The Genetic Domestication of *Clostridium carboxidivorans*

William Morris

Thesis Submitted to the University of Nottingham for the Degree of Doctor of Philosophy

October 2024

Abstract

The continued use of fossil fuels to produce energy and chemicals is becoming increasingly untenable. They are a finite resource responsible for the release greenhouse gases into the atmosphere, directly contributing to global warming and climate change. If the consumption of fossil fuels remains unabated then drastic environmental consequences are inevitable. A promising alternative is exploiting the potential of microorganisms as biological catalysts to produce renewable energy and chemicals. To this end, autotrophic bacteria that are capable of fixing inorganic carbon in the form of CO_2 and CO are of particular interest. *Clostridium carboxidivorans* is one such bacterium, an anaerobic acetogen capable of producing ethanol, butanol, hexanol, and their conjugate organic acids from CO or CO_2 and H_2. However, *C. carboxidivorans* possesses an expansive Restriction Modification System (RMS) rendering DNA transfer impossible.

In this study, the RMS of *C. carboxidivorans* is bypassed to deliver CRISPR-Cas9 vectors targeting RMS-associated nuclease-encoding genes to create a fully genetically domesticated strain, *C. carboxidivorans* Δ7RM. Then, a metabolic engineering approach is adopted whereby the gene *hytA* of the domesticated strain is deleted to increase autotrophic growth and ethanol, butanol, and hexanol production. Finally, an alternative Retrotransposition Activated Marker is developed for the ClosTron mutagenesis system. This is characterised in several members of *Clostridium*, including *C. carboxidivorans* Δ7RM.

COVID-19 Statement

The first lockdown saw the closure of the University while I was halfway through my third rotation, which I would then choose as my PhD project. I was unable to access the labs from 18/03/2020 to 10/10/2020 and was therefore unable to perform any research and produce data for my doctorate. Whilst I was able to spend this time doing background reading and preparing my literature review there was very limited scope for anything else, such as bioinformatics work, given that my project involves producing knockouts of specific genes for which wet lab work is essential.

Upon returning to the labs in November 2020 a rota system was in place due to laboratories and offices having reduced occupancy to follow covid guidelines which lasted until mid/late 2021. This followed strict alternating shift pattern of 06:00-14:00 and 14:30-22:00 which resulted in a reduction in productivity due to irregular and limited working hours where some experiments and protocols would have to be cut short during the day if they ran over time. It can be particularly hard to mitigate when working in microbiology as some strains can take a long time to grow and can behave unpredictably.

Covid restrictions also meant that I could only have meetings with my primary supervisor remotely through Microsoft Teams which restricted my ability to properly plan my research collaboratively. I found that remote discussions create friction hampering the “idea creating” process when compared to in-

person meetings as well as making any impromptu discussions that would happen day-to-day in a normal working environment impossible. Additionally, my access to individuals in the group with specialist knowledge on specific techniques of molecular/microbiology was limited due to only half the group having access to the building at the same time as I was.

Acknowledgements

I would like to acknowledge my supervisors Prof Nigel Minton, Dr Christopher Humphreys, and Prof Philippe Soucaille for having given me the opportunity to undertake this PhD. I would also like to thank them for their invaluable guidance and mentorship during my studies.

This work was funded by the BBSRC and for that I would like to thank them.

My gratitude also extends to my colleagues at the Synthetic Biology Research Centre and the friends I have made at the University of Nottingham, without whose support this thesis would not be possible.

I am also grateful for the encouragement of my friends, family, and partner whose belief in me has never wavered, providing a source of strength for me when I need it most.

List of Contents

Abbreviations

ACS	Acetyl CoA-synthase
ADP	Adenosine diphosphate
ALE	Adaptive Laboratory Experiment
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
aa	Amino acid
Ack	Acetate kinase
Adhe	Bifunctional acetaldehyde/alcohol dehydrogenase
Aor	Aldehyde:ferredoxin oxidoreductase
Bcd	Butyryl-CoA dehydrogenase
BLAST	Basic Local Alignment Search Tool
bp	Base pair
Buk	Butyrate kinase
CFU	Colony Forming Units
Cat	Chloramphenicol acetyltransferase
Cla^R	Clarithromycin resistant
Cm	Chloramphenicol
Cm^S	Chloramphenicol sensitive
Cm^R	Chloramphenicol resistant
CoA	Co-enzyme A
Codh	Carbon monoxide dehydrogenase
CoFeSP	Cobalt and iron containing corrinoid sulfur protein
cPCR	Colony PCR
CRISPR	Clustered Regularly Spaced Short Palindromic Repeats
CrRNA	CRISPR RNA
Crt	Crotonase
dH_2O	distilled H_2O

DNA	Deoxyribonucleic acid
DNP	DNA polymerase
DSMZ	German Collection of Microorganisms and Cell Cultures
EDTA	Ethylenediaminetetraacetic acid
Em	Erythromycin
Em^R	Erythromycin resistant
Em^S	Erythromycin sensitive
FTC	Formyl-THF Cyclohydrase
Fak	Fatty acid kinase
Fd	Oxidised ferredoxin
Fd^{2-}	Reduced ferredoxin
Fdh	Formate dehydrogenase
FTS	Formyl-THF Synthase
GSRP	Glycine synthase-reductase pathway
GTP	Guanosine-5'-triphosphate
Hbd	3-hydroxybutyryl-CoA dehydrogenase
IEP	Intron-Encoded Protein
KO	Knock Out
LCB	Lignocellulosic biomass
MATE	Multidrug and Toxic Compound Extrusion
MES	2-(N-morpholino)ethanesulfonic acid
MTR	Methylene-THF Reductase
MTD	Methylene-THF Dehydrogenase
NADP	Nicotinamide adenine dinucleotide phosphate
NEB	New England Biolabs
OD	Optical Density
PAM	Protospacer Adjacent Motif
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction

