC 7423 and resulting in high, consistent editing efficiencies (80-100%), for both deletions and insertions. This demonstrates RiboCas as a reliable genome editing system for use in *C. butyricum*, and avoids the need to utilise an endogenous system, particularly as one was not identified in NCTC 7423.

The successful use of ACE for insertion and expression of *bai* genes into the triple knockout *C. butyricum* strain demonstrates its utility as a vehicle for heterologous gene expression. Whilst insertion by ACE does require more experimental steps than

using RiboCas, the protocol has now been established and allows insertions of larger cargo, widening the range of possible applications for the bacteria. Whilst not all objectives could be achieved with regards to *bai* gene insertion, the demonstration of *bai* gene expression using the clostridial promoters characterised in this strain outlines the proof of concept for stable expression of genes. The *C. butyricum* auxotrophic strains could therefore provide opportunities for expression of other genes in the future and would allow targeting to the gut. For example, plasmid-based systems in a *C. butyricum* probiotic have been engineered to express glucagon-like peptide-1 (GLP-1), a peptide-based hormone, and demonstrated therapeutic effects such as improving motor dysfunction in a Parkinson's disease mouse model (Wang et al., 2023), and blood pressure in hypertensive rats (X. liang Wang et al., 2023). Applying these results to a stable genome-based expression system would reduce the risk of decrease in proportion of plasmid-containing cells during replication (Millan et al., 2014), and of horizontal gene transfer (Keese, 2008), improving its outlook as a therapeutic for use in humans.

Stable recombinant strains of other spore forming bacteria have previously been explored as drug delivery vehicles. For example, *C. sporogenes* has been engineered to express a novel nitroreductase for metabolism of a pro-drug that has been shown to have anti-cancer activity in preclinical evaluations (Heap et al., 2014; Mowday et al., 2022), administered as spores to survive anaerobic environments. It is hoped that this principle can be applied to harnessing the existing probiotic uses of *C. butyricum* (Hagihara et al., 2020) for delivery to the gut, for example in the MIYAIRI 588 strain.

6.2.2 For 7α-dehydroxylation

Unfortunately, it was not possible to transfer all major *bai* genes of the *C. scindens bai* operon into the *C. butyricum* model, despite several strategies being attempted. This was largely impeded by cloning toxicities and could not be overcome by use of different promoters or *bai* gene combinations. With hindsight perhaps it would have been pertinent to undertake the same strategy that was successful for Funabashi et al. (2020), utilising their three *bai* operon sections across the three *C. butyricum* loci. However, this still would not have guaranteed success. Firstly, their choice of promoters for expression in *C. sporogenes* may not have been transferable for use in *C. butyricum*, and changes in *bai* gene expression levels following replacement with

functionally confirmed promoters may have altered toxicity. Secondly, whilst adequate for expression on a plasmid, their largest *bai* operon section (*baiB-baiF*) would have likely yielded a complementation vector too large for conjugation, and is unlikely to have inserted into the genome given that insertion of the smaller fragment of *baiB-A2* was unsuccessful in this study. Future work for insertion of the *C. scindens bai* operon into *C. butyricum* could utilise the optimisation of RiboCas and ACE in this report, and insights gained from *bai* gene cloning and insertion in this report, alongside the strategy of Funabashi et al. (2020). This could entail using RiboCas or ACE for insertion of the two smaller *bai* operon fragments at two loci, along with the use of ACE/RiboCas for insertion of half of the larger fragment with a bookmark and insertion of the second half with RiboCas, to overcome size issues. Assessing these fragments alongside the range of promoters from this study may allow both cloning and insertion of the whole operon, something this study was unable to achieve.

Whilst this study was unable to mobilise the complete 7α-dehydroxylation pathway into *C. butyricum*, the putatively essential genes were inserted and confirmed to express. Whilst initial *in vitro* assays could not conclude whether this strain is capable of producing DCA, bile acid analysis is ongoing to assess if bile acid metabolism is imparted to this strain. In addition to its native protection against CDI (Hagihara et al., 2021, 2020, 2018; Hayashi et al., 2021), the use of *C. butyricum* as the chassis in this instance also imparts additional opportunities due to the presence of a choloylglycine hydrolase gene (FF104_09195). Choloylglycine hydrolases (CGH) function as a bile salt hydrolase, catalysing the hydrolysis of conjugated bile acids to produce the free primary bile acid, essential for 7α-dehydroxylation to then occur (Marchesini et al., 2011). If the engineered *C. butyricum* strain is capable of 7α-dehydroxylating CA, the presence of this gene could be extremely advantageous, extended above even native *C. scindens* which was shown to not deconjugate bile salts and instead to rely on the BSH activity of other bacteria such as *C. difficile* to generate free CA (Aguirre, Adegbite, & Sorg, 2022; Aguirre et al., 2021). Further work in *C. butyricum* should characterise the activity of the native CGH at bile acid deconjugation, in addition to confirming its role by generating a gene knockout using RiboCas.

Whilst the presence of a putative bile salt hydrolase is beneficial, the use of a chassis strain in general for *bai* gene expression does pose some limitations given that other

unknown genes required for 7α-dehydroxylation will be absent. Whilst it overcomes the genetic intractability of *C. scindens*, it may not be a complete replication of 7α-dehydroxylation. For example, in the model by Funabashi et al. (2020) the level of DCA production was significantly lower than the native level, and the authors suggest this could be due to the absence of ancillary genes required for metabolism, for example co-factor biogenesis. Until further characterisation of the pathway and gene identification has taken place this would be difficult to overcome, but identified genes could also be mobilised to *C. butyricum* if required. In a similar manner, use of a *C. butyricum* chassis would also remove the other proposed mechanism of *C. scindens* inhibition of CDI, which is the production of a tryptophan derived antimicrobial (J. D. Kang et al., 2019b). Whilst this may result in lower inhibition from a 7α-dehydroxylation chassis strain compared to native *C. scindens*, it could be highly beneficial in further elucidating the mechanisms of action of *C. scindens*; the chassis strain would allow the separation of non-bile acid mediated effects. Future work could also combine both proposed mechanisms in the chassis strain, but this would require the gene(s) responsible for production of this antimicrobial to be identified.

Finally, with regards to identifying *C. scindens* mechanisms of action the *C. butyricum* auxotrophic strain could also be used to investigate the hypotheses of protection by Stickland metabolism. Independently to 7α-dehydroxylation the genes for Stickland metabolism, for example proline or glycine reductases, could be transferred and expressed in *C. butyricum* using the same tools and principles as for the *bai* genes. Competition studies with *C. difficile* could be carried out to investigate inhibition, although the native inhibition by *C. butyricum* would have to be carefully considered to allow Stickland effects to be identified.

6.3 Summary and Future Perspectives

Presented in this dissertation are two strategies which, whilst approaching from different angles, could both aid in furthering the understanding of 7α-dehydroxylation, in both the pathway itself and its role in CDI resistance. The study of the native pathway in *P. hiranonis* was impeded by the inability to develop genetic tools to generate loss-of-function mutations, but did produce the *ΔbaiCD* knockout incapable of 7α-dehydroxylation. Moreover, the expression of *bai* genes in *C. butyricum*, whilst requiring further optimisation, could allow the study of bile acid intermediates as well

as the mobilisation of the 7α-dehydroxylation pathway to a chassis organism. Both strategies present the opportunity to study bile-acid mediated inhibition of *C. difficile* and could distinguish the effects of 7α-dehydroxylation from other potential mechanisms, such as production of antimicrobials or Stickland metabolism.

The bile acid metabolism field in general has experienced developments since the onset of this dissertation. The major development is the elucidation of the enzymes involved in the reductive arm of the 7α-dehydroxylation pathway by Funabashi et al. (2020), producing a complete pathway. The characterisation of the major *bai* enzymes involved could allow an optimised use of the pathway in future work, and provide a clearer background from which to understand the role of the peripheral *bai* genes. Recent studies have also begun exploring the diversity and abundance of other possible 7α-dehydroxylation species in the human gut, identifying for example 21 new metagenomic species containing a *bai* gene cluster (Vital, Rud, Rath, Pieper, & Schlüter, 2019). This highlights a possible shift towards exploring organisms with *bai* genes not necessarily arranged in the classical *bai* operon structure, with a range of novel structures identified cross several genera (Kim et al., 2022), exemplified in the *baiH* mutant generated in the newly identified *Faecalicatena contorta* (Jin et al., 2022). Further work is required to determine 7α-dehydroxylation in these newly identified organisms, and is not guaranteed, but has already been shown to be successful in the cases of *Eubacterium* sp. c-25 (I. Song et al., 2021) and *Dorea* sp. AM58-8 (Bai, Zhao, Gao, Zou, & Lei, 2022). This perhaps offers a new direction for bile acid metabolism work, rather than constricting the field to the few classical clostridial species that have been the focus since the outset. This new direction could present opportunities to circumvent the ongoing challenge in these organisms which is genetic manipulation, as reiterated in this dissertation.

The long-term aim of the field will be to allow prevention or treatment of CDI through the use of 7α-dehydroxylating gut microbes, reducing the reliance on antibiotics. However, the field of probiotics is still limited and there are currently no clinical probiotics in use for CDI. There is evidence for the success of some probiotic therapies in mouse models of CDI (X. Li et al., 2019; Shelby et al., 2020) and very small-scale, limited clinical trials in the use of probiotics to treat CDI, a review of which has been written by Gaines & Moore (2019). The research into *C. scindens* as a probiotic is

limited to mice studies (Buffie et al., 2015). Overall comprehensive research into probiotic use for CDI prevention or treatment is lacking, and the long-term safety implications are not well understood. This is a challenge that will need to be overcome before 7α-dehydroxylating bacteria can offer a therapeutic benefit. Moreover, even if future evidence demonstrates a high efficacy of treatment with 7α-dehydroxylating bacteria, the likelihood of a single magic bullet therapy is low. Future therapies are likely to employ a microbial consortia due to the complex interactions of the gut microbiota, and this is already being explored in mice models (Battaglioli et al., 2018; Buffie et al., 2015).

Finally, whilst the focus of this dissertation was bile acid metabolism pertaining to its role in CDI prevention, both the strategies explored in this report, as well as developments in the bile acid metabolism field, also present the opportunity to expand knowledge and possible treatments in other bile-acid related illnesses. Bile acids have been implicated in a range of metabolic and inflammatory diseases, of which a perturbed gut microbiome has been linked to many. For example, specific bile acid metabolites have been shown to contribute to the inflammation observed in inflammatory bowel disease (Thomas, Modos, Rushbrook, Powell, & Korcsmaros, 2022) and disrupted bile acid metabolism has been demonstrated in liver diseases and colon cancer (Chen, Takeda, & Sundrud, 2019). This demonstrates the possible further applications of studying the role of bile acids in the gut microbiota. It also highlights the need to consider the possible wide-ranging consequences of any bile-acid related therapies.

Bibliography

Abt, M. C., McKenney, P. T., & Pamer, E. G. (2016, October 1). Clostridium difficile colitis: Pathogenesis and host defence. *Nature Reviews Microbiology*, Vol. 14, pp. 609–620. https://doi.org/10.1038/nrmicro.2016.108

Adams, C. M., Eckenroth, B. E., Putnam, E. E., Doublié, S., & Shen, A. (2013). Structural and Functional Analysis of the CspB Protease Required for Clostridium Spore Germination. *PLoS Pathogens*, *9*(2). https://doi.org/10.1371/journal.ppat.1003165

Aguirre, A. M., Adegbite, A. O., & Sorg, J. A. (2022). Clostridioides difficile bile salt hydrolase activity has substrate specificity and affects biofilm formation. *Npj Biofilms and Microbiomes 2022 8:1*, *8*(1), 1–10. https://doi.org/10.1038/s41522-022-00358-0

Aguirre, A. M., Yalcinkaya, N., Wu, Q., Swennes, A., Tessier, M. E., Roberts, P., … Sorg, J. A. (2021). Bile acid-independent protection against Clostridioides difficile infection. *PLoS Pathogens*, *17*(10). https://doi.org/10.1371/JOURNAL.PPAT.1010015

Amy, J., Bulach, D., Knight, D., Riley, T., Johanesen, P., & Lyras, D. (2018). Identification of large cryptic plasmids in Clostridioides (Clostridium) difficile. *Plasmid*, *96–97*(April), 25–38. https://doi.org/10.1016/j.plasmid.2018.04.001

Annual epidemiological commentary: Gram-negative, MRSA, MSSA bacteraemia and C. difficile infections, up to and including financial year 2021 to 2022. (2022). Retrieved from https://www.gov.uk/government/statistics/mrsa-mssa-and-e-coli-bacteraemia-and-c-difficile-infection-annual-epidemiological-commentary/annual-epidemiological-commentary-gram-negative-mrsa-mssa-bacteraemia-and-c-difficile-infections-up-to-and-including-fina

Antharam, V. C., Li, E. C., Ishmael, A., Sharma, A., Mai, V., Rand, K. H., & Wang, G. P. (2013). Intestinal dysbiosis and depletion of butyrogenic bacteria in Clostridium difficile infection and nosocomial diarrhea. *Journal of Clinical Microbiology*, *51*(9), 2884–2892. https://doi.org/10.1128/JCM.00845-13

Antonopoulos, D. A., Huse, S. M., Morrison, H. G., Schmidt, T. M., Sogin, M. L., & Young, V. B. (2009). Reproducible community dynamics of the gastrointestinal microbiota following antibiotic perturbation. *Infection and Immunity*, *77*(6), 2367–2375. https://doi.org/10.1128/IAI.01520-08

Arndt, D., Marcu, A., Liang, Y., & Wishart, D. S. (2018). PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes. *Briefings in Bioinformatics*, *20*(4), 1560–1567. https://doi.org/10.1093/bib/bbx121

Arrieta-Ortiz, M. L., Immanuel, S. R. C., Turkarslan, S., Wu, W. J., Girinathan, B. P., Worley, J. N., … Baliga, N. S. (2021). Predictive regulatory and metabolic network models for systems analysis of Clostridioides difficile. *Cell Host and Microbe*, *29*(11), 1709-1723.e5. https://doi.org/10.1016/j.chom.2021.09.008

Aureli, P., Fenicia, L., Pasolini, B., Gianfrancesehi, M., McCroskey, L. M., Hatheway, C. L., … Hatheway, C. L. (1986). Two cases of type E infant botulism caused by neurotoxigenic Clostridium butyricum in Italy. *The Journal of Infectious Diseases*, *154*(2), 207–211. https://doi.org/10.1093/INFDIS/154.2.207

Bai, Y., Zhao, T., Gao, M., Zou, Y., & Lei, X. (2022). A Novel Gene Alignment in Dorea sp. AM58-8 Produces 7-Dehydroxy-3β Bile Acids from Primary Bile Acids. *Biochemistry*, *61*(24), 2870–2878. https://doi.org/10.1021/acs.biochem.2c00264

Ball, D. A., Taylor, R., Todd, S. J., Redmond, C., Couture-Tosi, E., Sylvestre, P., … Bullough, P. A. (2008). Structure of the exosporium and sublayers of spores of the Bacillus cereus family revealed

by electron crystallography. *Molecular Microbiology*, *68*(4), 947–958. https://doi.org/10.1111/j.1365-2958.2008.06206.x

Barra-Carrasco, J., & Paredes-Sabja, D. (2014, May 8). Clostridium difficile spores: A major threat to the hospital environment. *Future Microbiology*, Vol. 9, pp. 475–486. https://doi.org/10.2217/fmb.14.2

Bartlett, J. G., Chang, te W., Gurwith, M., Gorbach, S. L., & Onderdonk, A. B. (1978). Antibiotic-Associated Pseudomembranous Colitis Due to Toxin-Producing Clostridia. *New England Journal of Medicine*, *298*(10), 531–534. https://doi.org/10.1056/NEJM197803092981003

Bartlett JG. (1994). Clostridium difficile: history of its role as an enteric pathogen and the current state of knowledge about the organism. *Clinical Infectious Diseases*, *18*(4), 265–272.

Battaglioli, E. J., Hale, V. L., Chen, J., Jeraldo, P., Ruiz-Mojica, C., Schmidt, B. A., … Kashyap, P. C. (2018). Clostridioides difficile uses amino acids associated with gut microbial dysbiosis in a subset of patients with diarrhea. *Science Translational Medicine*, *10*(464). https://doi.org/10.1126/scitranslmed.aam7019

Bauer, M. P., Notermans, D. W., Van Benthem, B. H., Brazier, J. S., Wilcox, M. H., Rupnik, M., … Kuijper, E. J. (2011). Clostridium difficile infection in Europe: A hospital-based survey. *The Lancet*, *377*(9759), 63–73. https://doi.org/10.1016/S0140-6736(10)61266-4

Beer, L. A., Tatge, H., Schneider, C., Ruschig, M., Hust, M., Barton, J., … Gerhard, R. (2018). The binary toxin CDT of Clostridium difficile as a tool for intracellular delivery of bacterial glucosyltransferase domains. *Toxins*, *10*(6). https://doi.org/10.3390/toxins10060225

Bhattacharjee, D., McAllister, K. N., & Sorg, J. A. (2016). Germinants and their receptors in clostridia. *Journal of Bacteriology*, *198*(20), 2767–2775. https://doi.org/10.1128/JB.00405-16

Bhowmik, S., Chiu, H.-P., Jones, D. H., Chiu, H.-J., Miller, M. D., Xu, Q., … Lesley, S. A. (2016). Structure and Functional Characterization of a Bile Acid 7α Dehydratase BaiE in Secondary Bile Acid Synthesis HHS Public Access. *Proteins*, *84*(3), 316–331. https://doi.org/10.1002/prot.24971

Bhowmik, S., Jones, D. H., Chiu, H. P., Park, I. H., Chiu, H. J., Axelrod, H. L., … Lesley, S. A. (2014). Structural and functional characterization of BaiA, an enzyme involved in secondary bile acid synthesis in human gut microbe. *Proteins: Structure, Function and Bioinformatics*, *82*(2), 216–229. https://doi.org/10.1002/prot.24353

Bickle, T. A., & Kruger, D. H. (1993). *Biology of DNA Restriction*. *57*(2), 434–450.