PHA	Polyhydroxyalkanoate
PHB	poly-β-hydroxybutyrate
P^i	Inorganic phosphate
Pta	Phosphotransacetylase
Ptb	Phosphotransbutyrylase
Ptf	Phosphotransferase
RAM	Retrotransposition-Activated Marker
REBASE	Restriction Enzyme Database
RMS	Restriction Modification System
RPM	Revolutions Per Minute
SNP	Single Nucleotide Polymorphism
SMRTseq	Single-Molecule Real-Time sequencing
TIS	Transposon Insertion Sequencing
Tc	Tetracycline
THF	Tetrahydrofolate
Tm	Thiamphenicol
Tm^R	Thiamphenicol resistant
Tm^S	Thiamphenicol sensitive
TracrRNA	Trans-activating crRNA
WT	Wild type
WLP	Wood-Ljungdahl Pathway

List of Figures

List of Tables

Chapter 1 – Introduction

1.1. Modern biotechnology and its model Organisms

To meet the growing demand for the antibiotic penicillin brought on by the second world war, the US department of agriculture arranged a meeting with the pharmaceutical companies Merck, Squibb, Pfizer, and Lederle in 1941. The purpose of this meeting was to establish a cooperative project with the primary objective of improving penicillin production through strain development and process engineering (Buchholz & Collins, 2013). What followed was a large-scale and laborious effort of screening many hundreds of different strains of penicillin-producing microorganisms from sources all over the world. The most promising candidate fungal strain, NRRL 1951 (*Penicillium rubens*), was subjected to random mutagenesis via rounds of X-ray and ultraviolet radiation with subsequent selection of mutants displaying a high penicillin yielding phenotype (Quinn, 2013). This approach would be emblematic of how strain engineering was performed before the development of molecular cloning and genetic modification. However, what was to come would radically alter the landscape of biological research and sow the seeds for a new age of biotechnology.

Thirty years later the first example of introducing plasmid DNA into *Escherichia coli* through bacterial transformation was demonstrated (Cohen *et al.*, 1972) and was quickly applied when the first recombinant DNA molecules generated

in vitro were constructed to prove that the DNA within these constructs were functional and possessed the genetic properties of the DNA fragments they were originally made from (Cohen *et al.*, 1973). This meant, because of the universal nature of DNA, any genetically encoded product could theoretically be manufactured on an industrial scale through the fermentation of modified microorganisms. However vast the potential, this discovery was limited by the difficulties imposed by potential recombinant genes being limited to those with preexisting flanking restriction sites and a sparse number of non-standardised plasmids. Moreover, there was no efficient method for screening transformants on selection plates meaning that, although antibiotic selection would remove non-transformants, plasmids that have self-ligated, and therefore contain no insert, would be difficult to filter out. It would not be until the following decade that development of the Polymerase Chain Reaction (PCR) (Kaunitz, 2015) would allow for new cloning techniques to insert any gene into a plasmid, even with a limited knowledge of its sequence, through the addition of restriction recognition sites to the 5' end of primer pairs (Hoseini & Sauer, 2015) and for the screening of transconjugants via colony PCR (Bergkessel & Guthrie, 2013).

The innovations of molecular cloning, PCR, and commercially available automated sequencing machines set the stage for the modern age of biotechnology. This would lead to contemporary technologies that allow for the direct manipulation of an organism's genome, such as CRISPRCas9 (Ran *et al.*, 2013), the design of synthetic genetic circuits (Xie & Fussenegger, 2018), and increasingly sophisticated DNA sequencing technologies (Giani *et al.*, 2020). This has significantly reduced the time required for strain engineering where,

in contrast to the laborious processes of random mutagenesis and strain isolation, a direct and rational engineering approach can be taken regarding genetic manipulation for developing a new strain.

However, this is reliant upon the host organism's genetic tractability. Where model organisms in the past were chosen for the ease at which they could be maintained in a laboratory setting, their short generation time, and previous characterisation, a new requirement had emerged - an innate amenability to receiving foreign DNA. Some organisms have flourished under this new demand, such as and *E. coli* (Blount, 2015), *Bacillus subtilus* (Borriss *et al.*, 2018), and *Saccharomyces cerevisiae* (Parapouli *et al.*, 2020). Academic research and industrial application of these organisms has led to well-characterised genetic tools (Besada-Lombana *et al.*, 2018; Gao *et al.*, 2024; Wozniak & Simmons, 2022), full genome sequences (Blattner *et al.*, 1997; Goffeau *et al.*, 1996; Kunst *et al.*, 1997), and genome-scale metabolic models (Bernstein *et al.*, 2023; Bi *et al.*, 2023; Förster *et al.*, 2003). As a result, genetically recalcitrant organisms are a less attractive prospect for research and industrial biotechnology because rational design-build-test strain engineering strategies are either impossible or too laborious to implement in a timely manner, limiting indirect approaches such as random mutagenesis (Jeyachandran *et al.*, 2023) and adaptive laboratory evolution (Dragosits & Mattanovich, 2013).

1.1.1. Why non-model organisms matter

With around 20,000 named prokaryotes (Parte, 2018) and an estimated 2.2-4.3 million species worldwide (Louca *et al.*, 2019), there is vast potential for exploitation. Despite the practical difficulties associated with working with them, there is a strong case to be made for directing research efforts to increase the genetic tractability of high-potential non-model organisms so they can be utilised as chassis in industrial biotechnology.

Firstly, a microorganism engineered to produce a non-native molecule or even overexpress a native metabolite may experience inhibition in cell growth, cell viability, and product titres due to product toxicity. This is particularly an issue in biofuel production where solvents such as butanol and hexane can cause damage to the cell membrane and alter internal pH levels (Jin *et al.*, 2014). Chemical sensitivity and tolerance are arbitrated by complex mechanisms made up of the interactions between toxin-induced damage to biological molecules, biophysical changes in the cell membrane (affecting energy generation and metabolite transportation), and how the cell reacts to these change through stress responses (Nicolaou *et al.*, 2010). This makes product sensitivity a challenging problem to solve through a rational engineering approach with many variables to consider. Even when a tolerant strain is engineered, this does not necessarily coincide with increased product formation (Foo *et al.*, 2014; Lennen & Pfleger, 2013) and so it could be beneficial to turn to a species that exhibits an innate tolerance to the desired product and focus efforts developing the genetic tools for them.

Second, model organisms have been chosen for the ease at which they can cultured under laboratory conditions, however, conditions sustainable in a laboratory are limited compared to those available in industry. This means that the development of new industrial processes that utilise harsh environmental conditions such as high temperature, pH, and salinity and the organisms that can thrive in them, extremophiles, are not being explored to their fullest extent. These conditions have the potential to enable unique industrial processes that would otherwise be impossible to achieve with mesophilic organisms (species that inhabit moderate conditions). Therefore, there is a strong argument to be made to improve the genetic tractability of extremophiles of interest.

Thermophiles are organisms that have adapted to survive in high temperatures and the unique properties of their enzymes have been of particular interest to molecular biology. An example is *Taq* DNA polymerase from *Thermus aquaticus* that has been exploited to enable thermocycling due to its stability at high temperatures allowing it to withstand the ~95 °C during the denaturing step of PCR (Saiki *et al.*, 1988). Outside of their enzymes, several thermophiles have unique properties that could be exploited for industrial processes, *Caldicellulosiruptor bescii* and *Thermoanaerobacter mathranii* can utilise lignocellulosic biomass as a substrate for growth (Ahring *et al.*, 1999; Yang *et al.*, 2009), members of the archaeal *Sulfolobus* genus are adapted not only to temperatures of around 80 °C but are also tolerant to pH levels below 3, making them thermoacidophiles (Zeldes *et al.*, 2015), *Metallosphaera sedula* is capable of autotrophic growth on CO_2 and H_2 (Auernik & Kelly, 2010) as well as oxidising

iron- and sulur-containing compounds with potential applications in bioleaching (McCarthy *et al.*, 2018).

Halophiles are another group of extremophiles that have adapted to survive in environments with high salinity. This raises the possibility of replacing fresh water with seawater in industrial processes (X. Zhang *et al.*, 2018). Additionally, halophiles have been shown to accumulate high amounts of a group of biodegradable polyesters called polyhydroxyalkanoates (PHAs) that have applications as a material for plastics, textiles, cosmetics, medical implants, and drug carriers (Chen & Wu, 2005) (Park *et al.*, 2024). PHA can be generated with high productivity, with one study utilising *Haloferax mediterranei* demonstrating 87.5 % of dry cell weight being made up of PHA (Koller *et al.*, 2007). Moreover, high salinity discourages the growth of contaminating microorganisms which allows for fermentations of inexpensive waste substrates under non-sterile conditions. This is a concept currently being explored in *Halomonas* spp. to produce PHAs and poly-β-hydroxybutyrate (PHB) without an expensive substrate sterilisation process (Zhang *et al.*, 2024).

Organisms that survive on the extremities of the pH scale, acidophiles and alkaliphiles, also exhibit interesting properties with *Acidithiobacillus ferrooxidans* being an important microorganism in biomining where its ability to oxidise iron- and sulfur-containing compounds is utilised to extract metals such as copper from ores (Rawlings, 2002).

Finally, as well as product toxicity and environmental conditions, model organism are also limited in the variety of the carbon sources that can be used

as a feedstock. Organic carbon in the form of sugars, oils, and lignocellulosic biomass is the most common feedstock in industrial biotechnology (Ingle *et al.*, 2025; Vasileiadou, 2024), but their production can directly compete with food supply chains and encourage deforestation (Ajanovic, 2011). Consequently, there are environmental and economic incentives to utilise inorganic one-carbon compounds such as CO_2 and CO, which are often waste products from industry (Orsi *et al.*, 2023).

There are a number of inorganic carbon fixation pathways found in nature: the Calvin-Benson in cycle plants, algae, and cyanobacteria; the reductive tricarboxylic acid cycle in green sulfur bacteria; the Wood-Ljungdahl pathway found in acetogens; the 3-hydroxyproprionate pathway in the phylum *Chloroflexaota*; the Hydroxypropionate/4-hydroxybutyrate cycle in *M. sedula*; and the dicarboxylate/4-hydroxybutyrate cycle in *Ignicoccus hospitalis* (Garritano *et al.*, 2022; Santos Correa *et al.*, 2023). Whilst part of the Calvin-Benson cycle has been inserted into *E. coli* and *Pichia pastoris* to create full synthetic autotrophs (Gassler *et al.*, 2020; Gleizer *et al.*, 2019), these pathways could be exploited in their native organisms to valorise waste carbon and achieve a circular bio-economy.

If the issue of genetic intractability were to be solved, the unique properties of non-model organisms could be exploited to enable a strategy where a species that is naturally well-suited to an industrial process can be chosen as a platform strain. Rather than engage in ambitious and time-consuming metabolic rewiring of a model organism with no guarantee of success, smaller

adjustments could be made to fine-tune a non-model organism that already shows competency to that process.

1.2. Acetogens

Acetogens are group of microorganisms that have garnered considerable interest due to their innate ability as autotrophs to fix carbon in the form of CO_2 and CO to produce a variety of valuable compounds. Acetogens are anaerobic organisms that all produce acetate from acetyl-CoA, however acetyl-CoA can also act as a precursor for a diverse range of different compounds depending on the species. These include ethanol (Abrini *et al.*, 1994), butanol (Worden *et al.*, 1991), 2,3-butanediol (Köpke *et al.*, 2011), butyrate, hexanol, and hexanoate (Thunuguntla *et al.*, 2024) amongst other natural metabolites.

1.2.1. Gas Fermentation

Acetogens' natural ability to fix inorganic carbon to produce biofuels such as butanol means that their metabolism can be exploited to create a cyclical carbon chain for waste carbon produced in industry. Using carbon capture and storage technologies, CO_2 and CO can be used as a feedstock for fermentations with acetogenic bacteria to produce commodity chemicals, such as acetate and butyrate, and biofuels, ethanol and butanol, (Henstra *et al.*, 2007) which in turn can be used to power carbon-producing industries where waste carbon is captured again for the cycle to repeat (Köpke & Simpson, 2020).