Breaker, R. R. (2018). Riboswitches and translation control. *Cold Spring Harbor Perspectives in Biology*, *10*(11), 1–14. https://doi.org/10.1101/cshperspect.a032797

Buffie, C. G., Bucci, V., Stein, R. R., McKenney, P. T., Ling, L., Gobourne, A., … Pamer, E. G. (2015). Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature*, *517*(7533), 205–208. https://doi.org/10.1038/nature13828

Buffie, C. G., Jarchum, I., Equinda, M., Lipuma, L., Gobourne, A., Viale, A., … Pamer, E. G. (2012). Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to Clostridium difficile-induced colitis. *Infection and Immunity*, *80*(1), 62–73. https://doi.org/10.1128/IAI.05496-11

Burns, D. A., Heap, J. T., & Minton, N. P. (2010). Clostridium difficile spore germination: An update. *Research in Microbiology*, *161*(9), 730–734. https://doi.org/10.1016/j.resmic.2010.09.007

Burns, D. A., Heeg, D., Cartman, S. T., & Minton, N. P. (2011). Reconsidering the sporulation characteristics of hypervirulent clostridium difficile BI/NAP1/027. *PLoS ONE*, *6*(9). https://doi.org/10.1371/journal.pone.0024894

Cai, G., Jin, B., Monis, P., & Saint, C. (2013). A genetic and Metabolic Approach to Redirection of Biochemical Pathways of Clostridium butyricum for Enhancing Hydrogen Production. *Biotechnol. Bioeng*, *110*, 338–342. https://doi.org/10.1002/bit.24596/abstract

Cai, G., Jin, B., Saint, C., & Monis, P. (2011). Genetic manipulation of butyrate formation pathways in Clostridium butyricum. *Journal of Biotechnology*, *155*, 269–274. https://doi.org/10.1016/j.jbiotec.2011.07.004

Cairns, M. D., Preston, M. D., Hall, C. L., Gerding, D. N., Hawkey, P. M., Kato, H., … Wren, B. W. (2017). Comparative genome analysis and global phylogeny of the toxin variant clostridium difficile PCR Ribotype 017 Reveals the evolution of two independent sublineages. *Journal of Clinical Microbiology*, *55*(3), 865–876. https://doi.org/10.1128/JCM.01296-16

Cañadas, I. C., Groothuis, D., Zygouropoulou, M., Rodrigues, R., & Minton, N. P. (2019). RiboCas: A Universal CRISPR-Based Editing Tool for Clostridium. *ACS Synthetic Biology*, *8*(6), 1379–1390. https://doi.org/10.1021/acssynbio.9b00075

Carroll, A. C., & Wong, A. (2018). Plasmid persistence: costs, benefits, and the plasmid paradox. *Canadian Journal of Microbiology*, *64*(5), 293–304. https://doi.org/10.1139/cjm-2017-0609

Cassiano, M. H. A., & Silva-Rocha, R. (2020). Benchmarking Bacterial Promoter Prediction Tools: Potentialities and Limitations. *MSystems*, *5*(4). https://doi.org/10.1128/MSYSTEMS.00439-20/FORMAT/EPUB

Cassir, N., Benamar, S., & La Scola, B. (2016). Clostridium butyricum: From beneficial to a new emerging pathogen. *Clinical Microbiology and Infection*, Vol. 22, pp. 37–45. https://doi.org/10.1016/j.cmi.2015.10.014

CDC. (2019). *Antibiotic resistance threats in the United States, 2019. Atlanta, GA: U.S. Department of Health and Human Services, CDC*. https://doi.org/CS239559-B

Chen, F., Cocaign-Bousquet, M., Girbal, L., & Nouaille, S. (2022). 5'UTR sequences influence protein levels in Escherichia coli by regulating translation initiation and mRNA stability. *Frontiers in Microbiology*, *13*(December), 1–10. https://doi.org/10.3389/fmicb.2022.1088941

Chen, M. L., Takeda, K., & Sundrud, M. S. (2019). Emerging roles of bile acids in mucosal immunity and inflammation. *Mucosal Immunology*, *12*(4), 851–861. https://doi.org/10.1038/s41385-019-0162-4

Chen, X. J., Wang, Z. Q., Zhou, Z. Y., Zeng, N. Y., Huang, Q. F., Wang, Z. W., … Zhou, H. W. (2020). Characterization of peptacetobacter hominis gen. Nov., sp. nov., isolated from human faeces, and proposal for the reclassification of clostridium hiranonis within the genus peptacetobacter. *International Journal of Systematic and Evolutionary Microbiology*, *70*(5), 2988–2997. https://doi.org/10.1099/IJSEM.0.003925/CITE/REFWORKS

Coleman, J. P., White, W. B., Egestad, B., Sjövall, J., & Hylemon, P. B. (1987). Biosynthesis of a novel bile acid nucleotide and mechanism of 7α-dehydroxylation by an intestinal Eubacterium species. *Journal of Biological Chemistry*, *262*(10), 4701–4707. https://doi.org/10.1016/s0021-9258(18)61252-9

Cooper, C. C., Jump, R. L. P., & Chopra, T. (2016, December 1). Prevention of Infection Due to Clostridium difficile. *Infectious Disease Clinics of North America*, Vol. 30, pp. 999–1012. https://doi.org/10.1016/j.idc.2016.07.005

Cornely, O. A., Miller, M. A., Louie, T. J., Crook, D. W., & Gorbach, S. L. (2012). *Treatment of First Recurrence of Clostridium difficile Infection: Fidaxomicin Versus Vancomycin*. https://doi.org/10.1093/cid/cis462

Crobach, M. J. T., Vernon, J. J., Loo, V. G., Kong, L. Y., Péchiné, S., Wilcox, M. H., & Kuijper, E. J. (2018, April 1). Understanding clostridium difficile colonization. *Clinical Microbiology Reviews*,

Vol. 31. https://doi.org/10.1128/CMR.00021-17

Crogan, N. L., & Evans, B. C. (2007). Clostridium difficile: an emerging epidemic in nursing homes. *Geriatric Nursing (New York, N.Y.)*, Vol. 28, pp. 161–164. https://doi.org/10.1016/j.gerinurse.2007.04.005

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Research*, *14*(6), 1188–1190. https://doi.org/10.1101/gr.849004

Czepiel, J., Dróżdż, M., Pituch, H., Kuijper, E. J., Perucki, W., Mielimonka, A., … Biesiada, G. (2019). Clostridium difficile infection: review. *European Journal of Clinical Microbiology and Infectious Diseases*. https://doi.org/10.1007/s10096-019-03539-6

Dawson, J. A., Mallonee, D. H., Bjorkhem, I., & B, P. (1996). Expression and characterization of a C24 bile acid 7 alpha-dehydratase from Eubacterium sp. strain VPI 12708 in Escherichia coli. *Journal Of Lipid Research*, *37*, 1258–1267.

Deakin, L. J., Clare, S., Fagan, R. P., Dawson, L. F., Pickard, D. J., West, M. R., … Lawley, T. D. (2012). The Clostridium difficile spo0A gene is a persistence and transmission factor. *Infection and Immunity*, *80*(8), 2704–2711. https://doi.org/10.1128/IAI.00147-12

Dembek, M., Stabler, R. A., Witney, A. A., Wren, B. W., & Fairweather, N. F. (2013). Transcriptional Analysis of Temporal Gene Expression in Germinating Clostridium difficile 630 Endospores. *PLoS ONE*, *8*(5). https://doi.org/10.1371/journal.pone.0064011

Dempster, A. W. (2017). *The Antimicrobial and Bile Acid Mediated Control of Clostridium difficile Infection*. PhD Thesis (University of Nottingham).

Devendran, S., Shrestha, R., Alves, J. M. P., Wolf, P. G., Ly, L., Hernandez, A. G., … Ridlon, J. M. (2019). Clostridium scindens ATCC 35704: Integration of nutritional requirements, the complete genome sequence, and global transcriptional responses to bile acids. *Applied and Environmental Microbiology*, *85*(7). https://doi.org/10.1128/AEM.00052-19

Di Bella, S., Ascenzi, P., Siarakas, S., Petrosillo, N., & di Masi, A. (2016, May 1). Clostridium difficile toxins A and B: Insights into pathogenic properties and extraintestinal effects. *Toxins*, Vol. 8. https://doi.org/10.3390/toxins8050134

Di Gregorio, M. C., Cautela, J., & Galantini, L. (2021). Physiology and Physical Chemistry of Bile Acids. *International Journal of Molecular Sciences 2021, Vol. 22, Page 1780*, *22*(4), 1780. https://doi.org/10.3390/IJMS22041780

Doerner, K. C., Takamine, † Fusae, Lavoie, C. P., Mallonee, D. H., & Hylemon, P. B. (1997). Assessment of Fecal Bacteria with Bile Acid 7-Dehydroxylating Activity for the Presence of bai-Like Genes. In *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* (Vol. 63).

Dong, H., Zhang, Y., Dai, Z., & Li, Y. (2010). Engineering Clostridium strain to accept unmethylated DNA. *PLoS ONE*, *5*(2), 1–8. https://doi.org/10.1371/journal.pone.0009038

Dupuy, B., & Matamouros, S. (2006). Regulation of toxin and bacteriocin synthesis in Clostridium species by a new subgroup of RNA polymerase σ-factors. *Research in Microbiology*, *157*, 201–205. https://doi.org/10.1016/j.resmic.2005.11.004

Dykes, J. K., Lúquez, C., Raphael, B. H., McCroskey, L., & Maslanka, S. E. (2015). Laboratory Investigation of the First Case of Botulism Caused by Clostridium butyricum Type E Toxin in the United States. *Journal of Clinical Microbiology*, *53*(10), 3363. https://doi.org/10.1128/JCM.01351-15

Edwards, A. N., & Mcbride, S. M. (2014). Initiation of sporulation in Clostridium difficile: A twist on the classic model. *FEMS Microbiology Letters*, Vol. 358, pp. 110–118. https://doi.org/10.1111/1574-6968.12499

Egan, S. M. (2002). Growing Repertoire of AraC / XylS Activators GUEST COMMENTARY Growing Repertoire of AraC / XylS Activators. *Journal of Bacteriology*, *184*(20), 5529–5532. https://doi.org/10.1128/JB.184.20.5529

Elliott, B., Androga, G. O., Knight, D. R., & Riley, T. V. (2017, April 1). Clostridium difficile infection: Evolution, phylogeny and molecular epidemiology. *Infection, Genetics and Evolution*, Vol. 49, pp. 1–11. https://doi.org/10.1016/j.meegid.2016.12.018

Enright, E. F., Griffin, B. T., Gahan, C. G. M., & Joyce, S. A. (2018). Microbiome-mediated bile acid modification: Role in intestinal drug absorption and metabolism. *Pharmacological Research*, *133*, 170–186. https://doi.org/10.1016/j.phrs.2018.04.009

Evans, C. T., & Safdar, N. (2015). Current trends in the epidemiology and outcomes of clostridium difficile infection. *Clinical Infectious Diseases*, *60*, S66–S71. https://doi.org/10.1093/cid/civ140

Evfratov, S. A., Osterman, I. A., Komarova, E. S., Pogorelskaya, A. M., Rubtsova, M. P., Zatsepin, T. S., … Dontsova, O. A. (2017). Application of sorting and next generation sequencing to study 5'-UTR influence on translation efficiency in Escherichia coli. *Nucleic Acids Research*, *45*(6), 3487–3502. https://doi.org/10.1093/nar/gkw1141

Fagan, R. P., & Fairweather, N. F. (2011). Clostridium difficile has two parallel and essential sec secretion systems. *Journal of Biological Chemistry*, *286*(31), 27483–27493. https://doi.org/10.1074/jbc.M111.263889

Fimlaid, K. A., Jensen, O., Donnelly, M. L., Francis, M. B., Sorg, J. A., & Shen, A. (2015). Identification of a Novel Lipoprotein Regulator of Clostridium difficile Spore Germination. *PLoS Pathogens*, *11*(10). https://doi.org/10.1371/journal.ppat.1005239

Fimlaid, K. A., & Shen, A. (2015). Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes. *Current Opinion in Microbiology*, Vol. 24, pp. 88–95. https://doi.org/10.1016/j.mib.2015.01.006

Francis, M. B., Allen, C. A., Shrestha, R., & Sorg, J. A. (2013). Bile Acid Recognition by the Clostridium difficile Germinant Receptor, CspC, Is Important for Establishing Infection. *PLoS Pathogens*, *9*(5). https://doi.org/10.1371/journal.ppat.1003356

Francis, M. B., Allen, C. A., & Sorg, J. A. (2015). Spore cortex hydrolysis precedes dipicolinic acid release during Clostridium difficile spore germination. *Journal of Bacteriology*, *197*(14), 2276–2283. https://doi.org/10.1128/JB.02575-14

Funabashi, M., Grove, T. L., Wang, M., Varma, Y., McFadden, M. E., Brown, L. C., … Fischbach, M. A. (2020). A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature*, *582*(7813), 566–570. https://doi.org/10.1038/s41586-020-2396-4

Furuya-Kanamori, L., Marquess, J., Yakob, L., Riley, T. V., Paterson, D. L., Foster, N. F., … Clements, A. C. A. (2015). Asymptomatic Clostridium difficile colonization: epidemiology and clinical implications. *BMC Infectious Diseases*, *15*(1), 516. https://doi.org/10.1186/s12879-015-1258-4

Gaines, C., & Moore, J. (2019). Probiotics for the prevention of clostridium difficile-Associated diarrhea in adults and children. *Gastroenterology Nursing*, *42*(3), 299–301. https://doi.org/10.1097/SGA.0000000000000469

Gallegos, M. T., Schleif, R., Bairoch, A., Hofmann, K., & Ramos, J. L. (1997). Arac/XylS family of transcriptional regulators. *Microbiology and Molecular Biology Reviews*, *61*(4), 393–410. https://doi.org/10.1128/mmbr.61.4.393-410.1997

Garrett, E. M., Sekulovic, O., Wetzel, D., Jones, J. B., Edwards, A. N., Vargas-Cuebas, G., … Tamayo, R. (2019). Phase variation of a signal transduction system controls Clostridioides difficile colony morphology, motility, and virulence. *PLoS Biology*, *17*(10), 1–28. https://doi.org/10.1371/journal.pbio.3000379

Genth, H., Dreger, S. C., Huelsenbeck, J., & Just, I. (2008). Clostridium difficile toxins: More than mere inhibitors of Rho proteins. *International Journal of Biochemistry and Cell Biology*. https://doi.org/10.1016/j.biocel.2007.12.014

Ghoddusi, H. B., & Sherburn, R. (2010). Preliminary study on the isolation of Clostridium butyricum strains from natural sources in the UK and screening the isolates for presence of the type E botulinal toxin gene. *International Journal of Food Microbiology*, *142*(1–2), 202–206. https://doi.org/10.1016/J.IJFOODMICRO.2010.06.028

Giaretta, P. R., Rech, R. R., Guard, B. C., Blake, A. B., Blick, A. K., Steiner, J. M., … Suchodolski, J. S. (2018). Comparison of intestinal expression of the apical sodium-dependent bile acid transporter between dogs with and without chronic inflammatory enteropathy. *Journal of Veterinary Internal Medicine*, *32*(6), 1918–1926. https://doi.org/10.1111/jvim.15332

Giel, J. L., Sorg, J. A., Sonenshein, A. L., & Zhu, J. (2010). Metabolism of bile salts in mice influences spore germination in clostridium difficile. *PLoS ONE*, *5*(1). https://doi.org/10.1371/journal.pone.0008740

Girinathan, B. P., DiBenedetto, N., Worley, J. N., Peltier, J., Arrieta-Ortiz, M. L., Immanuel, S. R. C., … Bry, L. (2021). In vivo commensal control of Clostridioides difficile virulence. *Cell Host and Microbe*, *29*(11), 1693-1708.e7. https://doi.org/10.1016/j.chom.2021.09.007

Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., … Sorek, R. (2015). BREX is a novel phage resistance system widespread in microbial genomes. *The EMBO Journal*, *34*(2), 169–183. https://doi.org/10.15252/embj.201489455

Gordeeva, J., Morozova, N., Sierro, N., Isaev, A., Sinkunas, T., Tsvetkova, K., … Severinov, K. (2019). BREX system of Escherichia coli distinguishes self from non-self by methylation of a specific DNA site. *Nucleic Acids Research*, *47*(1), 253–265. https://doi.org/10.1093/nar/gky1125

Gu, S., Chen, Y., Zhang, X., Lu, H., Lv, T., Shen, P., … Li, L. (2015). *Identification of key taxa that favor intestinal colonization of Clostridium difficile in an adult Chinese population*. https://doi.org/10.1016/j.micinf.2015.09.008

Gupta, A., & Khanna, S. (2014, March 17). Community-acquired clostridium difficile infection: An increasing public health threat. *Infection and Drug Resistance*, Vol. 7, pp. 63–72. https://doi.org/10.2147/IDR.S46780

Gutelius, D., Hokeness, K., Logan, S. M., & Reid, C. W. (2014). Functional analysis of SleC from Clostridium difficile: An essential lytic transglycosylase involved in spore germination. *Microbiology (United Kingdom)*, *160*(PART 1), 209–216. https://doi.org/10.1099/mic.0.072454-0

Hafiz, S., & Oakley, C. L. (1976). Clostridium difficile isolation and characteristics. *Journal of Medical Microbiology*, *9*(2), 129–136. https://doi.org/10.1099/00222615-9-2-129

Hagihara, M., Ariyoshi, T., Kuroki, Y., Eguchi, S., Higashi, S., Mori, T., … Mikamo, H. (2021). Clostridium butyricum enhances colonization resistance against Clostridioides difficile by metabolic and immune modulation. *Scientific Reports*, *11*(1), 15007. https://doi.org/10.1038/S41598-021-94572-Z

Hagihara, M., Kuroki, Y., Ariyoshi, T., Yamagishi, Y., Takahashi, M., Mikamo, H., … Oka, K. (2020). *Clostridium butyricum Modulates the Microbiome to Protect Intestinal Barrier Function in Mice with Antibiotic-Induced Dysbiosis*. https://doi.org/10.1016/j.isci.2019.100772

Hagihara, M., Yamashita, R., Matsumoto, A., Mori, T., Kuroki, Y., Kudo, H., … Mikamo, H. (2018). *Anaerobes in the microbiome The impact of Clostridium butyricum MIYAIRI 588 on the murine gut microbiome and colonic tissue*. https://doi.org/10.1016/j.anaerobe.2018.07.012

Hall, I. C. (1935). INTESTINAL FLORA IN NEW-BORN INFANTS. *American Journal of Diseases*

of Children, *49*(2), 390. https://doi.org/10.1001/archpedi.1935.01970020105010

Harris, S. C., Devendran, S., Alves, J. M. P., Mythen, S. M., Hylemon, P. B., & Ridlon, J. M. (2018). Identification of a gene encoding a flavoprotein involved in bile acid metabolism by the human gut bacterium Clostridium scindens ATCC 35704. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, *1863*(3), 276–283. https://doi.org/10.1016/j.bbalip.2017.12.001

Hartman, A. H., Liu, H., & Melville, S. B. (2011). Construction and characterization of a lactose-inducible cromoter cystem for controlled gene expression in Clostridium perfringens. *Applied and Environmental Microbiology*, *77*(2), 471–478. https://doi.org/10.1128/AEM.01536-10

Hayashi, A., Nagao-Kitamoto, H., Kitamoto, S., Kim, C. H., & Kamada, N. (2021). The Butyrate-Producing Bacterium Clostridium butyricum Suppresses Clostridioides difficile Infection via Neutrophil- and Antimicrobial Cytokine–Dependent but GPR43/109a-Independent Mechanisms . *The Journal of Immunology*, *206*(7), 1576–1585. https://doi.org/10.4049/jimmunol.2000353

Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P., & Minton, N. P. (2007). The ClosTron: A universal gene knock-out system for the genus Clostridium. *Journal of Microbiological Methods*, *70*(3), 452–464. https://doi.org/10.1016/j.mimet.2007.05.021

Heap, J. T., Pennington, O. J., Cartman, S. T., & Minton, N. P. (2009). A modular system for Clostridium shuttle plasmids. *Journal of Microbiological Methods*. https://doi.org/10.1016/j.mimet.2009.05.004

Heap, J. T., Theys, J., Ehsaan, M., Kubiak, A. M., Dubois, L., Paesmans, K., … Minton, N. P. (2014). Spores of Clostridium engineered for clinical efficacy and safety cause regression and cure of tumors in vivo. *Oncotarget*, *5*(7), 1761–1769. https://doi.org/10.18632/oncotarget.1761

Heeg, D., Burns, D. A., Cartman, S. T., & Minton, N. P. (2012). Spores of clostridium difficile clinical isolates display a diverse germination response to bile salts. *PLoS ONE*, *7*(2). https://doi.org/10.1371/journal.pone.0032381

Heinlen, L., & Ballard, J. D. (2010). Clostridium difficile infection. *American Journal of the Medical Sciences*, *340*(3), 247–252. https://doi.org/10.1097/MAJ.0b013e3181e939d8

Henriques, A. O., & Moran, C. P. (2007). *Structure, Assembly, and Function of the Spore Surface Layers*. https://doi.org/10.1146/annurev.micro.61.080706.093224

Hidalgo-Cantabrana, C., Goh, Y. J., Pan, M., Sanozky-Dawes, R., & Barrangou, R. (2019). Genome editing using the endogenous type I CRISPR-Cas system in Lactobacillus crispatus. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(32), 15774–15783. https://doi.org/10.1073/PNAS.1905421116/SUPPL_FILE/PNAS.1905421116.SAPP.PDF

Houten, S. M., Watanabe, M., & Auwerx, J. (2006). New EMBO Member's Review Endocrine functions of bile acids. *The EMBO Journal*, *25*, 1419–1425. https://doi.org/10.1038/sj.emboj.7601049

Howerton, A., Patra, M., & Abel-Santos, E. (2013). A new strategy for the prevention of clostridium difficile infection. *Journal of Infectious Diseases*, *207*(10), 1498–1504. https://doi.org/10.1093/infdis/jit068

Howerton, A., Ramirez, N., & Abel-Santos, E. (2011). Mapping interactions between germinants and Clostridium difficile spores. *Journal of Bacteriology*, *193*(1), 274–282. https://doi.org/10.1128/JB.00980-10

Hromada, S., Qian, Y., Jacobson, T. B., Clark, R. L., Watson, L., Safdar, N., … Venturelli, O. S. (2021). Negative interactions determine Clostridioides difficile growth in synthetic human gut communities. *Molecular Systems Biology*, *17*(10). https://doi.org/10.15252/MSB.202110355

Huang, C. N., Liebl, W., & Ehrenreich, A. (2018). Restriction-deficient mutants and marker-less

genomic modification for metabolic engineering of the solvent producer Clostridium saccharobutylicum 06 Biological Sciences 0604 Genetics. *Biotechnology for Biofuels*, *11*(1), 1–13. https://doi.org/10.1186/s13068-018-1260-3

Hui, W., Zhang, W., Kwok, L., Zhang, H., Kong, J., & Sun, T. (2019). A Novel Bacteriophage Exclusion (BREX) System Encoded by. *Applied and Environmental Microbiology*, *85*(20), 1–12.