In addition to capturing CO_2, carbonaceous matter like municipal, food, animal and plant waste can be broken down into a mixture of CO, CO_2, H_2, and N_2 called syngas. This is made by reacting the organic matter with an oxidising agent at

temperatures between 500 °C and 1600 °C under high pressure in process called gasification (Molino *et al.*, 2016). Syngas has been typically processed into liquid hydrocarbons through a chemical reaction developed in 1925 called the Fischer-Tropsch process. By heating syngas to ~300 °C under high pressures in the presence of metal catalysts such as iron, cobalt, nickel, and ruthenium, carbon monoxide is converted to methane which subsequently polymerises into longer chain hydrocarbons (Hu *et al.*, 2012).

Gas fermentations can use syngas as a feedstock to produce commodity chemicals and biofuels instead of the energetically laborious Fischer-Tropsch process. This approach has an advantage over Fischer-Tropsch for the polymerisation of syngas due to the inherent advantages of biological systems. Typically, gas fermentations run at a much cooler temperature, such as 37 °C, require much lower pressure, and do not require metal catalysts. Whilst chemical reactions are generally much faster than their enzyme mediated equivalents, the specificity of enzymes, irreversibility of the reactions, and lower by-product formation leads to higher efficiencies than those found in a Fischer-Tropsch reactor. Enzymes also demonstrate resistance to small amounts of sulfur and chlorine-containing contaminants that can inactivate metal catalysts. Moreover, the Fischer-Tropsch process is sensitive to variability in the H_2:CO composition of the syngas being processed so this ratio needs to be maintained to keep the reaction running. Syngas derived from biomass typically has a H_2:CO ratio lower than the 2:1 that is required for the Fischer-Tropsch process, necessitating further processing using a water-step shift reaction to lower the CO concentration to adjust this ratio accordingly. In contrast, during

biological polymerisation of syngas, a constant ratio in not a prerequisite for successful fermentation and, although it can alter the concentration of fermentation products, a wide range of ratios can be utilised (Köpke & Simpson, 2020; Liew, Martin*, et al.*, 2016).

Gas fermentation has advantages over first generation biofuels that use fermentation feedstocks such as vegetable oils or cane sugar, directly competing with food crops for arable land. This forces farmers, typically in underdeveloped countries, to make a choice between using their land for producing biofuel or food and has been linked to deforestation and volatility in food prices (Ajanovic, 2011). It also has environmental advantages over second generation biofuels, which use lignocellulosic biomass (LCB) as a feedstock. LCB is waste plant matter from agriculture, sugar cane mills, and paper mills made up of long polymers of cellulose and hemicellulose tightly bound to lignin. This forms a very stable chemical structure and as such demands tremendous energy to be broken down. This is achieved by chemical, enzymatic, and mechanical treatments to produce fermentable monosaccharides (Kumar *et al.*, 2020). This means processing LCB is wasteful, energetically intensive, and requiring heavy use of chemicals for pretreatment.

These advantages, along with flexibility in substrate selection for syngas generation, highlight the importance for exploiting acetogens as microbial chassis for gas fermentation to generate green energy, reducing greenhouse gas emissions, and achieving a net-zero society.

1.2.2. The Wood-Ljungdahl Pathway

Acetogen's impressive metabolic profile is made possible by the pathway through which they fix inorganic carbon: the Wood-Ljungdahl pathway (WLP) (**figure 3**). The WLP is the most efficient of the known non-photosynthetic inorganic carbon fixing pathways and uses CO_2 and CO as starting materials to form acetyl-CoA via two main branches. The first of which is the Eastern (or Methyl) Branch where CO_2 is reduced to formate by formate dehydrogenase, which is reacted with tetrahydrofolate (THF) to form formyl-THF. Formyl-THF is condensed into methenyl-THF, reduced twice sequentially to methylene-THF and then methyl-THF. The methyl group is then transferred to CoFeSP (cobalt and iron containing corrinoid sulfur protein) where it is donated to Coenzyme A (CoA) along with one molecule of CO to form acetyl-CoA. In the absence of ambient CO, CO_2 is reduced in CO by carbon monoxide dehydrogenase (CODH) as part of the Western (or Carbonyl) Branch of the Wood-Ljungdahl pathway (Ragsdale & Pierce, 2008). Acetyl-CoA is then used in further metabolic pathways for generating biomass or acetate for ATP generation.