Humphreys, J. R., Debebe, B. J., Diggle, S. P., & Winzer, K. (2023). Clostridium beijerinckii strain degeneration is driven by the loss of Spo0A activity. *Frontiers in Microbiology*, *13*(January), 1–20. https://doi.org/10.3389/fmicb.2022.1075609

Ingle, P. (2017). *Expanding the Genome Editing Repertoire in Clostridium difficile for Improved Studies of Sporulation and Germination*. University of Nottingham.

Ingle, P., Groothuis, D., Rowe, P., Huang, H., Cockayne, A., Kuehne, S. A., … Minton, N. P. (2019). Generation of a fully erythromycin-sensitive strain of Clostridioides difficile using a novel CRISPR-Cas9 genome editing system. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-44458-y

Isa, K., Oka, K., Beauchamp, N., Sato, M., Wada, K., Ohtani, K., … Takahashi, M. (2016). Safety assessment of the Clostridium butyricum MIYAIRI 588® probiotic strain including evaluation of antimicrobial sensitivity and presence of Clostridium toxin genes in vitro and teratogenicity in vivo. *Human and Experimental Toxicology*, *35*(8), 818–832. https://doi.org/10.1177/0960327115607372

Jenison, R. D., Gill, S. C., Pardi, A., & Polisky, B. (1994). High-resolution molecular discrimination by RNA. *Science*, *263*(5152), 1425–1429. https://doi.org/10.1126/science.7510417

Jin, W. B., Li, T. T., Huo, D., Qu, S., Li, X. V., Arifuzzaman, M., … Guo, C. J. (2022). Genetic manipulation of gut microbes enables single-gene interrogation in a complex microbiome. *Cell*, *185*(3), 547-562.e22. https://doi.org/10.1016/J.CELL.2021.12.035

Johnston, C. D., Cotton, S. L., Rittling, S. R., Starr, J. R., Borisy, G. G., Dewhirst, F. E., & Lemon, K. P. (2019). Systematic evasion of the restriction-modification barrier in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *166*(23), 11454–11459. https://doi.org/10.1073/PNAS.1820256116/SUPPL_FILE/PNAS.1820256116.SAPP.PDF

Jones, B. V., Begley, M., Hill, C., Gahan, C. G. M., & Marchesi, J. R. (2008). Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(36), 13580–13585. https://doi.org/10.1073/pnas.0804437105

Kang, D. J., Ridlon, J. M., Moore, D. R., Barnes, S., & Hylemon, P. B. (2008). Clostridium scindens baiCD and baiH genes encode stereo-specific 7α/7β-hydroxy-3-oxo-Δ4-cholenoic acid oxidoreductases. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, *1781*(1–2), 16–25. https://doi.org/10.1016/j.bbalip.2007.10.008

Kang, J. D., Myers, C. J., Harris, S. C., Kakiyama, G., Lee, I. K., Yun, B. S., … Hylemon, P. B. (2019a). Bile Acid 7α-Dehydroxylating Gut Bacteria Secrete Antibiotics that Inhibit Clostridium difficile: Role of Secondary Bile Acids. *Cell Chemical Biology*, *26*(1), 27. https://doi.org/10.1016/J.CHEMBIOL.2018.10.003

Kang, J. D., Myers, C. J., Harris, S. C., Kakiyama, G., Lee, I. K., Yun, B. S., … Hylemon, P. B. (2019b). Bile Acid 7α-Dehydroxylating Gut Bacteria Secrete Antibiotics that Inhibit Clostridium difficile: Role of Secondary Bile Acids. *Cell Chemical Biology*, *26*(1), 27-34.e4. https://doi.org/10.1016/j.chembiol.2018.10.003

Karasawa, T., Ikoma, S., Yamakawa, K., & Nakamura, S. (1995). A defined growth medium for Clostridium difficile. *Microbiology*, *141*(2), 371–375. https://doi.org/10.1099/13500872-141-2-

Karberg, M., Guo, H., Zhong, J., Coon, R., Perutka, J., & Lambowitz, A. M. (2001). Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nature Biotechnology*, *19*(12), 1162–1167. https://doi.org/10.1038/nbt1201-1162

Karlin, S., Mrázek, J., Campbell, A., & Kaiser, D. (2001). Characterizations of highly expressed genes of four fast-growing bacteria. *Journal of Bacteriology*, *183*(17), 5025–5040. https://doi.org/10.1128/JB.183.17.5025-5040.2001

Keese, P. (2008). Risks from GMOs due to horizontal gene transfer. *Environmental Biosafety Research*, *7*(3), 123–149. https://doi.org/10.1051/ebr:2008014

Kelly, C. P. (2012, January 1). Can we identify patients at high risk of recurrent Clostridium difficile infection? *Clinical Microbiology and Infection*, Vol. 18, pp. 21–27. https://doi.org/10.1111/1469-0691.12046

Kim, K. H., Park, D., Jia, B., Baek, J. H., Hahn, Y., & Jeon, C. O. (2022). Identification and Characterization of Major Bile Acid 7α-Dehydroxylating Bacteria in the Human Gut. *MSystems*, *7*(4). https://doi.org/10.1128/msystems.00455-22

Kitahara, M., Sakamoto, M., & Benno, Y. (2001). PCR detection method of Clostridium scindens and C. hiranonis in human fecal samples. *Microbiology and Immunology*, *45*(3), 263–266. https://doi.org/10.1111/j.1348-0421.2001.tb02616.x

Kitahara, Maki, Takamine, F., Imamura, T., & Benno, Y. (2000). Assignment of Eubacterium sp. VPI 12708 and related strains with high bile acid 7α- dehydroxylating activity to Clostridium scindens and proposal of Clostridium hylemonae sp. nov., isolated from human faeces. In *International Journal of Systematic and Evolutionary Microbiology* (Vol. 50).

Kitahara, Maki, Takamine, F., Imamura, T., & Benno, Y. (2001). Clostridium hiranonis sp. nov., a human intestinal bacterium with bile acid 7α-dehydroxylating activity. In *International Journal of Systematic and Evolutionary Microbiology* (Vol. 51).

Kochan, T. J., Foley, M. H., Shoshiev, M. S., Somers, M. J., Carlson, P. E., & Hanna, P. C. (2018). Updates to Clostridium difficile spore germination. *Journal of Bacteriology*, *200*(16). https://doi.org/10.1128/JB.00218-18

Kochan, T. J., Shoshiev, M. S., Hastie, J. L., Somers, M. J., Plotnick, Y. M., Gutierrez-Munoz, D. F., … Hanna, P. C. (2018). Germinant Synergy Facilitates Clostridium difficile Spore Germination under Physiological Conditions. *MSphere*, *3*(5). https://doi.org/10.1128/msphere.00335-18

Koenigsknecht, M. J., Theriot, C. M., Bergin, I. L., Schumacher, C. A., Schloss, P. D., & Young, V. B. (2015). Dynamics and establishment of Clostridium difficile infection in the murine gastrointestinal tract. *Infection and Immunity*, *83*(3), 934–941. https://doi.org/10.1128/IAI.02768-14

Konstantakos, V., Nentidis, A., Krithara, A., & Paliouras, G. (2022). CRISPR–Cas9 gRNA efficiency prediction: an overview of predictive tools and the role of deep learning. *Nucleic Acids Research*, *50*(7), 3616–3637. https://doi.org/10.1093/NAR/GKAC192

Kosuri, S., Goodman, D. B., Cambray, G., Mutalik, V. K., Gao, Y., Arkin, A. P., … Church, G. M. (2013). Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(34), 14024–14029. https://doi.org/10.1073/pnas.1301301110

Kovács, K., Willson, B. J., Schwarz, K., Heap, J. T., Jackson, A., Bolam, D. N., … Minton, N. P. (2013). Secretion and assembly of functional mini-cellulosomes from synthetic chromosomal operons in Clostridium acetobutylicum ATCC 824. *Biotechnology for Biofuels*, *6*, 1. https://doi.org/10.1186/1754-6834-6-117

Kozakai, T., Shimofusa, Y., Nomura, I., & Suzuki, T. (2021). Construction of a reporter system for bifidobacteria using chloramphenicol acetyltransferase and its application for evaluation of promoters and terminators. *Bioscience of Microbiota, Food and Health*, *40*(2), 115–122. https://doi.org/10.12938/bmfh.2020-070

Kuehne, S A, Rood, J. I., & Lyras, D. (2019). Clostridial Genetics: Genetic Manipulation of the Pathogenic Clostridia. *Microbiology Spectrum*, *7*(3). https://doi.org/10.1128/microbiolspec.gpp3-0040-2018

Kuehne, Sarah A., & Minton, N. P. (2012). ClosTron-mediated engineering of Clostridium. *Bioengineered*, *3*(4), 247. https://doi.org/10.4161/BIOE.21004

Lance George, W., Goldstein, E. J. C., Sutter, V. L., Ludwig, S. L., & Finegold, S. M. (1978). ÆTIOLOGY OF ANTIMICROBIAL-AGENT-ASSOCIATED COLITIS. *The Lancet*, *311*(8068), 802–803. https://doi.org/10.1016/S0140-6736(78)93001-5

Lancet Infectious Diseases, T. (2019). *C difficile—a rose by any other name…*. https://doi.org/10.1016/j

Lato, D. F., & Golding, G. B. (2020). Spatial Patterns of Gene Expression in Bacterial Genomes. *Journal of Molecular Evolution*, *88*(6), 510–520. https://doi.org/10.1007/s00239-020-09951-3

Lawson, P. A., Citron, D. M., Tyrrell, K. L., & Finegold, S. M. (2016). Reclassification of Clostridium difficile as Clostridioides difficile (Hall and O'Toole 1935) Prévot 1938. *Anaerobe*, *40*, 95–99. https://doi.org/10.1016/j.anaerobe.2016.06.008

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

Lee, J. W., Cowley, E. S., Wolf, P. G., Doden, H. L., Murai, T., Caicedo, K. Y. O., … Ridlon, J. M. (2022). Formation of secondary allo-bile acids by novel enzymes from gut Firmicutes. *Gut Microbes*, *14*(1). https://doi.org/10.1080/19490976.2022.2132903

Lesiak, J. M., Liebl, W., & Ehrenreich, A. (2014). Development of an in vivo methylation system for the solventogen Clostridium saccharobutylicum NCP 262 and analysis of two endonuclease mutants. *Journal of Biotechnology*, *188*, 97–99. https://doi.org/10.1016/j.jbiotec.2014.07.005

Leslie, J. L., Vendrov, K. C., Jenior, M. L., & Young, V. B. (2019). The Gut Microbiota Is Associated with Clearance of Clostridium difficile Infection Independent of Adaptive Immunity. *MSphere*, *4*(1). https://doi.org/10.1128/mspheredirect.00698-18

Li, J., Adams, V., Bannam, T. L., Miyamoto, K., Garcia, J. P., Uzal, F. A., … McClane, B. A. (2013). Toxin Plasmids of Clostridium perfringens. *Microbiology and Molecular Biology Reviews*, *77*(2), 208–233. https://doi.org/10.1128/mmbr.00062-12

Li, Q., Chen, J., Minton, N. P., Zhang, Y., Wen, Z., Liu, J., … Yang, S. (2016). CRISPR-based genome editing and expression control systems in Clostridium acetobutylicum and Clostridium beijerinckii. *Biotechnology Journal*, *11*(7), 961–972. https://doi.org/10.1002/biot.201600053

Li, X., Chu, Q., Huang, Y., Xiao, Y., Song, L., Zhu, S., … Ren, Z. (2019). Consortium of Probiotics Attenuates Colonization of Clostridioides difficile. *Frontiers in Microbiology*, *10*(December), 1–12. https://doi.org/10.3389/fmicb.2019.02871

Lopatkin, A. J., Meredith, H. R., Srimani, J. K., Pfeiffer, C., Durrett, R., & You, L. (2017). Persistence and reversal of plasmid-mediated antibiotic resistance. *Nature Communications 2017 8:1*, *8*(1), 1–10. https://doi.org/10.1038/s41467-017-01532-1

Louie, T. J., Miller, M. A., Mullane, K. M., Weiss, K., Lentnek, A., Golan, Y., … Shue, Y.-K. (2011). Fidaxomicin versus Vancomycin for *Clostridium difficile* Infection. *New England Journal of Medicine*, *364*(5), 422–431. https://doi.org/10.1056/NEJMoa0910812

Luo, M. L., Leenay, R. T., & Beisel, C. L. (2016). Current and future prospects for CRISPR-based tools in bacteria. *Biotechnology and Bioengineering*, *113*(5), 930–943. https://doi.org/10.1002/bit.25851

Luyten, Y. A., Hausman, D. E., Young, J. C., Doyle, L. A., Higashi, K. M., Ubilla-Rodriguez, N. C., … Kaiser, B. K. (2022). Identification and characterization of the WYL BrxR protein and its gene as separable regulatory elements of a BREX phage restriction system. *Nucleic Acids Research*, *50*(2), 5190. https://doi.org/10.1093/nar/gkac311

Lynch, S. A., Desai, S. K., Sajja, H. K., & Gallivan, J. P. (2007). A High-Throughput Screen for Synthetic Riboswitches Reveals Mechanistic Insights into Their Function. *Chemistry and Biology*, *14*(2), 173–184. https://doi.org/10.1016/j.chembiol.2006.12.008

Lyras, D., O'Connor, J. R., Howarth, P. M., Sambol, S. P., Carter, G. P., Phumoonna, T., … Rood, J. I. (2009). Toxin B is essential for virulence of Clostridium difficile. *Nature*, *458*(7242), 1176–1179. https://doi.org/10.1038/nature07822

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

Madan, R., & Petri, W. A. (2012). Immune responses to Clostridium difficile infection. *Trends in Molecular Medicine*, *18*(11), 658–666. https://doi.org/10.1016/j.molmed.2012.09.005

Maikova, A., Kreis, V., Boutserin, A., Severinov, K., & Soutourina, O. (2019). Using an endogenous CRISPR-Cas system for genome editing in the human pathogen Clostridium difficile. *Applied and Environmental Microbiology*, *85*(20), e01416-19. https://doi.org/10.1128/AEM.01416-19

Mallonee, D H, Lijewski, M. A., & Hylemon, P. B. (1995). Expression in Escherichia coli and characterization of a bile acid-inducible 3 alpha-hydroxysteroid dehydrogenase from Eubacterium sp. strain VPI 12708. *Current Microbiology*, *30*(5), 259–263. https://doi.org/10.1007/bf00295498

Mallonee, Darrell H., & Hylemon, P. B. (1996a). Sequencing and expression of a gene encoding a bile acid transporter from Eubacterium sp. strain VPI 12708. *Journal of Bacteriology*, *178*(24), 7053–7058. https://doi.org/10.1128/jb.178.24.7053-7058.1996

Mallonee, Darrell H, Adams, J. L., & Hylemon, P. B. (1992). The Bile Acid-Inducible baiB Gene from Eubacterium sp. Strain VPI 12708 Encodes a Bile Acid-Coenzyme A Ligase. In *JOURNAL OF BACTERIOLOGY* (Vol. 174).

Mallonee, Darrell H, & Hylemon, P. B. (1996b). Sequencing and Expression of a Gene Encoding a Bile Acid Transporter from Eubacterium sp. Strain VPI 12708. In *JOURNAL OF BACTERIOLOGY* (Vol. 178).

Mallonee, Darrell H, & White, W. B. (1990). Cloning and Sequencing of a Bile Acid-Inducible Operon from. *Journal of Bacteriology*, *172*(12), 7011–7019.

Marchesini, M. I., Connolly, J., Delpino, M. V., Baldi, P. C., Mujer, C. V., DelVecchio, V. G., & Comerci, D. J. (2011). Brucella abortus choloylglycine hydrolase affects cell envelope composition and host cell internalization. *PLoS ONE*, *6*(12). https://doi.org/10.1371/journal.pone.0028480

Marion, S., Studer, N., Desharnais, L., Menin, L., Escrig, S., Meibom, A., … Bernier-Latmani, R. (2019). In vitro and in vivo characterization of Clostridium scindens bile acid transformations. *Gut Microbes*, *10*(4), 481–503. https://doi.org/10.1080/19490976.2018.1549420

Matteau, D. & Rodrigue, S. (2015). *Precise Identification of Genome-Wide Transcription Start Sites in Bacteria by 5'-Rapid Amplification of cDNA Ends (5'-RACE)* (Methods in).

McAllister, K. N., Bouillaut, L., Kahn, J. N., Self, W. T., & Sorg, J. A. (2017). Using CRISPR-Cas9-mediated genome editing to generate C. difficile mutants defective in selenoproteins synthesis. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-15236-5

McDonald, L. C., Gerding, D. N., Johnson, S., Bakken, J. S., Carroll, K. C., Coffin, S. E., … Wilcox, M. H. (2018). Clinical Practice Guidelines for Clostridium difficile Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clinical Infectious Diseases*, Vol. 66, pp. e1–e48. https://doi.org/10.1093/cid/cix1085

Meng, X., Karasawa, T., Zou, K., Kuang, X., Wang, X., Lu, C., … Nakamura, S. (1997). Characterization of a neurotoxigenic Clostridium butyricum strain isolated from the food implicated in an outbreak of food-borne type E botulism. *Journal of Clinical Microbiology*, *35*(8), 2160. https://doi.org/10.1128/jcm.35.8.2160-2162.1997

Millan, A. S., Peña-Miller, R., Toll-Riera, M., Halbert, Z. V., McLean, A. R., Cooper, B. S., & Maclean, R. C. (2014). Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids. *Nature Communications*, *5*, 1–11. https://doi.org/10.1038/ncomms6208

Minton, N., & Morris, J. G. (1981). Isolation and partial characterization of three cryptic plasmids from strains of Clostridium butyricum. *Journal of General Microbiology*, *127*(2), 325–331. https://doi.org/10.1099/00221287-127-2-325

Minton, N. P., Ehsaan, M., Humphreys, C. M., Little, G. T., Baker, J., Henstra, A. M., … Zhang, Y. (2016). A roadmap for gene system development in Clostridium. *Anaerobe*, *41*, 104–112. https://doi.org/10.1016/j.anaerobe.2016.05.011

Moore, P., Kyne, L., Martin, A., & Solomon, K. (2013). Germination efficiency of clinical Clostridium difficile spores and correlation with ribotype, disease severity and therapy failure. *Journal of Medical Microbiology*, *62*, 1405–1413. https://doi.org/10.1099/jmm.0.056614-0

Mowday, A. M., Dubois, L. J., Kubiak, A. M., Chan-Hyams, J. V. E., Guise, C. P., Ashoorzadeh, A., … Patterson, A. V. (2022). Use of an optimised enzyme/prodrug combination for Clostridia directed enzyme prodrug therapy induces a significant growth delay in necrotic tumours. *Cancer Gene Therapy*, *29*(2), 178–188. https://doi.org/10.1038/s41417-021-00296-7

Nakanishi, S., & Tanaka, M. (2010). Sequence analysis of a bacteriocinogenic plasmid of Clostridium butyricum and expression of the bacteriocin gene in Escherichia coli. *Anaerobe*, *16*(3), 253–257. https://doi.org/10.1016/j.anaerobe.2009.10.002

Narushima, S., Itoh, K., Miyamoto, Y., Park, S.-H., Nagata, K., Kuruma, K., & Uchida, K. (2006). Deoxycholic Acid Formation in Gnotobiotic Mice Associated with Human Intestinal Bacteria. *Lipids*, *41*(9), 835–843.