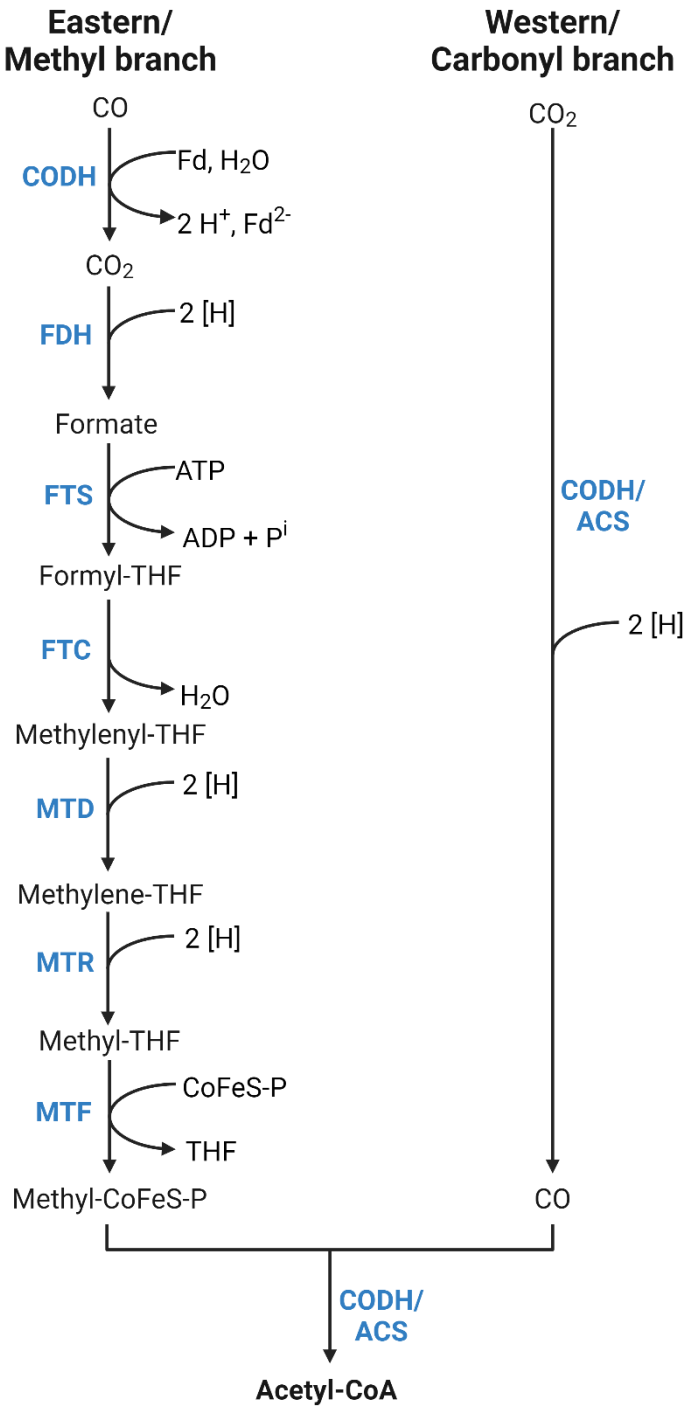

Figure 1 - The Wood-Ljungdahl Pathway.

CODH = Carbon Monoxide Dehydrogenase, FDH = Formate Dehydrogenase, THF = Tetrahydrofolate, FTS = Formyl-THF Synthase, FTC = Formyl-THF Cyclohydrase, MTD = Methylene-THF Dehydrogenase, MTR = Methylene-THF Reductase, MTF = Methyltransferase, ACS = Acetyl-CoA Synthase. Created with BioRender.com.

1.3. *Clostridium carboxidivorans*

Typical gas fermentation products consist of ethanol, acetate, and butanol but if the technology is to develop further then there needs to be a viable metabolic framework for the synthesis of longer and more complicated molecules. *C. carboxidivorans* (Liou *et al.*, 2005) belongs to an exclusive group of acetogens that can produce ethanol, butanol, and hexanol (as well as their equivalent carboxylic acids) and is an attractive candidate platform to produce higher molecules via gas fermentation. It is hoped that by elucidating the mechanisms behind the biosynthesis of longer carbon chain molecules, then these metabolic processes could be further exploited to produce more complicated molecules, either through natively or heterologously.

Moreover, hexanol and butanol are useful chemicals in and of themselves. Hexanol is generally non-toxic and useful in pharmaceuticals, cosmetics, perfumes, as a chemical precursor, detergents, as a solvent, in pesticides, and in the leather industry (Fernández-Naveira *et al.*, 2017). Since hexanol has a higher carbon content then ethanol and butanol, it will release more energy per molecule than both when used as a fuel. Butanol is used as a solvent in chemical processes, as a chemical precursor in the production of butyl acrylate and methacrylate, as well as an extractant in the cosmetic and pharmaceutical industries (Ndaba *et al.*, 2015).

1.3.1. *C. carboxidivorans'* metabolism

C. carboxidivorans' unique metabolic profile is made possible by the sequential addition of carbon to the acetyl group of acetyl-CoA in a series of reactions

known as reverse β-oxidation pathway (**figure 6**). Once acetyl-CoA has been formed in the final step of the Wood-Ljungdahl pathway, it can be converted into ethanol by a bifunctional acetaldehyde/alcohol dehydrogenase with acetaldehyde as an intermediate compound or into acetate by phosphotransacetylase followed by acetate kinase with acetyl-phosphate as an intermediate. Alternatively, acetyl-CoA can undergo a series of four reactions affecting the acetyl group carried out sequentially by the enzymes thiolase, 3-hydroxybutyryl-CoA dehydrogenase, crotonase, and butyryl-CoA dehydrogenase to form butyryl-CoA. This can be elongated even further by the same enzymes to form hexanoyl-CoA or converted into either butanol or butyrate. Hexanol and hexanoate can be formed from hexanoyl-CoA in an analogous way as butanol and butyrate using undetermined enzymes. (Vees *et al.*, 2022; Wirth & Dürre, 2021)