Narushima, S., Itoh, K., Takamine, F., & Uchida, K. (1999a). Absence of cecal secondary bile acids in gnotobiotic mice associated with two human intestinal bacteria with the ability to dehydroxylate bile acids in vitro. *Microbiology and Immunology*, *43*(9), 893–897. https://doi.org/10.1111/j.1348-0421.1999.tb01224.x

Narushima, S., Itoh, K., Takamine, F., & Uchida, K. (1999b). Absence of Cecal Secondary Bile Acids in Gnotobiotic Mice Associated with Two Human Intestinal Bacteria with the Ability to Dehydroxylate Bile Acids In Vitro. *Microbiology and Immunology*, *43*(9), 893–897. https://doi.org/10.1111/j.1348-0421.1999.tb01224.x

Nelson, R. L., Suda, K. J., & Evans, C. T. (2017). Antibiotic treatment for Clostridium difficile-associated diarrhoea in adults. *Cochrane Database of Systematic Reviews*, *2017*(3). https://doi.org/10.1002/14651858.CD004610.pub5

Ng, Y. K., Ehsaan, M., Philip, S., Collery, M. M., Janoir, C., Collignon, A., … 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). https://doi.org/10.1371/journal.pone.0056051

Pandak, W. M., & Kakiyama, G. (2019). The acidic pathway of bile acid synthesis: Not just an alternative pathway. *Liver Research*, *3*(2), 88–98. https://doi.org/10.1016/j.livres.2019.05.001

Paredes-Sabja, D., Shen, A., & Sorg, J. A. (2014). Clostridium difficile spore biology: Sporulation, germination, and spore structural proteins. *Trends in Microbiology*, Vol. 22, pp. 406–416. https://doi.org/10.1016/j.tim.2014.04.003

Pereira, F. C., Saujet, L., Tomé, A. R., Serrano, M., Monot, M., Couture-Tosi, E., … Henriques, A. O. (2013). The Spore Differentiation Pathway in the Enteric Pathogen Clostridium difficile. *PLoS Genetics*, *9*(10). https://doi.org/10.1371/journal.pgen.1003782

Pettit, L. J., Browne, H. P., Yu, L., Klaas Smits, W., Fagan, R. P., Barquist, L., … Lawley, T. D. (2014). *Functional genomics reveals that Clostridium difficile Spo0A coordinates sporulation, virulence and metabolism*. https://doi.org/10.1186/1471-2164-15-160

Poulalier-Delavelle, M., Baker, J. P., Millard, J., Winzer, K., & Minton, N. P. (2023). Endogenous CRISPR / Cas systems for genome engineering in the acetogens Acetobacterium woodii and Clostridium autoethanogenum. *Frontiers in Bioengineering and Biotechnology*, *11*(June), 1–17. https://doi.org/10.3389/fbioe.2023.1213236

Prensky, H., Gomez-Simmonds, A., Uhlemann, A.-C., & Lopatkin, A. J. (2021). Conjugation dynamics depend on both the plasmid acquisition cost and the fitness cost. *Molecular Systems Biology*, *17*(3), e9913. https://doi.org/10.15252/MSB.20209913

Public Health England. (2019). *Annual epidemiological commentary: bacteraemia, MSSA bacteraemia and C . difficile infections, up to and including financial year April 2018 to March 2019*. Retrieved from www.facebook.com/PublicHealthEngland

Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A., & Chou, C. P. (2016). Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Scientific Reports 2016 6:1*, *6*(1), 1–15. https://doi.org/10.1038/srep25666

Raibaud, P., Ducluzeau, R., Muller, M., & Sacquet, E. (1974). Le taurocholate de sodium, facteur de germination in vitro et in vivo. *Ann Microbiol (Paris)*, *Oct-Nov*(125B(3)), 381–391.

Rangel-Chavez, C., Galan-Vasquez, E., & Martinez-Antonio, A. (2017). Consensus architecture of promoters and transcription units in: Escherichia coli: Design principles for synthetic biology. *Molecular BioSystems*, *13*(4), 665–676. https://doi.org/10.1039/c6mb00789a

Reed, A. D., Fletcher, J. R., Huang, Y. Y., Thanissery, R., Rivera, A. J., Parsons, R. J., … Theriot, C. M. (2022). The Stickland Reaction Precursor trans -4-Hydroxy- l -Proline Differentially Impacts the Metabolism of Clostridioides difficile and Commensal Clostridia. *MSphere*, *7*(2). https://doi.org/10.1128/msphere.00926-21

Reed, A. D., Nethery, M. A., Stewart, A., Barrangou, R., & Theriot, C. M. (2020). Strain-dependent inhibition of clostridioides difficile by commensal clostridia carrying the bile acid-inducible (bai) operon. *Journal of Bacteriology*, *202*(11). https://doi.org/10.1128/JB.00039-20

Reeves, A. E., Theriot, C. M., Bergin, I. L., Huffnagle, G. B., Schloss, P. D., & Young, V. B. (2011). The interplay between microbiome dynamics and pathogen dynamics in a murine model of Clostridium difficile infection. *Gut Microbes*, *2*(3), 145–158. https://doi.org/10.4161/gmic.2.3.16333

Ridlon, J. M., Devendran, S., Alves, J. M., Doden, H., Wolf, P. G., Pereira, G. V., … Gaskins, H. R. (2020). The 'in vivo lifestyle' of bile acid 7α-dehydroxylating bacteria: comparative genomics, metatranscriptomic, and bile acid metabolomics analysis of a defined microbial community in gnotobiotic mice. *Gut Microbes*, *11*(3), 381–404.

https://doi.org/10.1080/19490976.2019.1618173

Ridlon, J. M., Harris, S. C., Bhowmik, S., Kang, D. J., & Hylemon, P. B. (2016, January 2). Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes*, Vol. 7, pp. 22–39. https://doi.org/10.1080/19490976.2015.1127483

Ridlon, J. M., & Hylemon, P. B. (2012). Identification and characterization of two bile acid coenzyme A transferases from Clostridium scindens, a bile acid 7α-dehydroxylating intestinal bacterium. *Journal of Lipid Research*, *53*(1), 66–76. https://doi.org/10.1194/jlr.M020313

Ridlon, J. M., Kang, D. J., & Hylemon, P. B. (2006, February). Bile salt biotransformations by human intestinal bacteria. *Journal of Lipid Research*, Vol. 47, pp. 241–259. https://doi.org/10.1194/jlr.R500013-JLR200

Ridlon, J. M., Kang, D. J., & Hylemon, P. B. (2010). Isolation and characterization of a bile acid inducible 7α-dehydroxylating operon in Clostridium hylemonae TN271. *Anaerobe*, *16*(2), 137–146. https://doi.org/10.1016/j.anaerobe.2009.05.004

Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2015). REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Research*, *43*(Database issue), D298-9. https://doi.org/10.1093/nar/gku1046

Rodriguez-Palacios, A., Borgmann, S., Kline, T. R., & LeJeune, J. T. (2013, June). Clostridium difficile in foods and animals: history and measures to reduce exposure. *Animal Health Research Reviews / Conference of Research Workers in Animal Diseases*, Vol. 14, pp. 11–29. https://doi.org/10.1017/S1466252312000229

Rodriguez, C., Taminiau, B., Van Broeck, J., Delmée, M., & Daube, G. (2016). Clostridium difficile in food and animals: A comprehensive review. In *Advances in Experimental Medicine and Biology* (Vol. 932, pp. 65–92). https://doi.org/10.1007/5584_2016_27

Rodriguez, C., Van Broeck, J., Taminiau, B., Delmée, M., & Daube, G. (2016, August 1). Clostridium difficile infection: Early history, diagnosis and molecular strain typing methods. *Microbial Pathogenesis*, Vol. 97, pp. 59–78. https://doi.org/10.1016/j.micpath.2016.05.018

Roer, L., Aarestrup, F. M., & Hasman, H. (2015). The EcoKI Type I Restriction-Modification System in Escherichia coli Affects but Is Not an Absolute Barrier for Conjugation. *Journal of Bacteriology*, *197*(2), 337. https://doi.org/10.1128/JB.02418-14

Rosenberg, E. Y., Bertenthal, D., Nilles, M. L., Bertrand, K. P., & Nikaido, H. (2003). Bile salts and fatty acids induce the expression of Escherichia coli AcrAB multidrug efflux pump through their interaction with Rob regulatory protein. *Molecular Microbiology*, *48*(6), 1609–1619. https://doi.org/10.1046/j.1365-2958.2003.03531.x

Rupnik, M., Wilcox, M. H., & Gerding, D. N. (2009). Clostridium difficile infection: New developments in epidemiology and pathogenesis. *Nature Reviews Microbiology*, Vol. 7, pp. 526–536. https://doi.org/10.1038/nrmicro2164

San Millan, A., & Maclean, R. C. (2019). Fitness costs of plasmids: A limit to plasmid transmission. *Microbial Transmission*, (19), 65–79. https://doi.org/10.1128/9781555819743.ch4

Sánchez-Hurtado, K., Corretge, M., Mutlu, E., McIlhagger, R., Starr, J. M., & Poxton, I. R. (2008). Systemic antibody response to Clostridium difficile in colonized patients with and without symptoms and matched controls. *Journal of Medical Microbiology*, *57*(6), 717–724. https://doi.org/10.1099/jmm.0.47713-0

Sarathchandra, S., Barker, A., & Wolf, J. (1973). Studies on the germination of Clostridium butyricum. In *Spore research* (pp. 207–231). London, United Kingdom: Academic Press.

Sartelli, M., Di Bella, S., McFarland, L. V., Khanna, S., Furuya-Kanamori, L., Abuzeid, N., … Catena,

F. (2019). 2019 update of the WSES guidelines for management of Clostridioides (Clostridium) difficile infection in surgical patients. *World Journal of Emergency Surgery*, *14*(1), 8. https://doi.org/10.1186/s13017-019-0228-3

Saujet, L., Pereira, F. C., Henriques, A. O., & Martin-Verstraete, I. (2014). The regulatory network controlling spore formation in Clostridium difficile. *FEMS Microbiology Letters*, *358*(1), 1–10. https://doi.org/10.1111/1574-6968.12540

Schäffler, H., & Breitrück, A. (2018, April 10). Clostridium difficile - From colonization to infection. *Frontiers in Microbiology*, Vol. 9, p. 646. https://doi.org/10.3389/fmicb.2018.00646

Schubert, A. M., Sinani, H., & Schloss, P. D. (2015). Antibiotic-induced alterations of the murine gut microbiota and subsequent effects on colonization resistance against Clostridium difficile. *MBio*, *6*(4). https://doi.org/10.1128/mBio.00974-15

Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., … Parkhill, J. (2006). The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nature Genetics*, *38*(7), 779–786. https://doi.org/10.1038/ng1830

Seddon, S. V, & Borriello, S. P. (1989). A chemically defined and minimal medium for Clostridium dificile. In *Letters in Applied Microbiology*.

Seki, H., Shiohara, M., Matsumura, T., Miyagawa, N., Tanaka, M., Komiyama, A., & Kurata, S. (2003). Prevention of antibiotic-associated diarrhea in children by Clostridium butyricum MIYAIRI. *Pediatrics International*, *45*(1), 86–90. https://doi.org/10.1046/j.1442-200X.2003.01671.x

Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., … Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(25), 10098–10103. https://doi.org/10.1073/PNAS.1104144108/-/DCSUPPLEMENTAL

Setlow, P. (2014). Germination of spores of Bacillus species: What we know and do not know. *Journal of Bacteriology*, Vol. 196, pp. 1297–1305. https://doi.org/10.1128/JB.01455-13

Setlow, P., Wang, S., & Li, Y.-Q. (2017). Germination of Spores of the Orders Bacillales and Clostridiales. *Annual Review of Microbiology*, *71*(1), 459–477. https://doi.org/10.1146/annurev-micro-090816-093558

Seys, F. M., Rowe, P., Bolt, E. L., Humphreys, C. M., & Minton, N. P. (2020). A gold standard, CRISPR/Cas9-based complementation strategy reliant on 24 nucleotide bookmark sequences. *Genes*, *11*(4). https://doi.org/10.3390/genes11040458

Shah, S. A., Erdmann, S., Mojica, F. J. M., & Garrett, R. A. (2013). Protospacer recognition motifs: Mixed identities and functional diversity. *RNA Biology*, *10*(5), 891. https://doi.org/10.4161/RNA.23764

Shaw, W. V. (1975). Chloramphenicol Acetyltransferase from Chloramphenicol-Resistant Bacteria. *Methods in Enzymology*, *43*(C), 737–755. https://doi.org/10.1016/0076-6879(75)43141-X

Shelby, R. D., Janzow, G. E., Mashburn-Warren, L., Galley, J., Tengberg, N., Navarro, J., … Besner, G. E. (2020). A novel probiotic therapeutic in a murine model of Clostridioides difficile colitis. *Gut Microbes*, *12*(1), 1–13. https://doi.org/10.1080/19490976.2020.1814119

Silverstein, R. M. (1975). The determination of the molar extinction coefficient of reduced DTNB. *Analytical Biochemistry*, *63*(1), 281–282.

Smith, L. D., & King, E. O. (1962). Occurrence of Clostridium difficile in infections of man. *Journal of Bacteriology*, *84*, 65–67. https://doi.org/10.1002/path.1700840108

Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H., & Kuijper, E. J. (2016). Clostridium difficile infection. *Nature Reviews Disease Primers*, *2*(1), 16020. https://doi.org/10.1038/nrdp.2016.20

So, J. S., Oh, K., & Shin, Y. (2022). Growth stimulation of Clostridium butyricum in the presence of Lactobacillus brevis JL16 and Lactobacillus parabuchneri MH44. *Food Science and Technology (Brazil)*, *42*, 1–8. https://doi.org/10.1590/fst.50521

Solomon, K. (2013). The host immune response to Clostridium difficile infection. *Therapeutic Advances in Infectious Disease*, *1*(1), 19–35. https://doi.org/10.1177/2049936112472173

Song, I., Gotoh, Y., Ogura, Y., Hayashi, T., Fukiya, S., & Yokota, A. (2021). Comparative genomic and physiological analysis against clostridium scindens reveals eubacterium sp. C-25 as an atypical deoxycholic acid producer of the human gut microbiota. *Microorganisms*, *9*(11). https://doi.org/10.3390/microorganisms9112254

Song, Y., Nikoloff, J. M., Fu, G., Chen, J., Li, Q., Xie, N., … Zhang, D. (2016). Promoter Screening from Bacillus subtilis in Various Conditions Hunting for Synthetic Biology and Industrial Applications. *PLoS ONE*, *11*(7). https://doi.org/10.1371/JOURNAL.PONE.0158447

Sorg, J. A., & Sonenshein, A. L. (2008). Bile salts and glycine as cogerminants for Clostridium difficile spores. *Journal of Bacteriology*, *190*(7), 2505–2512. https://doi.org/10.1128/JB.01765-07

Sorg, J. A., & Sonenshein, A. L. (2009). Chenodeoxycholate is an inhibitor of Clostridium difficile spore germination. *Journal of Bacteriology*, *191*(3), 1115–1117. https://doi.org/10.1128/JB.01260-08

Sorg, J. A., & Sonenshein, A. L. (2010). Inhibiting the initiation of Clostridium difficile spore germination using analogs of chenodeoxycholic acid, a bile acid. *Journal of Bacteriology*, *192*(19), 4983–4990. https://doi.org/10.1128/JB.00610-10

Staels, B., & Fonseca, V. A. (2009). Bile acids and metabolic regulation: mechanisms and clinical responses to bile acid sequestration. *Diabetes Care*. https://doi.org/10.2337/dc09-s355

Stewart, F. J., Panne, D., Bickle, T. A., & Raleigh, E. A. (2000). Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *Journal of Molecular Biology*, *298*(4), 611–622. https://doi.org/10.1006/jmbi.2000.3697

Stone, N. E., Nunnally, A. E., Jimenez, V., Cope, E. K., Sahl, J. W., Sheridan, K., … Wagner, D. M. (2019, August 1). Domestic canines do not display evidence of gut microbial dysbiosis in the presence of Clostridioides (Clostridium) difficile, despite cellular susceptibility to its toxins. *Anaerobe*, Vol. 58, pp. 53–72. https://doi.org/10.1016/j.anaerobe.2019.03.017

Stone, N. E., Sidak-Loftis, L. C., Sahl, J. W., Vazquez, A. J., Wiggins, K. B., Gillece, J. D., … Wagner, D. M. (2016). More than 50% of clostridium difficile isolates from pet dogs in Flagstaff, USA, carry toxigenic genotypes. *PLoS ONE*, *11*(10), 1–21. https://doi.org/10.1371/journal.pone.0164504

Streett, H. E., Kalis, K. M., & Papoutsakis, E. T. (2019). A Strongly Fluorescing Anaerobic Reporter and Protein-Tagging System for Clostridium Organisms Based on the Fluorescence-Activating and Absorption-Shifting Tag Protein (FAST). *Applied and Environmental Microbiology*, *85*(14), 1–15. https://doi.org/10.1128/AEM.00622-19

Studer, N., Desharnais, L., Beutler, M., Brugiroux, S., Terrazos, M. A., Menin, L., … Hapfelmeier, S. (2016). Functional intestinal bile acid 7α-dehydroxylation by Clostridium scindens associated with protection from Clostridium difficile infection in a gnotobiotic mouse model. *Frontiers in Cellular and Infection Microbiology*, *6*(DEC). https://doi.org/10.3389/fcimb.2016.00191

Sundaresan, A., Le Ngoc, M., Wew, M. U., Ramkumar, V., Raninga, P., Sum, R., & Cheong, I. (2023). A design of experiments screen reveals that Clostridium novyi-NT spore germinant sensing is stereoflexible for valine and its analogs. *Communications Biology*, *6*(1), 1–15.

https://doi.org/10.1038/s42003-023-04496-9

Swann, J. R., Want, E. J., Geier, F. M., Spagou, K., Wilson, I. D., Sidaway, J. E., … Holmes, E. (2011). Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(SUPPL. 1), 4523–4530. https://doi.org/10.1073/pnas.1006734107

Talukdar, P. K., Olguín-Araneda, V., Alnoman, M., Paredes-Sabja, D., & Sarker, M. R. (2015). Updates on the sporulation process in Clostridium species. *Research in Microbiology*, Vol. 166, pp. 225–235. https://doi.org/10.1016/j.resmic.2014.12.001

Thanassi, D. G., Cheng, L. W., & Nikaido, H. (1997). Active efflux of bile salts by Escherichia coli. *Journal of Bacteriology*, *179*(8), 2512–2518. https://doi.org/10.1128/jb.179.8.2512-2518.1997

Thanissery, R., McLaren, M. R., Rivera, A., Reed, A. D., Betrapally, N. S., Burdette, T., … Theriot, C. M. (2020). C. difficile carriage in animals and the associated changes in the host fecal microbiota. *Anaerobe*, *66*. https://doi.org/10.1016/j.anaerobe.2020.102279

Thanissery, Rajani, Winston, J. A., & Theriot, C. M. (2017). Inhibition of spore germination, growth, and toxin activity of clinically relevant C. difficile strains by gut microbiota derived secondary bile acids. *Anaerobe*. https://doi.org/10.1016/j.anaerobe.2017.03.004

The European Commission. (2014). C(2014) 9345. *Official Journal of the European Union*, *588*(March), 30–33.

Theriot, C. M., Bowman, A. A., & Young, V. B. (2016). Antibiotic-Induced Alterations of the Gut Microbiota Alter Secondary Bile Acid Production and Allow for Clostridium difficile Spore Germination and Outgrowth in the Large Intestine . *MSphere*, *1*(1). https://doi.org/10.1128/msphere.00045-15

Theriot, C. M., Koenigsknecht, M. J., Carlson, P. E., Hatton, G. E., Nelson, A. M., Li, B., … Young, V. B. (2014). Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms4114

Theriot, C. M., & Young, V. B. (2015). Interactions Between the Gastrointestinal Microbiome and Clostridium difficile. *Annual Review of Microbiology*, *69*(1), 445–461. https://doi.org/10.1146/annurev-micro-091014-104115

Thomas, J. P., Modos, D., Rushbrook, S. M., Powell, N., & Korcsmaros, T. (2022). The Emerging Role of Bile Acids in the Pathogenesis of Inflammatory Bowel Disease. *Frontiers in Immunology*, *13*(February), 1–14. https://doi.org/10.3389/fimmu.2022.829525

Ticho, A. L., Malhotra, P., Dudeja, P. K., Gill, R. K., & Alrefai, W. A. (2019). Bile acid receptors and gastrointestinal functions. *Liver Research*, *3*(1), 31–39. https://doi.org/10.1016/j.livres.2019.01.001

Topp, S., Reynoso, C. M. K., Seeliger, J. C., Goldlust, I. S., Desai, S. K., Murat, D., … Gallivan, J. P. (2010). Synthetic Riboswitches That Induce Gene Expression in Diverse Bacterial Species †. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, *76*(23), 7881–7884. https://doi.org/10.1128/AEM.01537-10

Truong, Q. L., Cho, Y., Barate, A. K., Kim, S., Watarai, M., & Hahn, T. W. (2015). Mutation of purD and purF genes further attenuates Brucella abortus strain RB51. *Microbial Pathogenesis*, *79*, 1–7. https://doi.org/10.1016/J.MICPATH.2014.12.003

Tummala, S. B., Welker, N. E., & Papoutsakis, E. T. (1999). Development and characterization of a gene expression reporter system for Clostridium acetobutylicum ATCC 824. *Applied and Environmental Microbiology*, *65*(9), 3793–3799. https://doi.org/10.1128/aem.65.9.3793-3799.1999

Underwood, S., Guan, S., Vijayasubhash, V., Baines, S. D., Graham, L., Lewis, R. J., … Stephenson, K. (2009). Characterization of the sporulation initiation pathway of Clostridium difficile and its role in toxin production. *Journal of Bacteriology*, *191*(23), 7296–7305. https://doi.org/10.1128/JB.00882-09

Urdaneta, V., & Casadesús, J. (2017, October 3). Interactions between bacteria and bile salts in the gastrointestinal and hepatobiliary tracts. *Frontiers in Medicine*, Vol. 4, p. 163. https://doi.org/10.3389/fmed.2017.00163

V. Solovyev, A. S. (2011). Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies. *Nova Science Publishers*, (February), 61–78. Retrieved from https://www.researchgate.net/publication/259450599_V_Solovyev_A_Salamov_2011_Automatic_Annotation_of_Microbial_Genomes_and_Metagenomic_Sequences_In_Metagenomics_and_its_Applications_in_Agriculture_Biomedicine_and_Environmental_Studies_Ed_RW_Li_Nova_Sc

Van Nood, E., Vrieze, A., Nieuwdorp, M., Fuentes, S., Zoetendal, E. G., De Vos, W. M., … Keller, J. J. (2013). Duodenal infusion of donor feces for recurrent clostridium difficile. *New England Journal of Medicine*, *368*(5), 407–415. https://doi.org/10.1056/NEJMoa1205037

Vasu, K., & Nagaraja, V. (2013). Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. *Microbiology and Molecular Biology Reviews*, *77*(1), 53–72. https://doi.org/10.1128/mmbr.00044-12

Vidal, J. E., Ma, M., Saputo, J., Garcia, J., Uzal, F. A., & 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*, *83*(1), 179–194. https://doi.org/10.1111/j.1365-2958.2011.07925.x

Vincent, C., Stephens, D. A., Loo, V. G., Edens, T. J., Behr, M. A., Dewar, K., & Manges, A. R. (2013). *Reductions in intestinal Clostridiales precede the development of nosocomial Clostridium difficile infection*. https://doi.org/10.1186/2049-2618-1-18

Vital, M., Rud, T., Rath, S., Pieper, D. H., & Schlüter, D. (2019). Diversity of Bacteria Exhibiting Bile Acid-inducible 7α-dehydroxylation Genes in the Human Gut. *Computational and Structural Biotechnology Journal*, *17*, 1016–1019. https://doi.org/10.1016/j.csbj.2019.07.012

Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, et al. (2009). Genus I. Clostridium Prazmowski, 1880. In *Bergeys Manual of Systematic Bacteriology. The Firmicutes* (9th ed.). New York, Springer.

Wang, S., Hong, W., Dong, S., Zhang, Z. T., Zhang, J., Wang, L., & Wang, Y. (2018). Genome engineering of Clostridium difficile using the CRISPR-Cas9 system. *Clinical Microbiology and Infection*, *24*(10), 1095–1099. https://doi.org/10.1016/j.cmi.2018.03.026

Wang, X. liang, Chen, W. jie, Jin, R., Xu, X., Wei, J., Huang, H., … Chen, T. tao. (2023). Engineered probiotics Clostridium butyricum-pMTL007-GLP-1 improves blood pressure via producing GLP-1 and modulating gut microbiota in spontaneous hypertension rat models. *Microbial Biotechnology*, *16*(4), 799–812. https://doi.org/10.1111/1751-7915.14196

Wang, Y., Chen, W. jie, Han, Y. yang, Xu, X., Yang, A. xia, Wei, J., … Chen, T. tao. (2023). Neuroprotective effect of engineered Clostridium butyricum-pMTL007-GLP-1 on Parkinson's disease mice models via promoting mitophagy. *Bioengineering and Translational Medicine*, *8*(3), 1–15. https://doi.org/10.1002/btm2.10505

Wein, T., Hülter, N. F., Mizrahi, I., & Dagan, T. (2019). Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. *Nature Communications*, *10*(1), 1–13. https://doi.org/10.1038/s41467-019-10600-7

Wells, J. E., & Hylemon, P. B. (2000). Identification and characterization of a bile acid 7α-dehydroxylation operon in Clostridium sp. strain TO-931, a highly active 7α-dehydroxylating strain isolated from human feces. *Applied and Environmental Microbiology*, *66*(3), 1107–1113. https://doi.org/10.1128/AEM.66.3.1107-1113.2000

White, B. A., Lipsky, R. L., Fricke, R. J., & Hylemon, P. B. (1980). Bile acid induction specificity of 7 alpha-dehydroxylase activity in an intestinal Eubacterium species. *Steroids*, *35*(1), 103–109. https://doi.org/10.1016/0039-128x(80)90115-4

Wilcox, M. H., Shetty, N., Fawley, W. N., Shemko, M., Coen, P., Birtles, A., … Wren, M. W. D. (2012). Changing epidemiology of clostridium difficile infection following the introduction of a national ribotyping-based surveillance scheme in England. *Clinical Infectious Diseases*, Vol. 55, pp. 1056–1063. https://doi.org/10.1093/cid/cis614

Wilson, K. H. (1983). Efficiency of various bile salt preparations for stimulation of Clostridium difficile spore germination. *Journal of Clinical Microbiology*, *18*(4), 1017–1019.

Wilson, Kenneth H, Kennedy, M. J., & Fekety, F. R. (1982). Use of Sodium Taurocholate to Enhance Spore Recovery on a Medium Selective for Clostridium difficile. In *JOURNAL OF CLINICAL MICROBIOLOGY* (Vol. 15).

Winston, J. A., & Theriot, C. M. (2016). Impact of microbial derived secondary bile acids on colonization resistance against Clostridium difficile in the gastrointestinal tract. *Anaerobe*, *41*, 44–50. https://doi.org/10.1016/j.anaerobe.2016.05.003

Woods, C., Humphreys, C. M., Rodrigues, R. M., Ingle, P., Rowe, P., Henstra, A. M., … Minton, N. P. (2019). A novel conjugal donor strain for improved DNA transfer into Clostridium spp. *Anaerobe*, *59*, 184–191. https://doi.org/10.1016/j.anaerobe.2019.06.020

Wydau-Dematteis, S., Louis, M., Zahr, N., Lai-Kuen, R., Saubaméa, B., Butel, M. J., & Pons, J. L. (2015). The functional dlt operon of Clostridium butyricum controls the d-alanylation of cell wall components and influences cell septation and vancomycin-induced lysis. *Anaerobe*, *35*, 105–114. https://doi.org/10.1016/J.ANAEROBE.2015.09.001

Xu, T., Li, Y., Shi, Z., Hemme, C. L., Li, Y., Zhu, Y., … Zhou, J. (2015). Efficient genome editing in clostridium cellulolyticum via CRISPR-Cas9 nickase. *Applied and Environmental Microbiology*, *81*(13), 4423–4431. https://doi.org/10.1128/AEM.00873-15

Yoon, S. H., Ha, S. min, Lim, J., Kwon, S., & Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek*, *110*(10), 1281–1286. https://doi.org/10.1007/S10482-017-0844-4

Yu, Q., Lepp, D., Gohari, I. M., Wu, T., Zhou, H., Yin, X., … 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*, *85*(6), 1–17. https://doi.org/10.1128/IAI.00975-16

Zeng, H., Umar, S., Rust, B., Lazarova, D., & Bordonaro, M. (2019). Secondary bile acids and short chain fatty acids in the colon: A focus on colonic microbiome, cell proliferation, inflammation, and cancer. *International Journal of Molecular Sciences*, *20*(5). https://doi.org/10.3390/ijms20051214

Zhang, J., Zong, W., Hong, W., Zhang, Z. T., & Wang, Y. (2018). Exploiting endogenous CRISPR-Cas system for multiplex genome editing in Clostridium tyrobutyricum and engineer the strain for high-level butanol production. *Metabolic Engineering*, *47*, 49–59. https://doi.org/10.1016/J.YMBEN.2018.03.007

Zhang, L., Dong, D., Jiang, C., Li, Z., Wang, X., & Peng, Y. (2015). *Molecular biology, genetics and biotechnology Insight into alteration of gut microbiota in Clostridium difficile infection and asymptomatic C. difficile colonization*. https://doi.org/10.1016/j.anaerobe.2015.03.008

Zhou, X., Wang, X., Luo, H., Wang, Y., Wang, Y., Tu, T., … Zhang, J. (2021). Exploiting heterologous and endogenous CRISPR-Cas systems for genome editing in the probiotic Clostridium butyricum. *Biotechnol Bioeng*, *118*, 2448–2459. https://doi.org/10.1002/bit.27753

Zhu, D., Sorg, J. A., & Sun, X. (2018). Clostridioides difficile biology: Sporulation, germination, and corresponding therapies for C. difficile infection. *Frontiers in Cellular and Infection Microbiology*, *8*(FEB). https://doi.org/10.3389/fcimb.2018.00029

Appendix

A.1 Plasmids

Thesis section	Plasmid name	Relevant Features	Use	Source
3.2.4	pMTL81151	Modular shuttle vector, Cm^R/Tm^R, ColE1+*traJ* replicon, MCS	General cloning and transfer into Clostridium.	Heap et al. (2009)
3.2.4	pMTL82151	Modular shuttle vector, Cm^R/Tm^R, ColE1+*traJ*/pBP1 replicons, MCS	General cloning and transfer into Clostridium.	Heap et al. (2009)
3.2.4	pMTL83151	Modular shuttle vector, Cm^R/Tm^R, ColE1+*traJ*/pCB102 replicons, MCS	General cloning and transfer into Clostridium.	Heap et al. (2009)
3.2.4	pMTL84151	Modular shuttle vector, Cm^R/Tm^R, ColE1+*traJ*/pCD6 replicons, MCS	General cloning and transfer into Clostridium.	Heap et al. (2009)
3.2.4	pMTL85151	Modular shuttle vector, Cm^R/Tm^R, ColE1+*traJ*/pIM13 replicons, MCS	General cloning and transfer into Clostridium.	Heap et al. (2009)
3.2.4	pMTL86151	Modular shuttle vector, Cm^R/Tm^R, ColE1+*traJ*/pIP404 replicons, MCS	General cloning and transfer into Clostridium.	James Millard, Nottingham
3.2.4	pMTL87151	Modular shuttle vector, Cm^R/Tm^R, ColE1+*traJ*/p19 replicons, MCS	General cloning and transfer into Clostridium.	James Millard, Nottingham
3.2.4	pMTL88151	Modular shuttle vector, Cm^R/Tm^R, ColE1+*traJ*/pUB110 replicons, MCS	General cloning and transfer into Clostridium.	James Millard, Nottingham
3.2.4	pMTL81151	Modular shuttle vector, Cm^R/Tm^R, ColE1+*traJ* replicon, MCS	General cloning and transfer into Clostridium.	Heap et al. (2009)
3.2.4	pMTL82151	Modular shuttle vector, Cm^R/Tm^R, ColE1+*traJ*/pBP1 replicons, MCS	General cloning and transfer into Clostridium.	Heap et al. (2009)
3.3.3	pRECas1_MCS	PMTL83151 vector containing *cas9* from *S. pyogenes* under the control of riboswitch P*fdx-E*, and an application specific module for insertion of relevant editing template and appropriate guide RNA, under the control of P*araE*.	Backbone to generate application specific plasmids for insertion/deletion by inducible CRISPR-Cas9.	Cañadas et al. (2019)
3.3.3	pRECas1_CD6_MCS	PMTL84151 vector containing *cas9* from *S. pyogenes* under the control of riboswitch P*fdx-E*, and an application specific module for insertion of relevant editing template and appropriate guide RNA, under the control of P*araE*.	Backbone for CRISPR-Cas9 applications in *P. hiranonis*	This study
3.3.3	pRECas1_MCS_B	pRECas1_CD6_MCS with sgRNA and HAs for *baiB* KO in *P. hiranonis*	CRISPR-Cas9 KO plasmid.	This study
3.3.3	pRECas1_MCS_CD	pRECas1_CD6_MCS with sgRNA and HAs for *baiCD* KO in *P. hiranonis*	CRISPR-Cas9 KO plasmid.	This study
3.3.3	pRECas1_MCS_E	pRECas1_CD6_MCS with sgRNA and HAs for *baiE* KO in *P. hiranonis*	CRISPR-Cas9 KO plasmid.	This study
3.3.3	pRECas1_MCS_A2	pRECas1_CD6_MCS with sgRNA and HAs for *baiA2* KO in *P. hiranonis*	CRISPR-Cas9 KO plasmid.	This study
3.3.3	pRECas1_MCS_F	pRECas1_CD6_MCS with sgRNA and HAs for *baiF* KO in *P. hiranonis*	CRISPR-Cas9 KO plasmid.	This study

3.3.3	pRECas1_MCS_G	pRECas1_CD6_MCS with sgRNA and HAs for *baiG* KO in *P. hiranonis*	CRISPR-Cas9 KO plasmid.	This study
3.3.3	pRECas1_MCS_H	pRECas1_CD6_MCS with sgRNA and HAs for *baiH* KO in *P. hiranonis*	CRISPR-Cas9 KO plasmid.	This study
3.3.3	pRECas1_MCS_CD2	pRECas1_CD6_MCS with sgRNA and HAs for *baiCD* KO in *P. hiranonis*	CRISPR-Cas9 KO plasmid.	This study
3.3.3	pRECas1_MCS_CD3	pRECas1_CD6_MCS with sgRNA and HAs for *baiCD* KO in *P. hiranonis*	CRISPR-Cas9 KO plasmid.	This study
3.3.4	pMTL_HZ13_Pfdx	ACE expression vector	Template for *Fdx* promoter.	Laboratory collection.
3.3.4	psbrcas9_630pyrE_ptcdb_PI	CRISPR KO vector	Template for *Thl* promoter	Laboratory collection.
3.3.4	pRECas2_MCS	PMTL83151 vector containing *cas9* from *S. pyogenes* under the control of riboswitch P*fdx-E*, and an application specific module for insertion of relevant editing template and appropriate guide RNA, under the control of P*j23119*.	Template for *j23119* promoter	Laboratory collection.
3.3.4	pMTL84151_FAST_neg	pMTL84151 vector with a promoterless *FAST*.	Reporter plasmid.	This study
3.3.4	pMTL84151_FAST_fdx	pMTL84151 vector with P*fdx* controlling expression of *FAST*.	Reporter plasmid.	This study
3.3.4	pMTL84151_FAST_fdxE	pMTL84151 vector with P*fdxE* controlling expression of *FAST*.	Reporter plasmid.	This study
3.3.4	pMTL84151_FAST_araE	pMTL84151 vector with P*araE* controlling expression of *FAST*.	Reporter plasmid.	This study
3.3.4	pMTL84151_FAST_thl	pMTL84151 vector with P*thl* controlling expression of *FAST*.	Reporter plasmid.	This study
3.3.4	pMTL84151_FAST_j23	pMTL84151 vector with P*j23119* controlling expression of *FAST*.	Reporter plasmid.	This study
3.3.4	pMTL8315_ptcdB_FAST	pMTL83151 with P*tcdB* controlling expression of *FAST*.	Template for *FAST*.	Chris Humphreys, Nottingham
3.4.1	pMTL84151_FAST_secG	pMTL84151 vector with P*secG* from *P. hiranonis* controlling expression of *FAST*.	Reporter plasmid.	This study
3.4.1	pMTL84151_FAST_thrS	pMTL84151 vector with P*thrS* from *P. hiranonis* controlling expression of *FAST*.	Reporter plasmid.	This study
3.4.1	pMTL84151_FAST_trxA	pMTL84151 vector with P*trxA* from *P. hiranonis* controlling expression of *FAST*.	Reporter plasmid.	This study
3.4.1	pMTL84151_FAST_fba	pMTL84151 vector with P*fba* from *P. hiranonis* controlling expression of *FAST*.	Reporter plasmid.	This study
3.4.1	pMTL84151_FAST_groS	pMTL84151 vector with P*groS* from *P. hiranonis* controlling expression of *FAST*.	Reporter plasmid.	This study
3.4.1	pMTL84151_FAST_pgk	pMTL84151 vector with P*pgk* from *P. hiranonis* controlling expression of *FAST*.	Reporter plasmid.	This study
3.4.2	pCHRE1	pRECas1_CD6_MCS with P*fdxE* exchanged for P*pgk* (*P. hiranonis*) and P*araE* exchanged for P*secG* (*P. hiranonis*).	Backbone for CRISPR-Cas9 applications using native promoters in *P. hiranonis*.	This study

3.4.2	pCHRE1_B	pCHRE1 with the guide RNA and homology arms for a *baiB* KO in *P. hiranonis.*	CRISPR-Cas9 KO plasmid.	This study
3.4.2	pCHRE1_CD	pCHRE1 with the guide RNA and homology arms for a *baiCD* KO in *P. hiranonis.*	CRISPR-Cas9 KO plasmid.	This study
3.4.2	pCHRE1_F	pCHRE1 with the guide RNA and homology arms for a *baiF* KO in *P. hiranonis.*	CRISPR-Cas9 KO plasmid.	This study
3.4.2	vFS40	*nCas9* under a theophylline inducible promoter.	Template for *ncas9*.	Francois Seys.
3.4.2	pCHnRE1	pCHRE1 with *ncas9*	Backbone for CRISPR-Cas9 applications in *P. hiranonis*	This study
3.4.2	pCHnRE1_CD	pCHnRE1 with sgRNA and HAs for baiCD KO in P. hiranonis	CRISPR-Cas9 KO plasmid.	This study
3.4.3	pMTL84151_PsecG_RBE	pMTL84151 vector with P*secG* from *P. hiranonis*, plus Riboswitch E, controlling expression of *FAST*.	Reporter plasmid.	This study
3.4.3	pMTL84151_PsecG_RBF	pMTL84151 vector with P*secG* from *P. hiranonis*, plus Riboswitch F, controlling expression of *FAST*.	Reporter plasmid.	This study
3.4.3	pMTL84151_PsecG_RBG	pMTL84151 vector with P*secG* from *P. hiranonis*, plus Riboswitch G, controlling expression of *FAST*.	Reporter plasmid.	This study
3.4.3	pMTL84151_Ppgk_RBE	pMTL84151 vector with P*pgk* from *P. hiranonis*, plus Riboswitch E, controlling expression of *FAST*.	Reporter plasmid.	This study
3.4.3	pMTL84151_Ppgk_RBF	pMTL84151 vector with P*pgk* from *P. hiranonis*, plus Riboswitch F, controlling expression of *FAST*.	Reporter plasmid.	This study
3.4.3	pMTL84151_Ppgk_RBG	pMTL84151 vector with P*pgk* from *P. hiranonis*, plus Riboswitch G, controlling expression of *FAST*.	Reporter plasmid.	This study
3.4.3	ML_Gapdh_RBE	Reporter plasmid with *gapd* plus Riboswitch E, controlling expression of *GFP*.	Template for Riboswitch E.	Matthew Lau, Nottingham
3.4.3	ML_Gapdh_RBF	Reporter plasmid with *gapd* plus Riboswitch F, controlling expression of *GFP*.	Template for Riboswitch F.	Matthew Lau, Nottingham
3.4.3	ML_Gapdh_RBG	Reporter plasmid with *gapd* plus Riboswitch G, controlling expression of *GFP*.	Template for Riboswitch G.	Matthew Lau, Nottingham
3.4.4	pCHRE1_RB	pRECas1_CD6_MCS with P*fdxE* exchanged for P*secG*RBE and P*araE* exchanged for P*pgk*.	Backbone for CRISPR-Cas9 applications using inducible native promoters in *P. hiranonis*.	This study
3.4.4	pCHRB_A	pCHRE1_RB with sgRNA and HAs for *baiA* KO in *P. hiranonis*	CRISPR-Cas9 KO plasmid.	This study
3.4.4	pCHRB_CD	pCHRE1_RB with sgRNA and HAs for *baiCD* KO in *P. hiranonis*	CRISPR-Cas9 KO plasmid.	This study
3.4.4	pCHRB_F	pCHRE1_RB with sgRNA and HAs for *baiF* KO in *P. hiranonis*	CRISPR-Cas9 KO plasmid.	This study
3.4.4	pCHRB_PE	pCHRE1_RB with sgRNA and HAs for *pyrE* KO in *P. hiranonis*	CRISPR-Cas9 KO plasmid.	This study
3.5.1	pMTL84151_barAB_FAST	pMTL84151 vector with *barAB* from *P. hiranonis* controlling expression of *FAST*.	Reporter plasmid.	This study
3.5.3	pMTL84151_CDcomp	pMTL84151 vector with *baiCD* from *P. hiranonis* under the expression of P*pgk*	Complementation expression plasmid	This study