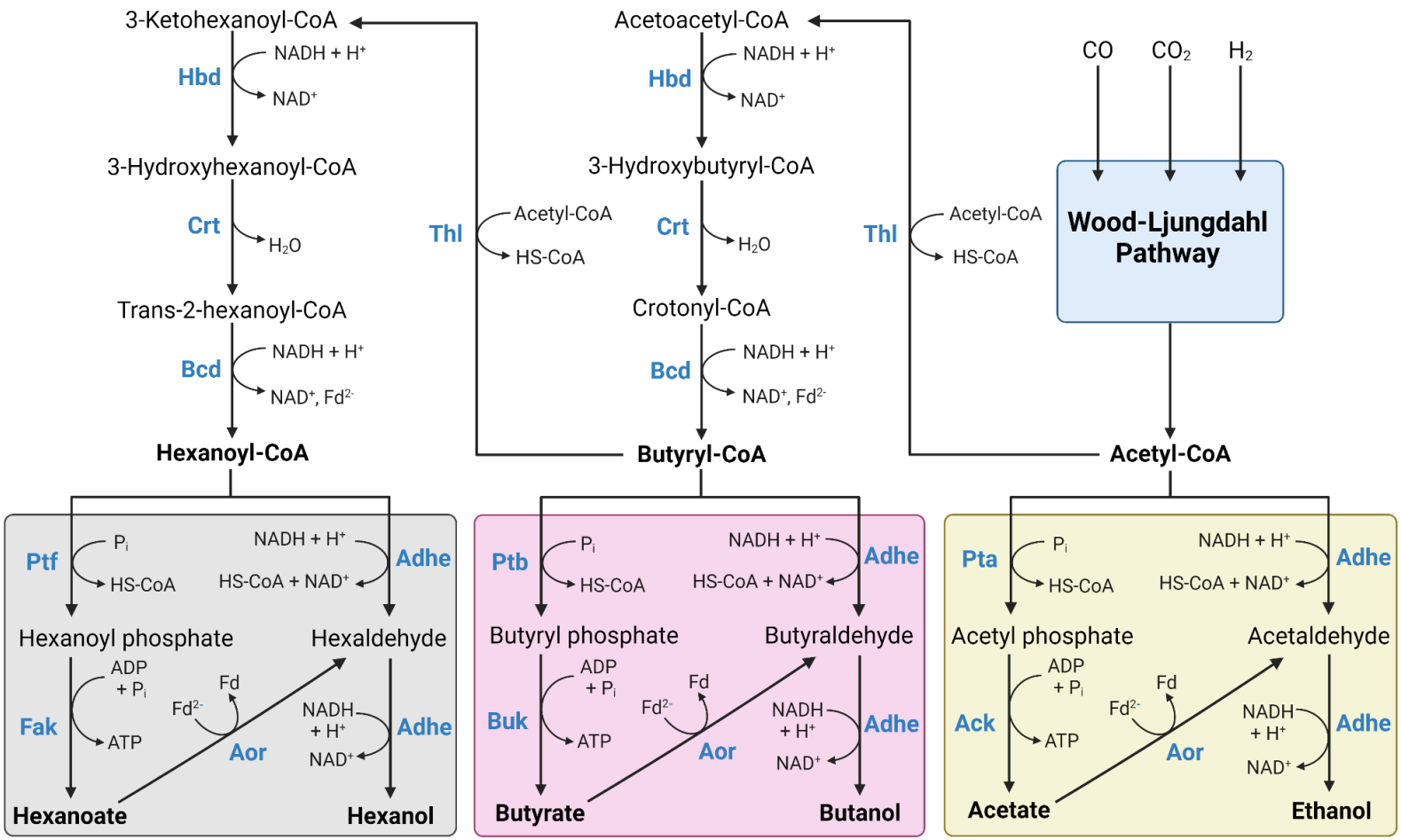

Figure 2 - The central metabolism of acidogenesis and solventogenesis from inorganic carbon in C. carboxidivorans P7 (Vees et al., 2022).

Ack = Acetate Kinase, AdhE = bifunctional acetaldehyde/alcohol dehydrogenase, Aor = aldehyde:ferredoxin oxidoreductase, Bcd = butyryl-CoA dehydrogenase, Buk = butyrate kinase, Crt = crotonase, Fak = fatty acid kinase, Fd = oxidised ferredoxin, Fd $^{2-}$ = reduced ferredoxin, Hbd = 3-hydroxybutyryl-CoA dehydrogenase, Pta = phosphotransacetylase, Ptb = phosphotransbutyrylase, Ptf = phosphotransferase, Thl = thiolase. Created with BioRender.com.

1.4. Barriers to DNA delivery

The main challenge to overcome for the genetic engineering of non-model organisms is the barriers imposed by the cell toward the acceptance of foreign DNA. The ability to deliver DNA into the cell is a fundamental requirement for any strain engineering project. Without a method for vector delivery, no genetic tools can be developed and no direct genetic modifications to the organism can be made.

There are number of strategies available to introduce DNA into a cell. The most common methods being electroporation, where by an electrical current is applied to the cell causing pores in the membrane to open allowing the entry of DNA, and conjugation, horizontal cell to cell gene transfer where a donor cell transfers a plasmid to a recipient cell through a pore in a physical junction formed between them.

During electroporation the cell wall or cell membrane presents the first barrier to DNA transfer by blocking physical entry into the cell. This is particularly problematic in Gram-positive bacteria, which have a thick peptidoglycan cell wall that can prove difficult to penetrate (Pyne *et al.*, 2014) with transformation frequencies of Gram-positive being orders of magnitude lower than that of Gram-negative (Pyne *et al.*, 2013). Strategies to chemically weaken the Gram-positive cell wall prior to transformation to increase efficiency have been attempted by treating the cells with various compounds such as DL-threonine

(Zhu *et al.*, 2005) and glycine (Cui *et al.*, 2012), as well as enzymatic treatments (Scott & Rood, 1989).

Where attempts to develop a working electroporation protocol for an organism fails, conjugation presents an alternative route for plasmid delivery, but this can encounter its own obstacles. Conjugation can be impeded by a phenomenon called 'surface exclusion'. If a recipient cell is already harbouring a conjugative element, then subsequent transfers of the same element into that cell are inhibited (Gago-Córdoba *et al.*, 2019). This is most well studied for F type plasmids in *E. coli* where two genes, *traS* and *traT*, impedes horizontal transfer of DNA in two different ways. *traS* encodes an inner membrane protein that prevents access to DNA into the cell after a mating pair has established whereas *traT* changes the outer surface of the cell in a way that inhibits the attachment of the conjugative pilus (Frost *et al.*, 1994).

Whilst each method of DNA transfer can come with its own impediments, restriction modification systems affect all DNA delivery techniques and pose a far larger obstacle to overcome.

1.4.1. Restriction Modification Systems

Restriction-modification (RM) systems are a rudimentary form of immune system found within prokaryotic organisms that allow the cell to protect itself from harmful foreign DNA introduced by invaders, such as bacteriophages. First described by Arber and Dussoix (1962), they consist of a nuclease and a methyltransferase, the nuclease targets specific DNA motifs that, if not bearing the correct methylation pattern, are identified as non-self and cleaved. The

methyltransferase complements this nuclease by methylating specific bases of the same DNA motifs in the host's genome to protect it from its own restriction enzymes (**figure 1**).