4.2.4	pRECas_p19_MCS	PMTL87151 vector containing *cas9* from *S. pyogenes* under the control of riboswitch P*fdx-E*, and an application specific module for insertion of relevant editing template and appropriate guide RNA, under the control of P*araE*.	Backbone for CRISPR-Cas9 applications in *C. butyricum*	This study
4.2.4	pRECas_p19_pglx+	pRECas_p19_MCS with sgRNA and HAs for *pglx* complementation in *C. butyricum*	CRISPR-Cas9 insertion plasmid.	This study
4.2.4	pRECas_p19_pglx	pRECas_p19_MCS with sgRNA and HAs for *pglx* KO in *C. butyricum*	CRISPR-Cas9 KO plasmid.	This study
4.3.1	pRECas_p19_pyrE	pRECas_p19_MCS with sgRNA and HAs for *pyrE* KO in *C. butyricum*	CRISPR-Cas9 KO plasmid.	This study
4.3.1	pRECas_p19_purD	pRECas_p19_MCS with sgRNA and HAs for *purD* KO in *C. butyricum*	CRISPR-Cas9 KO plasmid.	This study
4.3.1	pRECas_p19_hisI	pRECas_p19_MCS with sgRNA and HAs for *hisI* KO in *C. butyricum*	CRISPR-Cas9 KO plasmid.	This study
4.3.3	pMTL82151_PErepair	pMTL82151 with repair fragment for *pyrE* truncation in *C. butyricum*	ACE repair vector	This study
4.3.3	pMTL82151_PDrepair	pMTL82151 with repair fragment for *purD* truncation in *C. butyricum*	ACE repair vector	This study
4.3.3	pMTL82151_HIrepair	pMTL82151 with repair fragment for *hisI* truncation in *C. butyricum*	ACE repair vector	This study
4.3.3	pMTL-YN2C	ACE complementation vector based on pMTL82151.	ACE complementation vector	Laboratory collection.
4.3.3	pMTL-YN2C_PE	ACE complementation vector for cargo insertion at *pyrE* in *C. butyricum*	ACE complementation vector	This study
4.3.3	pMTL-YN2C_PD	ACE complementation vector for cargo insertion at *purD* in *C. butyricum*	ACE complementation vector	This study
4.3.3	pMTL-YN2C_HI	ACE complementation vector for cargo insertion at *hisI* in *C. butyricum*	ACE complementation vector	This study
4.4.1	pMTL-HZ6	ACE complementation vector for insertion of lactose inducible *tcdR*/P*tcdB* system	Template for lactose inducible *tcdR*/P*tcdB* system	Laboratory collection.
4.4.1	pMTL87151_ ptcdB_FAST	pMTL87151 with FAST under the control of P*tcdB*	Reporter plasmid.	This study
4.4.1	pMTL-HZ13	ACE expression vector	Template for lactose inducible *tcdR*/P*tcdB* system	Laboratory collection
4.4.1	pMTL-YN2C_PE_ lactcdR	lactose inducible system for insertion at *pyrE* with ACE	ACE complementation vector	This study
4.4.1	pMTL_HZ13_ LHA	HZ13 with swapped LHA	ACE complementation vector	This study
4.4.1	pMTL_HZ13_ LHA_RHA	pMTL_HZ13_ LHA with swapped RHA	ACE complementation vector	This study
4.4.1	pMTL82151_lactcdR	pMTL82151 with *bgaR*/P*bgaL*.*tcdR*/P*tcdB*	Reporter plasmid.	This study
4.4.1	pMTL-CW21	pMTL vector with Himar1C9 under the control of the inducible P*tet* system	Template for P*tet*	Terry Bilverstone, Nottingham
4.4.1	pMTL87151_FAST_neg	pMTL87151 with a promoterless *FAST*.	Reporter plasmid.	This study
4.4.1	pMTL87151_FAST_araE	pMTL87151 vector with P*araE* controlling expression of *FAST*.	Reporter plasmid.	This study
4.4.1	pMTL87151_FAST_fdxE	pMTL87151 vector with P*fdxE* controlling expression of *FAST*.	Reporter plasmid.	This study
4.4.1	pMTL87151_FAST_tet	pMTL87151 vector with P*tet* controlling expression of *FAST*.	Reporter plasmid.	This study
4.4.2	pMTL82254	pMTL82251 with a promoterless *catP*.	Reporter plasmid.	Laboratory collection
4.4.2	pMTL82251	Modular shuttle vector, Erm^R, ColE1+t*raJ*/pBP1 replicons, MCS	General cloning and transfer into Clostridium.	Heap et al. (2009)

4.4.2	pMTL82251_CAT_araE	pMTL82251 vector with P*araE* controlling expression of *catP*	Reporter plasmid.	This study
4.4.2	pMTL-IC101	pMTL82251 vector with P*fdx* controlling expression of *catP*	Reporter plasmid.	Laboratory collection
4.4.2	pMTL-IC111-E	pMTL82251 vector with P*fdxE* controlling expression of *catP*	Reporter plasmid.	Laboratory collection
4.4.2	pMTL-HZ1	pMTL82251 vector with the inducible LAC system controlling expression of *catP*	Reporter plasmid.	Laboratory collection
4.4.2	pMTL-ICE1	pMTL82251 vector with RiboLac E controlling expression of *catP*	Reporter plasmid.	Laboratory collection
5.2.1	pSV1.0	pMTL82251 vector with *barB, barA, baiB, baiCD, baiE* from *C. scindens*	Shuttle vector	This study
5.2.1	pSV2.0	pMTL82251 vector with *baiA2, baiF, baiG, baiH, baiI* from *C. scindens*	Shuttle vector	This study
5.2.1	pCV1.0	pMTL87151 vector with *cas9* and sgRNA targeting upstream of *pyrE* in *C. butyricum*	Cas9 vector	This study
5.2.1	pCV2.0	pMTL87151 vector with *cas9* and sgRNA targeting BM8	Cas9 vector	This study
5.2.1	pSV2.1	pMTL82251 vector with *baiA2, baiF, baiG* (first half) from *C. scindens*	Shuttle vector	This study
5.2.1	pSV3.0	pMTL82251 vector with *baiG* (second half), *baiH, baiI* from *C. scindens*	Shuttle vector	This study
5.2.1	pCV3.0	pMTL87151 vector with *cas9* and sgRNA targeting BM4	Cas9 vector	This study
5.2.1	pSV2.2	pMTL82251 vector with *baiA2, baiF* (first half) from *C. scindens*	Shuttle vector	This study
5.2.1	pSV3.1	pMTL82251 vector with *baiF* (second half), *baiG, baiH, baiI* from *C. scindens*	Shuttle vector	This study
5.2.2	pMTL87151_baiG	pMTL87151 with *baiG* from *C. scindens*	Cloning	This study
5.2.2	pMTL87151_baiF	pMTL87151 with *baiF* from *C. scindens*	Cloning	This study
5.2.2	pMTL-YN2CPE_bBbA2	pMTL-YN2C_PE with *barB-baiA2* from *C. scindens*	ACE complementation vector	This study
5.3.1	pMTL87151_baiBCD	pMTL87151 with *baiB* and *baiCD* from *C. scindens*	Cloning	This study
5.3.1	pMTL87151_baiEAF	pMTL87151 with *baiE, baiA2 and baiF* from *C. scindens*	Cloning	This study
5.3.1	pMTL87151_baiGHI	pMTL87151 with *baiG, baiH and baiI* from *C. scindens*	Cloning	This study
5.3.1	pMTL-YN2CPD_fOID_ins3	pMTL-YN2C_PD with *baiG, baiH and baiI* from *C. scindens* under the control of P*fdx*OID	ACE complementation vector	This study
5.3.1	pMTL-YN2CHI_fdxE_ins3	pMTL-YN2C_HI with *baiG, baiH and baiI* from *C. scindens* under the control of P*fdxE*	ACE complementation vector	This study
5.3.2	pJ201_lacIQ_fdxOID	pJ201 expression vector	Template for P*fdx*OID	Kovacs et al 2013
5.3.2	pMTL-YN2CPE_baiG	pMTL-YN2C_PE with *baiG* from *C. scindens* under the control of P*fdxE*	ACE complementation vector	This study
5.3.2	pMTL-YN2CPD_aeBE1	pMTL-YN2C_PD with *baiB* from *C. scindens* under the control of P*araE*	ACE complementation vector	This study
5.3.2	pMTL-YN2CPD_aeBE2	pMTL-YN2C_PD with *baiB* and *baiCD* from *C. scindens* under the control of P*araE*	ACE complementation vector	This study

5.3.2	pMTL-YN2CPD_aeBE3	pMTL-YN2C_PD with *baiB, baiCD and baiE* from *C. scindens* under the control of P*araE*	ACE complementation vector	This study
5.3.2	pMTL-YN2CHI_aeBE4	pMTL-YN2C_HI with *baiJ* from *C. scindens* under the control of P*araE*	ACE complementation vector	This study
5.3.2	pMTL-YN2CPD_lacBE1	pMTL-YN2C_PD with *baiB* from *C. scindens* under the control of LAC	ACE complementation vector	This study
5.3.2	pMTL-YN2CPD_lacBE2	pMTL-YN2C_PD with *baiB* and *baiCD* from *C. scindens* under the control of LAC	ACE complementation vector	This study
5.3.2	pRECasp19_HIins	For insertion at *hisI* in *C. butyricum* by RiboCas	CRISPR-Cas9 insertion plasmid.	This study
5.3.2	pRECas_HIins_BE4	pRECasp19_HIins with *baiJ* under the control of P*araE*.	CRISPR-Cas9 insertion plasmid.	This study

A.2 Oligonucleotides

Thesis section	Primer name	Sequence	Plasmid/ Function
3.3.2	84_proto11_5	tcgacgtcacgcgtccatggagatcAATTTCAACGTTAAACACTTCTTTGAGTTTTTCAATTACTTatttaggcctgcagacatg caagcttggca	p84151_11-5
3.3.2	84_proto11_DR	tcgacgtcacgcgtccatggagatccaaatCAACGTTAAACACTTCTTTGAGTTTTTCAATTACTTatttaggcctgcagacatgcaa gcttggca	p84151_11-DR
3.3.2	84_proto23_5	tcgacgtcacgcgtccatggagatcACTACTAAACCAGAAACTGTAACTCTTAAAAATAATATAATatttaggcctgcagaca tgcaagcttggca	p84151_23-5
3.3.2	84_proto23_DR	tcgacgtcacgcgtccatggagatccaaatTAAACCAGAAACTGTAACTCTTAAAAATAATATAATatttaggcctgcagacatgc aagcttggca	p84151_23-DR
3.3.2	84_proto26_5	tcgacgtcacgcgtccatggagatcGATATGAAAATAAATCAAGAAATAGAAGCATTAAAAGCTCgcttaggcctgcagaca tgcaagcttggca	p84151_26-5
3.3.2	84_proto26_DR	tcgacgtcacgcgtccatggagatctaaatGAAAATAAATCAAGAAATAGAAGCATTAAAAGCTCgcttaggcctgcagacatgc aagcttggca	p84151_26-DR
3.3.3	sgRNA_Rev	ggccgacgtcataaaaataagaagcctgcaaatgcaggcttcttatttttataaaaaaagcaccgactcggtgccacttttcaagttg	sgRNAs
3.3.3	baiA2_LF3	atatatGACGTCaggtgttggtacaggtctaagc	pRECas1_MCS_A2
3.3.3	BaiA2_BM8_LR1	CCATTTTTTATCGGAGCTCCGGACgttcattgtcttatctccttctactttaattaagattttaaaag	pRECas1_MCS_A2
3.3.3	BaiA2_BM8_RF1	GTCCGGAGCTCCGATAAAAAATGGccatcataagatttactttaatttaaaactgtaattag	pRECas1_MCS_A2
3.3.3	baiA2_RR1	atatatGGCGCGCCatttatcatccatcctgtgaatcc	pRECas1_MCS_A2
3.3.3	baiB_LF1	atatatGACGTCataatagaaccatatccttgtttgacc	pRECas1_MCS_B
3.3.3	BaiB_BM8_LR1	CCATTTTTTATCGGAGCTCCGGACattcatatcggtattttgcctccc	pRECas1_MCS_B
3.3.3	BaiB_BM8_RF1	GTCCGGAGCTCCGATAAAAAATGGgtttgctaggatataaattcagttaactatctgc	pRECas1_MCS_B
3.3.3	baiB_RR1	atatatGGCGCGCCtagccatttctcctgtgtttatacg	pRECas1_MCS_B
3.3.3	baiCD_LF2	atatatGACGTCgctcctcatacagcagcatttaatgg	pRECas1_MCS_CD
3.3.3	BaiCD_BM8_LR1	CCATTTTTTATCGGAGCTCCGGACactcattttagacctcctaaatttacc	pRECas1_MCS_CD
3.3.3	BaiCD_BM8_RF1	GTCCGGAGCTCCGATAAAAAATGGagtatataatttaaaaattatataattatataaattaaaag	pRECas1_MCS_CD
3.3.3	baiCD_RR2	atatatGGCGCGCCtctgggctaactcttgagaatacg	pRECas1_MCS_CD
3.3.3	baiE_LF1	atatatGACGTCtgtacaagaaaccctcacttatgc	pRECas1_MCS_E
3.3.3	BaiE_BM8_LR1	CCATTTTTTATCGGAGCTCCGGACagtcattttattcgcctctttcttg	pRECas1_MCS_E

3.3.3	BaiE_BM8_RF1	GTCCGGAGCTCCGATAAAAAATGGgaaaaataataactgattgctaataaacaagatataaac	pRECas1_MCS_E
3.3.3	baiE_RR1	atatatGGCGCGCCacgcttacagttgtagctgttatacc	pRECas1_MCS_E
3.3.3	baiF_LF1	atatatGACGTCacagtggtacataaaagaaactggatatg	pRECas1_MCS_F
3.3.3	baiF_BM8_LR1	CCATTTTTTATCGGAGCTCCGGACagccattttatttctccttttctttttaaagaac	pRECas1_MCS_F
3.3.3	baiF_BM8_RF1	GTCCGGAGCTCCGATAAAAAATGGgataaataagaaacgtattaaataataaaatataaatgtcg	pRECas1_MCS_F
3.3.3	baiF_RR1	atatatGGCGCGCCctggctgagtagctctaac	pRECas1_MCS_F
3.3.3	baiG_LF1	atatatGACGTCatacggacaggttggagatcc	pRECas1_MCS_G
3.3.3	baiG_BM8_LR1	CCATTTTTTATCGGAGCTCCGGACtgtcatttcaattctccttgtataaaatttaagc	pRECas1_MCS_G
3.3.3	baiG_BM8_RF1	GTCCGGAGCTCCGATAAAAAATGGatagtattcacaataggtaacgtattaacagc	pRECas1_MCS_G
3.3.3	baiG_RR1	atatatggcgcgcccctcctatcattctttcatccc	pRECas1_MCS_G
3.3.3	baiH_LF1	atatatgacgtcgggattaatgcaggctaacatgac	pRECas1_MCS_H
3.3.3	baiH_BM8_LR1	CCATTTTTTATCGGAGCTCCGGACatccatgatgttgttttttcctttcg	pRECas1_MCS_H
3.3.3	baiH_BM8_RF1	GTCCGGAGCTCCGATAAAAAATGGttagttcctcagttcagagaatttgc	pRECas1_MCS_H
3.3.3	baiH_RR1	atatatggcgcgccgctattgcagctgcactc	pRECas1_MCS_H
3.3.3	sg_BaiB_F1	ttttcGTCGACctgctgtatgaggagcaccagttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct	pRECas1_MCS_B
3.3.3	sg_BaiCD_F1	ttttcGTCGACcattgtataggatgtgaccagttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct	pRECas1_MCS_CD
3.3.3	sg_BaiE_F1	ttttcGTCGACaaaaacgttggaataaacgggttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct	pRECas1_MCS_E
3.3.3	sg_BaiA2_F1	ttttcGTCGACattggagtgttaacaacaccgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct	pRECas1_MCS_A2
3.3.3	sg_BaiF_F1	ttttcGTCGACaaaaaacctgacaaccagaggttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct	pRECas1_MCS_F
3.3.3	sg_BaiG_F1	ttttcGTCGACttactagtaggtcttatagggttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct	pRECas1_MCS_G
3.3.3	sg_BaiH_F1	ttttcGTCGACtttatatggcaacgacactggttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct	pRECas1_MCS_H
3.3.4	sg_BaiCD2_F1	ttttcGTCGACggtgttggacgtataaacacgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct	pRECas1_MCS_CD2
3.3.4	sg_BaiCD3_F1	ttttcGTCGACgacgtagctgacttatcaaggttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct	pRECas1_MCS_CD3
3.3.4	ClProm_FAST_NdeI_F1	atatATCatATGGAACACGTAGCATTTGG	All Clostridial promoter + FAST plasmids

3.3.4	ClProm_FAST_AscI_R1	atatatggcgcgccTCATACCCTCTTAACGAAAACCCAGTAG	All Clostridial promoter + FAST plasmids
3.3.4	ClProm_Pthiol_NdeI_R2	atataCATATGtctaactaacctcctaaattttgatacggggtaacagataaaccatttcaatctatttcataagttccatagtttatccctaatttata	pMTL84151_FAST_thl
3.3.4	ClProm_Pthiol_NotI_F2	tatatGCGGCCGCtttttaacaaaatatattgataaaaataataatagtgggtataattaagttgttagagaaaacgtataaattagggataaactatgg	pMTL84151_FAST_thl
3.3.4	ClProm_ParaE_NotI_F1	atatatGCGGCCGCtttatatttagtcccttgccttgc	pMTL84151_FAST_araE
3.3.4	ClProm_ParaE_NdeI_R1	atatatCATATGgaaaactcctccttaagatttatatatgtgg	pMTL84151_FAST_araE
3.3.4	ClProm_PfdxE_NdeI_F1	atatatcatatgcttgttgttacctccttagcag	pMTL84151_FAST_fdxE
3.3.4	ClProm_PfdxE_NotI_R1	atatatgcggccgcgtgtagtagcctgtgaaataagtaag	pMTL84151_FAST_fdxE
3.3.4	ClProm_Pfdx_NotI_F1	atatatgcggccgcgtgtagtagcctgtgaaataagtaag	pMTL84151_FAST_fdx
3.3.4	ClProm_Pfdx_NdeI_R1	atatatcatatgtaacacacctccttaaaaattacacaac	pMTL84151_FAST_fdx
3.3.4	ClProm_Pj23119_NotI_F1	atatatGCGGCCGCttgacagctagctcagtcctaggtataata	pMTL84151_FAST_j23
3.3.4	ClProm_Pj23119_NdeI_R1	atatatCATATGtattatacctaggactgagctagctgtcaa	pMTL84151_FAST_j23
3.4.1	PsecG_FAST_Not_F1	atatatgcggccgctttgaaaaaatg	pMTL84151_FAST_secG
3.4.1	PthrS_FAST_Not_F1	atatatgcggccgccg	pMTL84151_FAST_thrS
3.4.1	PtrxA_FAST_Not_F1	atatatGCGGCCGCTCAG	pMTL84151_FAST_trxA
3.4.1	PtrxA_FAST_Xho_R1	atatatctcgagTCATACCCTCTTAACGAAAACC	pMTL84151_FAST_trxA
3.4.1	Pfba_hifi_F1	caggaaacagctatgaccgcggccgcttatagtaattttatagtgaataaaaataccg	pMTL84151_FAST_fba
3.4.1	Pfba_hifi_R1	cgtgttccataataaataccacctttttcttttttattatttttttc	pMTL84151_FAST_fba
3.4.1	FAST_fba_F1	ggtatttattatggaacacgtagcatttg	pMTL84151_FAST_fba
3.4.1	FAST_fba_R1	agcttgcatgtctgcaggcctcgagtcataccctcttaacgaaaac	pMTL84151_FAST_fba
3.4.1	PgroS_FAST_Not_F1	atatatgcggccgcgattaaaaac	pMTL84151_FAST_groS
3.4.1	Ppgk_FAST_Not_F1	atatatgcggccgcaatataaaacttg	pMTL84151_FAST_pgk
3.4.1	PrpoB_FAST_Not_F1	atatatgcggccgcatttaatgg	pMTL84151_FAST_rpoB
3.4.1	CHIRprom_FAST_Xho_R1	tatatactcgagTCATACCCTCTTAACG	All geneblock native *P. hiranonis* promoter plasmids
3.4.2	CHRE1_cas9_F1	caggaaacagctatgaccgcggccgctgtatccatatgaccatgattac	pCHRE1
3.4.2	CHRE1_cas9_R2	gaaattacttatggataagaaatactcaataggc	pCHRE1
3.4.2	CHRE1_Ppgk_F1	ccatctcgagaagtaatttctccttttatttgaag	pCHRE1
3.4.2	CHRE1_Ppgk_R1	ggatttttcgtctagatttaaaattgttaacctatatacactttataatataaaac	pCHRE1

3.4.2	CHRE1_PsecG_F1	taaatctagacgaaaaatcctccctttac	pCHRE1
3.4.2	CHRE1_PsecG_R1	gcgcgccattgacgtctatgtcgacactgttccctcctgtttatc	pCHRE1
3.4.2	pCHn_nCas9_F1	caggaaacagctatgaccgcggccgctcagtcacctcctagctg	pCHnRE1
3.4.2	pCHn_nCas9_R1	gaaattacttatggataagaaatactcaataggc	pCHnRE1
3.4.2	pCHn_Ppgk_F1	tcttatccataagtaatttctccttttatttgaag	pCHnRE1
3.4.2	pCHn_Ppgk_R1	ttttttcaaatctagatttaaaattgttaacctatatac	pCHnRE1
3.4.2	pCHn_PsecG_F1	ttaaatctagatttgaaaaaatgctatttaataaaaaagatag	pCHnRE1
3.4.2	pCHn_PsecG_R1	gcgcgccattgacgtctatgtcgacactgttccctcctgtttatc	pCHnRE1
3.4.3	PgkRB_prom_F1	ctcggtacccggggatcctctagagtcgacaatataaaacttgaatacatttaaaccttaatatg	pMTL84151_Ppgk_Riboswitch reporter plasmids
3.4.3	PgkRB_prom_R1	tgagtcgtattgtttatttgaagtatttttttaaaagtatgac	pMTL84151_Ppgk_Riboswitch reporter plasmids
3.4.3	PgkRBE_RBE_F1	tacttcaaataaacaatacgactcactatagg	pMTL84151_Ppgk_RBE
3.4.3	PgkRBE_RBE_R1	ctacgtgttccatcttgttgttacctccttag	pMTL84151_Ppgk_RBE
3.4.3	PgkRBE_FAST_F1	ggtaacaacaagatggaacacgtagcatttg	pMTL84151_Ppgk_RBE
3.4.3	PromRB_FAST_R1	cgagatctccatggacgcgtgacgtctataccctcttaacgaaaac	pMTL84151_PsecG/Ppgk_Riboswitch reporter plasmids
3.4.3	PgkRBF_RBF_F1	tacttcaaataaacaatacgactcactatagg	pMTL84151_Ppgk_RBF
3.4.3	PgkRBF_RBF_R1	ctacgtgttccattaagttacctccttagcag	pMTL84151_Ppgk_RBF
3.4.3	PgkRBF_FAST_F1	ggaggtaacttaatggaacacgtagcatttg	pMTL84151_Ppgk_RBF
3.4.3	PpgkRBG_RBG_F1	tacttcaaataaacaatacgactcactataggtg	pMTL84151_Ppgk_RBG
3.4.3	PpgkRBG_RBG_R1	ctacgtgttccattaacacacctccttagcag	pMTL84151_Ppgk_RBG
3.4.3	PpgkRBG_FAST_F1	ggaggtgtgttaatggaacacgtagcatttg	pMTL84151_Ppgk_RBG
3.4.3	SecGRB_prom_F1	ctcggtacccggggatcctctagagtcgactttgaaaaaatgctatttaataaaaaagatag	pMTL84151_PsecG_Riboswitch reporter plasmids
3.4.3	SecGRB_prom_R1	tgagtcgtattggtttatcgaatatattcttacattttag	pMTL84151_PsecG_Riboswitch reporter plasmids
3.4.3	SecGRBE_RBE_F1	atattcgataaaccaatacgactcactatagg	pMTL84151_PsecG_RBE
3.4.3	SecGRBE_RBE_R1	ctacgtgttccatcttgttgttacctccttag	pMTL84151_PsecG_RBE
3.4.3	SecGRBE_FAST_F1	ggtaacaacaagatggaacacgtagcatttg	pMTL84151_PsecG_RBE
3.4.3	SecGRBF_RBF_F1	atattcgataaaccaatacgactcactatagg	pMTL84151_PsecG_RBF

3.4.3	SecGRBF_RBF_R1	ctacgtgttccattaagttacctccttagcag	pMTL84151_PsecG_RBF
3.4.3	SecGRBF_FAST_F1	ggaggtaacttaatggaacacgtagcatttg	pMTL84151_PsecG_RBF
3.4.3	SecGRBG_RBG_F1	atattcgataaaccaatacgactcactataggtg	pMTL84151_PsecG_RBG
3.4.3	SecGRBG_RBG_R1	ctacgtgttccattaacacacctccttagcag	pMTL84151_PsecG_RBG
3.4.3	SecGRBG_FAST_F1	ggaggtgtgttaatggaacacgtagcatttg	pMTL84151_PsecG_RBG
3.4.4	pCHRB_pgK_F1	catttttttcaaaaatataaaacttgaatacatttaaaccttaatatg	pCHRE1_RB
3.4.4	pCHRB_pgK_R1	ggtcacatcctatacaatggtcgacaagtaatttctccttttatttgaag	pCHRE1_RB
3.4.4	pCHRB_sec_R1	gttttatatttttgaaaaaatgctatttaataaaaaagatagaaaatgtc	pCHRE1_RB
3.4.4	pCHRB_Sec_F1	tcttatccatcttgttgttacctccttag	pCHRE1_RB
3.4.4	pCHRB_cas9_F1	gaccgatcgggccccctgcaggtcagtcacctcctagctgactc	pCHRE1_RB
3.4.4	pCHRB_cas9_R1	taacaacaagatggataagaaatactcaatag	pCHRE1_RB
3.4.4	pCHRB_PE_sg_F1	ataaaaggagaaattacttgtcgaccttggaagcgaacttcaggagttttagagctagaaatagcaag	pCHRB_PE
3.4.4	pCHRB_PE_sg_R1	tagtaaaattgacgtcataaaaataagaagc	pCHRB_PE
3.4.4	pCHRB_PE_LHA_F1	ttatgacgtcaattttactaaaaaatgtttgcaaatatattaaattg	pCHRB_PE
3.4.4	pCHRB_PE_LHA_R1	ttatatattagtttgctgcacctcttac	pCHRB_PE
3.4.4	pCHRB_PE_RHA_F1	tgcagcaaactaatatataaatatattggggtgattg	pCHRB_PE
3.4.4	pCHRB_PE_RHA_R1	tttttagacttaagggcgggcgcgcctagaatatatcctcttagc	pCHRB_PE
3.5.1	84bar_bar_F1	tacccggggatcctctagagttatttacgatttttatctacatgagtgtgc	pMTL84151_barAB_FAST
3.5.1	84bar_bar_R1	cgtgttccatatcggtattttgcctccc	pMTL84151_barAB_FAST
3.5.1	84bar_FAST_F1	aaataccgatATGGAACACGTAGCATTTG	pMTL84151_barAB_FAST
3.5.1	84bar_FAST_R1	agcttgcatgtctgcaggcctcataccctcttaacgaaaacc	pMTL84151_barAB_FAST
3.5.3	CDcomp_prom_F1	tacccggggatcctctagagTTTAAAATTGTTAACCTATATACACTTTATAATATAAAACTTG	pMTL84151_CDcomp
3.5.3	CDcomp_prom_R1	cgtaactcatAAGTAATTTCTCCTTTTATTTGAAG	pMTL84151_CDcomp
3.5.3	CDcomp_bai_F1	gaaattacttATGAGTTACGACGCACTTTTTTC	pMTL84151_CDcomp
3.5.3	CDcomp_bai_R1	agcttgcatgtctgcaggccTTATATACTCATACCTACTTCGTAACCTTC	pMTL84151_CDcomp
4.2.4	pglXKO_RE1a_F1	ggcgcgccttgatatgaaaaatttaag	pRECas_p19_pglx
4.2.4	pglXKO_RE1a_R1	gagctcccgctgcttttaaatattttg	pRECas_p19_pglx
4.2.4	pglXKO_RE1b_F1	tttaaaagcagcgggagctccaagattcgtcaacgtaaataaatg	pRECas_p19_pglx

4.2.4	pglXKO_RE1b_R1	atagaatatcgtcgacgaaaactcctcc	pRECas_p19_pglx
4.2.4	pglXKO_LHA_F1	ttatttttatcagtcttcatttggctctg	pRECas_p19_pglx
4.2.4	pglXKO_LHA_R1	ccatttttttatcggagctccggacaaaatataaaaggtaagtgatttttgg	pRECas_p19_pglx
4.2.4	pglXKO_RHA_F1	gtccggagctccgataaaaaatggatccactcttcttcctcc	pRECas_p19_pglx
4.2.4	pglXKO_RHA_R1	ttttcatatcaaggcgcgcccatcaaaaggagagcagag	pRECas_p19_pglx
4.2.4	pglXKO_sg_F1	tttcgtcgacgatattctatagatgcagaagttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct	pRECas_p19_pglx
4.2.4	pglXKO_sg_R1	atgaagactgataaaaataagaagcctgcaaatgcaggcttcttatttttataaaaaaagcaccgactcggtgccacttttttcaagttg	pRECas_p19_pglx
4.2.4	PglxComp_sg_F1	tataaatcttaaggaggagttttcgTCCGGAGCTCCGATAAAAAAgttttagagctagaaatagcaagtt	pRECas_p19_pglx+
4.2.4	Pglxcomp_sg_R1	tatgttaaagataaaaataagaagcctgcaaatg	pRECas_p19_pglx+
4.2.4	Pglxcomp_pglx_F1	tcttatttttatctttaacatatcaagaaattgcag	pRECas_p19_pglx+
4.2.4	Pglxcomp_pglx_R1	ttaaatttttcatatcaaggtaccactctttcatcaatgaagg	pRECas_p19_pglx+
4.3.1	BUT_PyrEtrunc_LF2	atatatGACGTCatttgtcccatgcattgttcatttatac	pRECas_p19_pyrE
4.3.1	BUT_PyrETrunc_LR1	aaagagccaataatacaatatatagggaaaaataaggtgatgtactag	pRECas_p19_pyrE
4.3.1	BUT_PyrETrunc_RF1	tattgtattattggctctttaatataggcattgtttc	pRECas_p19_pyrE
4.3.1	BUT_PyrEtrunc_RR2	atatatGGCGCGCCatagtaggaattatttaataaaaaagtttgaagttaacg	pRECas_p19_pyrE
4.3.1	BUT_PyrEtrunc_sg_F1	ttttcGTCGACgatagaatggaaagaggccagttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct	pRECas_p19_pyrE
4.3.1	BUT_RE1_sgRNA_R1	ggccGACGTCataaaaataagaagcctgcaaatgcaggcttcttatttttataaaaaaagcaccgactcggtgccacttttttcaagttg	pRECas_p19_pyrE
4.3.1	BUT_PurDtrnc_sg_F1	ttttcGTCGACgctggagctaaattcgaagagttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct	pRECas_p19_purD
4.3.1	BUT_PurDtrnc_LF1	atatatgacgtcagtagacgcactacatatgttaatacag	pRECas_p19_purD
4.3.1	BUT_PurDtrnc_LR1	agtttatctttaatagaataaattataagaaacaggtttggtttatttaactaaac	pRECas_p19_purD
4.3.1	BUT_PurDtrnc_RF1	tattctattaaagataaactccttttttagttatcattattccaaag	pRECas_p19_purD
4.3.1	BUT_PurDtrnc_RR1	atatatGGCGCGCCctgcagaatatgctacatttacagatg	pRECas_p19_purD
4.3.1	BUT_HisItrnc_sg_F1	ttttcGTCGACttagtattggtagagcaaaagttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgct	pRECas_p19_hisI
4.3.1	BUT_HisItrnc_LF1	atatatgacgtcgttaacatccatggatgctgatg	pRECas_p19_hisI
4.3.1	BUT_HisItrnc_LR1	ttattaattagtctacctttatttccttaacatattgaaaatgtc	pRECas_p19_hisI
4.3.1	BUT_HisItrnc_RF1	aaaggtagactaattaataattcatgaatttataaatggaatgagcag	pRECas_p19_hisI
4.3.1	BUT_HisItrnc_RR1	atatatGGCGCGCCgcgatatcaactggcattataactc	pRECas_p19_hisI

4.3.3	BUT_PyrErepair_LF1	atatatCCTGCAGGtcacctatgattcttcctctaccataagc	pMTL82151_PErepair
4.3.3	BUT_PyrErepair_RR1	atatatGGCGCGCCcagacttggagaattttatgcagaagc	pMTL82151_PErepair
4.3.3	BUT_PurDrepair_LF1	atatatCCTGCAGGatgatacatctctactttgggagac	pMTL82151_PDrepair
4.3.3	BUT_PurDrepair_RR1	atatatGGCGCGCCttgtcattgaagaattccttgaagg	pMTL82151_PDrepair
4.3.3	BUT_HisIrepair_LF1	atatatCCTGCAGgacggtggatgaagttaagc	pMTL82151_HIrepair
4.3.3	BUT_HisIrepair_RR1	atatatGGCGCGCCgcttataatgtacatttgttgctattcc	pMTL82151_HIrepair
4.3.3	BUT_PyrEcomp_LR1	atatatgcggccgctacaatatatagggaaaaataaggtgatgtactagtgac	pMTL-YN2C_PE
4.3.3	BUT_PyrEcomp_RF1	atatagctagcttattttgcaccatattgagaatagtaatcatcaattc	pMTL-YN2C_PE
4.3.3	BUT_PurDcomp_LR1	tagcggccgctagaataaattataagaaacaggtttg	pMTL-YN2C_PD
4.3.3	BUT_PurDcomp_RF1	atatatgctagcttatttaaatatgcctatatcagttctacaataagaatc	pMTL-YN2C_PD
4.3.3	BUT_HisIcomp_LR1	tagcggccgcttaaacatttgttatcggctttc	pMTL-YN2C_HI
4.3.3	BUT_HisIcomp_RF1	atatatgctagcttaataattcatgaatttataaatggaatgagcag	pMTL-YN2C_HI
4.4.1	lacFAST_lac_Not_F1	atatatGCGGCCGCTAGCGCGGCCGCTAATTTAG	Unable to clone plasmid
4.4.1	lacFAST_lac_Nde_R1	atatatcatatgatcatcctctcttatatTTTAAAATAATTATGTATTCATGAAAC	Unable to clone plasmid
4.4.1	lacFAST_FAST_Nde_F1	atatatCatATGGAACACGTAGCATTTG	Unable to clone plasmid
4.4.1	lacFAST_FAST_XhoI_F1	atatatctcgagGGGTTTTCGTTAAGAGGGTATGA	Unable to clone plasmid
4.4.1	82_lactcdR_Sal_F1	atatatgtcgacTAATTTAGATATTAATTCTAAATTAAGTGAAATTAATATAGTAATTATATTAAAAGTCTAATTAAGAC	Unable to clone plasmid
4.4.1	82_lactcdR_AatII_R1	atatatgacgtcaggcgattaagttgggtaacg	Unable to clone plasmid
4.4.1	82_lactcdrHF_F1	caggaaacagctatgaccgcggccgcTAATTTAGATATTAATTCTAAATTAAGTGAAATTAATATAGTAATTATATTAAAAGTCTAATTAAGAC	Unable to clone plasmid
4.4.1	82_lactcdrHF_R1	agcttgcatgtctgcaggcctcgagATCTCCATGGACGCGTGAC	Unable to clone plasmid
4.4.1	87_lactcdrFAST_lac_Not_F1	atataGCGGCCGCTTTAGATATTAATTC	Unable to clone plasmid
4.4.1	87_lactcdrFAST_lac_Bsa_R1	tataggtctctgttcaaaattttctcctttactataatattttattg	Unable to clone plasmid
4.4.1	87_lactcdrFAST_FAST_Bsa_F1	tataggtctcaGAACACGTAGCATTTGGAAG	Unable to clone plasmid
4.4.1	87_lactcdrFAST_FAST_Xho_R1	atatatctcgagATATATAAAATAAATGTGCCTTAACATCTAAG	Unable to clone plasmid
4.4.1	PE_lactcdrHF_LHA_F1	gaccgatcgggccccctgcaggTCACCTATGATTCTTCCTC	Unable to clone plasmid