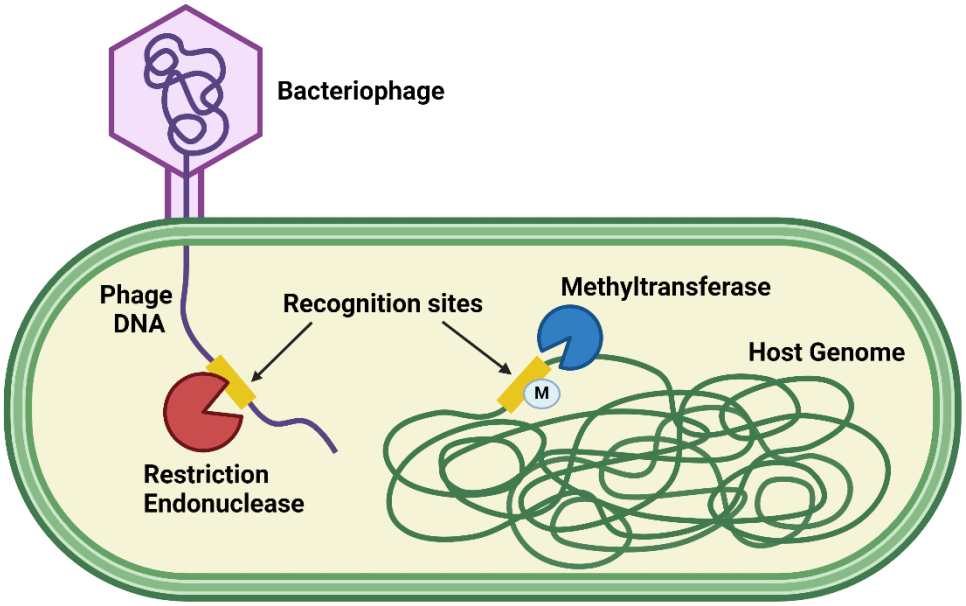

Figure 3 - The basic mechanism of Restriction Modification Systems.

A methyltransferase targeting a specific sequence methylates host DNA, protecting it from a restriction endonuclease that targets the same sequence. This nuclease cleaves unmethylated DNA of extracellular origin, such as in a phage attack, protecting the cell. Created with BioRender.com.

Restriction modification systems can be characterised based on their genetic organisation (**figure 2**) and protein structure and are broadly placed into four different categories:

Type I restriction modification systems consist of multiple subunits forming a single protein complex made up of a restriction (R) subunit, which cleaves DNA, a methylation (M) subunit, which catalyses the methylation of DNA, and the specificity (S) subunit, which determines the DNA sequence that is targeted by

the complex (Roberts, Belfort*, et al.*, 2003). These sub-units can form two different complexes determining their function: for endonuclease activity two R's two M's, and one S form a pentameric protein (Dryden *et al.*, 1997; Janscak *et al.*, 1998), and for methylation a trimer is formed of two M units and one S (Dryden *et al.*, 1993; Taylor *et al.*, 1992). This means that methylation can occur independently of the existence of the restriction sub-unit. All discovered type I restriction modification systems methylate adenine to form N6-methyladenine, require ATP hydrolysis for restriction, and can be further divided into the sub-categories A, B, C, and D based on sequence homology (Murray, 2000).

Type II systems differ from type I systems in that the nuclease and methyltransferase components of the system typically act independently from one another as separate enzymes and type II methyltransferases can methylate DNA to N6-methylcytosine, N4-methylcytosine, or N6-methyladenine (Roberts, Vincze*, et al.*, 2003). The nucleases from type II systems come in many different varieties and can be categorised into eleven distinct sub-types. This classification is based on several key features, such as if the DNA recognition sequence of the nuclease is palindromic or asymmetric, and whether the site of DNA restriction is inside, outside, or flanking the recognition sequence. Another criterion is the number of recognition sequence copies needed for cutting. Additionally, the requirement for methylation of the recognition sequence, and if restriction and methylation activity is performed by a single enzyme in a 'fused' system (Roberts, Belfort*, et al.*, 2003).

Type III systems are defined by the restriction and methylation components being encoded by separate genes within an operon that make up two sub-units of a single bifunctional enzyme (Wilson, 1991). Like type I systems, ATP is a required cofactor for DNA cleavage (Saha & Rao, 1995), methylation is possible independent of the restriction sub-unit (Hadi *et al.*, 1983), and adenine is methylated to form N6-methyladenine (Rao *et al.*, 2013).

Type IV systems differ from all other groups in that, rather than cleaving unmethylated DNA, they only target DNA that has been methylated incorrectly and they consist only of a nuclease with no methylation component (Stewart *et al.*, 2000).

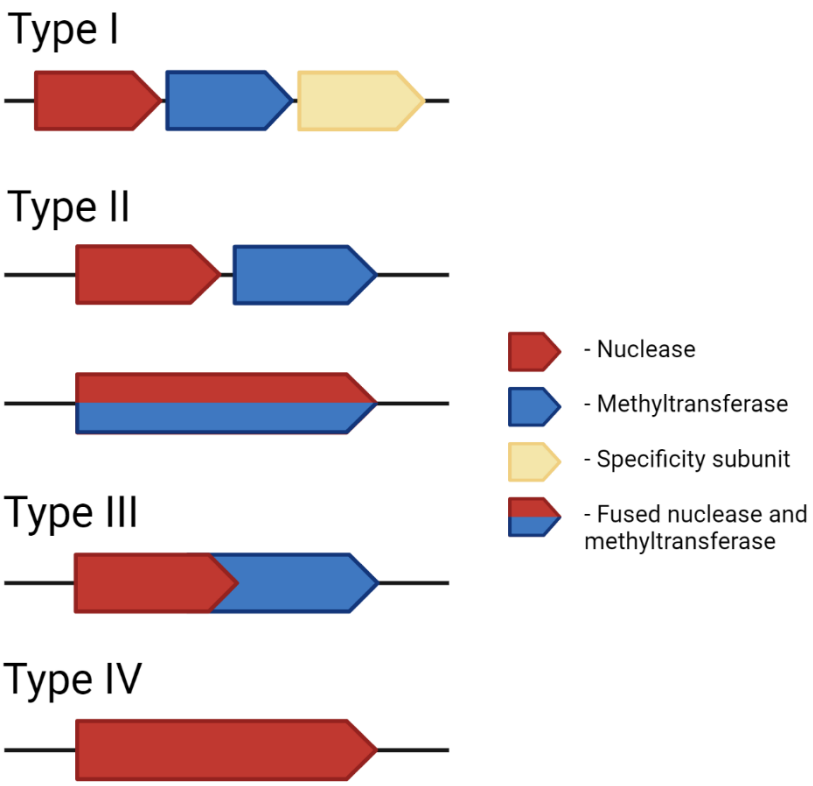

Figure 4 - The general genetic organisation of the various types of RM systems. Created with BioRender.com

1.5. Overcoming Restriction Modification Systems

The Restriction Enzyme Database (REBASE) is collection of restriction modification systems identified from the genomic sequencing data of bacteria and archaea. It contains information of restriction enzyme and methyltransferase sequences, the microorganisms they are found in, the recognition sequences they target, and their methylation activity. Of the 71,103 chromosomes analysed to date, 64,742 have been found to harbour at least one type of restriction-modification system (Roberts, Vincze*, et al.*, 2003). This highlights the ubiquitous nature of these systems within prokaryotes and as such developing an effective strategy to overcome them is of paramount importance if strain engineering of non-model organisms is to be feasible.