4.4.1	PE_lactcdrHF_LHA_F1	tagcggccgcTACAATATATAGGGAAAAATAAGGTG	Unable to clone plasmid
4.4.1	PE_lactcdrHF_lac_F1	atatattgtaGCGGCCGCTAATTTAGATATTAATTC	Unable to clone plasmid
4.4.1	PE_lactcdrHF_lac_F1	tggtgcaaaaTAACTCGAGATCTCCATGG	Unable to clone plasmid
4.4.1	PE_lactcdrHF_RHA_F1	tctcgagttaTTTTGCACCATATTGAGAATAG	Unable to clone plasmid
4.4.1	PE_lactcdrHF_RHA_F1	tttttagacttaagggcgggcgcgccCAGACTTGGAGAATTTTATGC	Unable to clone plasmid
4.4.1	PD_lactcdrHF_LHA_F1	gaccgatcgggccccctgcaggatgatacatctctactttgg	Unable to clone plasmid
4.4.1	PD_lactcdrHF_LHA_F1	tagcggccgctagaataaattataagaaacaggtttg	Unable to clone plasmid
4.4.1	PD_lactcdrHF_lac_F1	atttattctagcggccgctaatttagatattaattc	Unable to clone plasmid
4.4.1	PD_lactcdrHF_lac_F1	atttaaataactcgagatctccatggac	Unable to clone plasmid
4.4.1	PD_lactcdrHF_RHA_F1	gaccgatcgggccccctgcaggatgatacatctctactttgg	Unable to clone plasmid
4.4.1	PD_lactcdrHF_RHA_F1	tagcggccgctagaataaattataagaaacaggtttg	Unable to clone plasmid
4.4.1	HI_lactcdrHF_LHA_F1	gaccgatcgggccccctgcagggacggtggatgaagttaag	Unable to clone plasmid
4.4.1	HI_lactcdrHF_LHA_F1	tagcggccgcttaaacatttgttatcggctttc	Unable to clone plasmid
4.4.1	HI_lactcdrHF_lac_F1	aaatgtttaagcggccgctaatttagatattaattc	Unable to clone plasmid
4.4.1	HI_lactcdrHF_lac_F1	gaattattaagtcgactctagaggatcc	Unable to clone plasmid
4.4.1	HI_lactcdrHF_RHA_F1	tagagtcgacttaataattcatgaatttataaatggaatg	Unable to clone plasmid
4.4.1	HI_lactcdrHF_RHA_F1	ttaaatttttcatatcaaggcgcgccgcttataatgtacatttgttgc	Unable to clone plasmid
4.4.1	HZ13new_RHA_XhoI_F1	atatatctcgagttattttgcaccatattgagaatagtaatcat	pMTL_HZ13_ LHA_RHA
4.4.1	tcdR_scr_F1	atatatcatatgcaaaagtctttttatgaattaattgttttagcaagaaataactc	Screening for the LAC system
4.4.1	CW21_Tet_F1_Not	atatatgcggccgcttaagacccactttcacatttaagttg	pMTL87151_FAST_tet
4.4.1	CW21_tet_R1_Nde	tatatacatatgatgtatttcctcctcttcaatatatttaaggtc	pMTL87151_FAST_tet
5.2.1	CV2_BM8_R1	ggccGGCGCGCCataaaaataagaagcctgcaaatgcaggcttcttatttttataaaaaaagcaccgactcggtgccacttttttcaagttg	pCV2.0
5.2.1	CV2_BM8_F1	ttttcGTCGACTCCGGAGCTCCGATAAAAAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggc accgagtcggtgct	pCV2.0
5.2.1	CV1_guide_F1	ttttcGTCGACATAATAAACTAGAAGAATAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggc accgagtcggtgct	pCV1.0
5.2.1	CV1_guide_R1	ggccGGCGCGCCataaaaataagaagcctgcaaatgcaggcttcttatttttataaaaaaagcaccgactcggtgccacttttcaagttg	pCV1.0
5.2.1	SV1_LHA_F1	caggaaacagctatgaccgcggccgcttattttgcaccatattgagaatag	pSV1.0
5.2.1	SV1_LHA_R1	gaaatattaggttgaagatgttacaacagc	pSV1.0

5.2.1	SV1_BaiOp_F1	catcttcaacctaatatttctcaggttttcg	pSV1.0
5.2.1	SV1_BaiOp_BM8_R1	aaagctttgaCCATTTTTTATCGGAGCTCCGGACttatttgtgcatgttcatcg	pSV1.0
5.2.1	SV1_RHA_BM8_F1	gcacaaataatGTCCGGAGCTCCGATAAAAAATGGcaaagctttataaaaactttgttaaatc	pSV1.0
5.2.1	SV1_RHA_R1	agcttgcatgtctgcaggcctcgagaaaataaaaaatgaaagttataaatcataaaaac	pSV1.0
5.2.1	SV2_LHA_F1	caggaaacagctatgaccgcggccgcaccgattacttaaagagc	pSV2.0
5.2.1	SV2_LHA_R1	tacaatgttcttatttgtgcatgttcatcg	pSV2.0
5.2.1	SV2_BaiOp_F1	gcacaaataagaacattgtaaaagaaaggc	pSV2.0
5.2.1	SV2_BaiOp_R1	aaagctttgattaaaaatcacatgtatcccac	pSV2.0
5.2.1	SV2_RHA_F1	tgatttttaatcaaagctttataaaaactttgttaaatc	pSV2.0
5.2.1	SV2_RHA_R1	agcttgcatgtctgcaggcctcgagaaaataaaaaatgaaagttataaatcataaaaac	pSV2.0
5.2.1	CV3_BM4_R1	ggccGGCGCGCCataaaaataagaagcctgcaaatgcaggcttcttatttttataaaaaaagcaccgactcggtgccacttttcaagttg	pCV3.0
5.2.1	CV3_BM4_F1	ttttcGTCGACAGGGTTGTGGGTTGTACGGAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggca ccgagtcggtgct	pCV3.0
5.2.1	SV2.1_LHA_F1	caggaaacagctatgaccgcaccgattacttaaagagc	pSV2.1
5.2.1	SV2.1_LHA_R1	tacaatgttcttatttgtgcatgttcatcg	pSV2.1
5.2.1	SV2.1_BaiOp_F1	gcacaaataagaacattgtaaaagaaaggcaggagtaagattatg	pSV2.1
5.2.1	SV2.1_BaiOp_BM4_R1	aaagctttgaCCTTCCGTACAACCCACAACCCTCgaaaagcgccaccagcgc	pSV2.1
5.2.1	SV2.1_RHA_BM4_F1	ggcgcttttcGAGGGTTGTGGGTTGTACGGAAGGtcaaagctttataaaaactttgttaaatc	pSV2.1
5.2.1	SV2.1_RHA_R1	agcttgcatgtctgcaggcctcgagaaaataaaaaatgaaagttataaatcataaaaac	pSV2.1
5.2.1	SV3_LHA_F1	caggaaacagctatgaccgcggccgctccagcggcgccgttgtg	pSV3.0
5.2.1	SV3_LHA_R1	ttaccaggatgaaaagcgccaccagcgc	pSV3.0
5.2.1	SV3_BaiOp_F1	ggcgcttttcatcctggtaatggtagaaaag	pSV3.0
5.2.1	SV3_BaiOp_R1	aaagctttgattaaaaatcacatgtatcccac	pSV3.0
5.2.1	SV3_RHA_F1	tgatttttaatcaaagctttataaaaactttgttaaatc	pSV3.0
5.2.1	SV3_RHA_R1	agcttgcatgtctgcaggcctcgagaaaataaaaaatgaaagttataaatcataaaaac	pSV3.0
5.2.1	SV2.2_LHA_F1	caggaaacagctatgaccgcggccgcaccgattacttaaagagc	pSV2.2
5.2.1	SV2.2_LHA_R1	tacaatgttcttatttgtgcatgttcatcg	pSV2.2
5.2.1	SV2.2_BaiOp_F1	gcacaaataagaacattgtaaaagaaaggcag	pSV2.2

5.2.1	SV2.2_RHA_BM4_F1	atgatccagtGAGGGTTGTGGGTTGTACGGAAGGtcaaagctttataaaaactttgttaaatc	pSV2.2
5.2.1	SV2.2_BaiOp_BM4_R1	aaagctttgaCCTTCCGTACAACCCACAACCCTCactggatcatacgtccgtc	pSV2.2
5.2.1	SV2.2_RHA_R1	agcttgcatgtctgcaggcctcgagaaaataaaaaatgaaagttataaatcataaaaac	pSV2.2
5.2.1	SV3.1_LHA_F1	caggaaacagctatgaccgcggccgctaaatcctaagattgccatcg	pSV3.1
5.2.1	SV3.1_LHA_R1	tctgtggcgtactggatcatacgtccgtc	pSV3.1
5.2.1	SV3.1_BaiOp_F1	atgatccagtacgccacagacggcgtga	pSV3.1
5.2.1	SV3.1_BaiOp_R1	aaagctttgattaaaaatcacatgtatcccactcttctggcatc	pSV3.1
5.2.1	SV3.1_RHA_F1	tgatttttaatcaaagctttataaaaactttgttaaatc	pSV3.1
5.2.1	SV3.1_RHA_R1	agcttgcatgtctgcaggcctcgagaaaataaaaaatgaaagttataaatcataaaaac	pSV3.1
5.3.2	PE_fdxE_G_prom_F1	tacccggggatcctctagagtcgacataaaaaaattgtagataaattttataaaatagttttatc	pMTL-YN2CPE_baiG
5.3.2	PE_fdxE_G_prom_R1	cggtgctcatcttgttgttacctccttag	pMTL-YN2CPE_baiG
5.3.2	PE_fdxE_G_bai_F1	taacaacaagatgagcaccgtagccaatc	pMTL-YN2CPE_baiG
5.3.2	PE_fdxE_G_bai_R1	agcttgcatgtctgcaggcctcgagttatgcctctttcttctgatagattc	pMTL-YN2CPE_baiG
5.3.2	PD_LAC_BE_prom_F1	cctgtttcttataatttattctagcataaaaaaattgtagataaattttataaaatagttttatctac	pMTL-YN2CPD_lacBE1/BE2
5.3.2	PD_LAC_BE_prom_R1	ttttttgtgcatatgaccctcccaatacatttaaaataattatg	pMTL-YN2CPD_lacBE1/BE2
5.3.2	PD_LAC_BE1_bai_F1	ttgggagggtcatatgcacaaaaaatcaacgtgtgagag	pMTL-YN2CPD_lacBE1
5.3.2	PD_LAC_BE1_bai_R1	tgccaagcttgcatgtctgcaggcctcataccccgcgggcaat	pMTL-YN2CPD_lacBE1
5.3.2	PD_LAC_BE_bai_F1	ttgggagggtcatatgcacaaaaaatcaacgtg	pMTL-YN2CPD_lacBE2
5.3.2	PD_LAC_BE2_bai_R1	tgccaagcttgcatgtctgcaggccctagattgccattcctgc	pMTL-YN2CPD_lacBE2
5.3.2	PD_LAC_BE3_bai_R1	tgccaagcttgcatgtctgcaggccttatttgtgcatgttcatcg	Unable to clone plasmid
5.3.2	HI_LAC_BE4_prom_F1	ccgataacaaatgtttaagcataaaaaaattgtagataaattttataaaatagttttatctac	Unable to clone plasmid
5.3.2	HI_LAC_BE4_prom_R1	aacttgccatatgaccctcccaatacatttaaaataattatg	Unable to clone plasmid
5.3.2	HI_LAC_BE4_bai_F1	ggagggtcatatggcaagttatacaccc	Unable to clone plasmid
5.3.2	HI_LAC_BE4_bai_R1	agcttgcatgtctgcaggccttacagcatctctctctgg	Unable to clone plasmid
5.3.2	PD_araE_BE_prom_F1	ttcttataatttattctagcataaaaaaattgtagataaattttataaaatagttttatc	pMTL-YN2CPD_aeBE1/BE2/BE3
5.3.2	PD_araE_BE_prom_R1	ttttgtgcatatggaaaactcctccttaag	pMTL-YN2CPD_aeBE1/BE2/BE3

5.3.2	PD_araE_BE1_bai_F1	agttttccatatgcacaaaaaatcaacgtgtgag	pMTL-YN2CPD_aeBE1
5.3.2	PD_araE_BE1_bai_R1	agcttgcatgtctgcaggcctataccccgcgggcaat	pMTL-YN2CPD_aeBE1
5.3.2	PD_araE_BE_bai_F1	agttttccatATGCACAAAAAATCAACGTG	pMTL-YN2CPD_aeBE2
5.3.2	PD_araE_BE2_bai_F1	agcttgcatgtctgcaggccCTAGATTGCCATTCCTGC	pMTL-YN2CPD_aeBE2
5.3.2	PD_araE_BE3_bai_F1	agcttgcatgtctgcaggccTTATTTGTGCATGTTCATCG	pMTL-YN2CPD_aeBE3
5.3.2	HI_araE_BE4_prom_F1	ccgataacaaatgtttaagcATAAAAAAATTGTAGATAAATTTTATAAAATAGTTTTATC	pMTL-YN2CHI_aeBE4
5.3.2	HI_araE_BE4_prom_R1	aacttgccatATGGAAAACTCCTCCTTAAG	pMTL-YN2CHI_aeBE4
5.3.2	HI_araE_BE4_bai_F1	agttttccatATGGCAAGTTATACACCC	pMTL-YN2CHI_aeBE4
5.3.2	HI_araE_BE4_bai_R1	agcttgcatgtctgcaggccTTACAGCATCTCTCTCTGg	pMTL-YN2CHI_aeBE4
5.3.2	RE1_HIA_guide_F1	tataaatcttaaggaggagttttcgtgtacatcaataaatcttaagttttagagctagaaatagcaagtt	pRECasp19_HIins
5.3.2	RE1_HIA_guide_R1	aatactaaaatagataaaaataagaagcctgcaaatg	pRECasp19_HIins
5.3.2	RE1_HIA_LHA_F1	tcttatttttatctattttagtattggtagagcaaaag	pRECasp19_HIins
5.3.2	RE1_HIA_LHA_R1	ttatatgacgtcataaaaagaaaatggaatgagc	pRECasp19_HIins
5.3.2	RE1_HIA_RHA_F1	attttctttttatgacgtcatataaagctcaggaaagaaag	pRECasp19_HIins
5.3.2	RE1_HIA_RHA_R1	attccttaaatttttcatatcaaggcacaaagtatcataatatcttctc	pRECasp19_HIins
5.3.2	RE1_HIB_MCS_F1	tcattccattttctttttatgacgtcgccattcgccattcaggc	pRECasp19_HIins
5.3.2	RE1_HIB_MCS_R1	ttctttcctgagctttatatgacgtgcggccgcactagtctgc	pRECasp19_HIins
5.3.2	Hiins_aeJ_F1	atctccatggacgcgtgacgTTTATATTTAGTCCCTTGCCTTG	pRECas_HIins_BE4
5.3.2	Hiins_aeJ_R1	tacccggggatcctctagagTTACAGCATCTCTCTCTGgc	pRECas_HIins_BE4
5.4.2	cD_recA_F1	GCC CTT AGT ATA GCA GCA GAA G	cDNA screening of *recA* in *C. butyricum*
5.4.2	cD_recA_R1	TCA GGT GAA CCG AAC ATT ACT C	cDNA screening of *recA* in *C. butyricum*
5.4.2	cD_baiG_F1	GCA CCG TAG CCA ATC CTA AT	cDNA screening of *baiG* (*C. scindens*)
5.4.2	cD_baiG_R1	GCT GTT CCG ATA CCT ACG ATA AA	cDNA screening of *baiG* (*C. scindens*)
5.4.2	cD_baiB_F1	CAG CTG GCA TGG TAC CTT ATA G	cDNA screening of *baiB* (*C. scindens*)
5.4.2	cD_baiB_R1	GCT TCT GAT CGT CTC ATC ATC C	cDNA screening of *baiB* (*C. scindens*)

5.4.2	cD_baiCD_F1	GTT CAG GCA ATC CGG TCT AAT A	cDNA screening of *baiCD (C. scindens)*
5.4.2	cD_baiCD_R1	CCC TGG TAC TCT AAG CAC AAT C	cDNA screening of *baiCD* (*C. scindens*)
5.4.2	cD_baiE_F1	GAC GGA AAG ATG TGG GAT GAA	cDNA screening of *baiE* (*C. scindens*)
5.4.2	cD_baiE_R1	GAT GTA CCA CTG GCC TTC TAT TT	cDNA screening of *baiE* (*C. scindens*)
5.4.2	cD_baiJ_F1	GGG CAT GAG GTA CAG GAT TTA G	cDNA screening of *baiJ* (*C. scindens*)
5.4.2	cD_baiJ_R1	CCA ACG ACT CTT CCA TCT TCT C	cDNA screening of *baiJ* (*C. scindens*)
Screening	ColE1_tra_F1	CTGTGGATAACCGTATTACC	pMTL vectors
	ColE1_tra_F2	CCATCAAGAAGAGCGAC	pMTL vectors
	pCD6_R1	GACTTTAAGCCTACGAATACC	pMTL vectors
	p19_R1	gatcacaatatgcaaagttgtcc	pMTL vectors
	sgRNA_F4	CCTTTCATTTACAATTCATACG	pMTL vectors
	pCD6_F1	GGAGTTTGAACCAATATTGG	pMTL vectors
	catP_R1	CAAGTTTATCGCTCTAATGAAC	pMTL vectors
	pCB102_R1	CTGTTATGCCTTTTGACTATC	pMTL vectors
	pBP1_R1	CTTCATTAAATGCCTTAGAATC	pMTL vectors
	M13_F	tgtaaaacgacggccagt	pMTL vectors
	MCS_F1	CTTCCCAACAGTTGCGC	pMTL vectors
	Cas9scr1R	gaaacttaatcatatgcgctaagg	pRECas1 vectors
	Cas9scr1	atggataagaaatactcaataggcttag	pRECas1 vectors
	CHIR_PEscr_F1	aatagctgaccagcttgg	*pyrE P. hiranonis*
	CHIR_PEscr_R1	ggaacagatgttctatagaagaac	*pyrE P. hiranonis*
	Ch_scr_F1	acaataaatcaaaagagaattgattttaaaagagg	*P hiranonis bai* operon
	Ch_scr_F2	acctagaaatttaagacctgaaagtatagttag	*P hiranonis bai* operon
	Ch_scr_F3	caatggaaaggtaaacagagataagatg	*P hiranonis bai* operon
	Ch_scr_F4	gttgactgcatagatgagttaatgg	*P hiranonis bai* operon

Ch_scr_F5	tccagctataccaggtatagatagtg	*P hiranonis bai* operon
Ch_scr_F6	aagaagtaactgaatatttagcagcag	*P hiranonis bai* operon
Ch_scr_F7	aacgttacaggtgtattcaatgg	*P hiranonis bai* operon
Ch_scr_F8	gtacgacagactaggtctttctg	*P hiranonis bai* operon
Ch_scr_F9	cccttctgaaatatggagaggtg	*P hiranonis bai* operon
Ch_scr_F10	cttgtggtggattaataatgcagg	*P hiranonis bai* operon
Ch_scr_F11	ttcgtaccactatttgcaaacac	*P hiranonis bai* operon
Ch_scr_F12	gggtaatacaactgctttagattctg	*P hiranonis bai* operon
Ch_scr_F13	ccatgtgcaagctgtgg	*P hiranonis bai* operon
Ch_scr_R1	agtcatctgagggtttataactgc	*P hiranonis bai* operon
Ch_scr_R2	tgcactggtaagcagcc	*P hiranonis bai* operon
Ch_scr_R3	gaagtgtataactgtagcaccac	*P hiranonis bai* operon
Ch_scr_R4	aagtatgtctctgttgtccatacc	*P hiranonis bai* operon
Ch_scr_R5	ccataggtcctatgaatactgaacc	*P hiranonis bai* operon
Ch_scr_R6	tctaatccttcgtactgagtacctc	*P hiranonis bai* operon
Ch_scr_R7	ggaattgatgtctgtgcacc	*P hiranonis bai* operon
CHIR_baiop_scr_F1	tcaagtggctcatcagattcac	*P hiranonis bai* operon
CHIR_baiop_scr_R1	aattgtttctccagcttttggc	*P hiranonis bai* operon
BUT_pglxscr_F1	gcataataacagttgctctatcagtagag	*pglx C. butyricum*
BUT_pglxscr_R1	gtataggctgtaaatatccagggg	*pglx C. butyricum*
pCBU2_F1	gcatgatcttcttcgaagagatg	*pCB102* in *C. butyricum*
pCBU2_R1	ctctgctagatcatctagctcc	*pCB102* in *C. butyricum*
BUT_pyrE_F1	tggaacaattaactttataaatccaatggc	*pyrE* in *C. butyricum*
BUT_pyrE_R3	tgtcaactccgtataaagcaactc	*pyrE* in *C. butyricum*
BUT_pyrE_R2	ccatcccattgtttaaacataattgtttc	*pyrE* in *C. butyricum*
BUT_PurD_F1	tatttttatttaatacgtatatgaggagtattaatatgc	*purD* in *C. butyricum*
BUT_PurD_R1	agttaattattaagaaggaagttctaattgg	*purD* in *C. butyricum*
BUT_HisI_F1	agatgcagctttagcagc	*hisI* in *C. butyricum*

BUT_HisI_R1	ctgagctttatatagaatgtatcaattcc	*hisI* in *C. butyricum*
Univ-0027-F	GCGAGAGTTTGATCCTGGCTCAG	16S sequencing
Univ-1492-R	CGCGGTTACCTTGTTACGACTT	16S sequencing
BUT_HIscr_F1	caatatgttaaggaaataaaggtagactgtg	*his I* in *C. butyricum* (RiboCas insertions)
BUT_HIscr_R1	ggaaatcacattgtgaaccaactg	*his I* in *C. butyricum* (RiboCas insertions)