Given that restriction-modification systems are an adaptation to protect the cell from bacteriophages, it is reasonable to assume that bacteriophages have developed counter-adaptations in return. Therefore, by deducing how bacteriophages bypass restriction-modification systems we can deploy similar strategies when delivering foreign DNA into genetically recalcitrant organisms.

1.5.1. Mimicking Methylation Patterns

One strategy is to mimic the methylation pattern of the target organism's restriction-modification system to protect the invasive DNA from nuclease activity. The methyltransferase M.BsuRI found in *B. subtilis* methylates the central cytosine of the sequence 5'-GGCC-3' as part of a restriction-modification system. It has been found that the bacteriophages SPR, SPβ, ϕ3T, and ρ11 each produce their own methyltransferase that mimics this activity.

Interestingly, these methyltransferase genes show little homology (Noyer-Weidner *et al.*, 1983) despite having the same sequence specificity and serves as an example of convergent evolution.

The same strategy can be deployed in a laboratory setting by first determining which methylation patterns appear in the target organism by Single-Molecule Real-Time sequencing (SMRTseq). SMRTseq determines the nucleotide sequence of a DNA sample by utilising a single DNA polymerase housed in a small well alongside nucleotides tagged with four different base-dependent fluorescent dyes. Light is emitted as the polymerase incorporates the tagged nucleotides into the template DNA the spectra of which determines which base is being annealed. The secondary structure of the template DNA effects the kinetics of the polymerase, and it has been shown that methylated DNA increases the duration between pulses of fluorescence distinguishing methylated and unmethylated DNA as well as different forms of methylation of the same base. This allows the entire methylome of an organism to be established in parallel with its genomic nucleotide sequence (Clark *et al.*, 2012; Flusberg *et al.*, 2010).

Knowing the target organism's native methylation patterns and its nuclease target site specificity from REBASE, it can be deduced which methylation patterns in a methylome are part of a restriction-modification system. If these target sequences appear on the vector to be delivered, it can be protected from nuclease activity by treating the plasmid *in vitro* with appropriate commercial methyltransferases prior to transformation. However, these are not always available for purchase in which case the methyltransferases (or isoschimozers

thereof) from the target organism can be cloned into *E. coli* for *in vivo* methylation of the plasmid (Pyne *et al.*, 2014). The vector is methylated by an *E. coli* strain harbouring a separate plasmid containing methyltransferase genes conferring appropriate DNA methylation. Once the vector is extracted, it can then be transformed into the target host with orders of magnitude increased efficiency compared to an unmethylated vector. This strategy has been deployed for transformation of *Clostridium acetobutylicum* (Mermelstein & Papoutsakis, 1993), *Clostridium pasteurianum* (Pyne *et al.*, 2013), *Bifidobacterium adolescentis*, and *Lactococcus lactis* (Yasui *et al.*, 2009).

When protecting against the methyl-targeting type IV nucleases inadvertent methylation of the vector must be avoided. This is achieved by using strains of *E. coli* that are deficient in their native methyltransferase-encoding genes (Johnston *et al.*, 2019) as a conjugal donor or when treating a plasmid *in vivo*. An example of this is the *E. coli* sExpress conjugal donor strain, which is a NEB Express strain (originally developed for protein expression) modified to be a conjugal donor through the introduction of the conjugation enabling plasmid R702. The NEB Express strain lacks the methyltransferase-encoding gene *dcm* and so would not methylate the second cytosine of the recognition motif '5-CCWGG-3' on shuttle vectors prior to conjugation. sExpress was shown to produce superior conjugation efficiency compared to the natural *dcm*-containing conjugal donor strain CA434 when delivering a shuttle vector to *Clostridium autoethanogenum* and *Clostridium sporogenes* due to evasion of the recipient organism's Type IV restriction systems (Woods *et al.*, 2019).

1.5.2. Restriction site avoidance

The simplest way to escape nuclease activity is to avoid having the restriction site altogether. The target DNA specificity of nucleases can serve as an evolutionary disadvantage to the cell, because even a single base change in a phage's genome can render a restriction-modification system completely ineffective. This can be observed in nature with many phages lacking 6 base pair palindromic sequences (Sharp, 1986), a typical motif for the recognition sites of Type II systems, often with only one base pair difference from common palindromic sequences found in their target host organism's restriction-modification system (Blaisdell *et al.*, 1996).

This principle has been applied in a study where the recognition sites of the restriction-modification system found in *Staphylococcus aureus* were deduced via SMRTseq and REBASE. This allowed for these sites to be identified and removed from an *E. coli-S. aureus* shuttle vector pEPSA5 via 6 single base mutations whilst maintaining plasmid functionality. When used in conjunction with a *dcm*- strain of *E. coli*, transformation efficiencies increased 70,000-fold when compared to the unmodified vector (Johnston *et al.*, 2019).

1.5.3. Other Methods of RM Evasion

As well as mimicking the methylation profile of their target host, phages can protect their DNA through other kinds of nucleotide modifications. Hydroxymethylation, glucosylation, and other processes add bulky groups to nucleotides acting as a more generalised strategy for inhibiting the interactions of nucleases with phage DNA (Warren, 1980; Weigele & Raleigh, 2016). A well-

studied example of this is the *mom* operon found in the *E. coli* phage Mu. The product of which modifies adenine in the target sequence 5'-G/C-A-G/C-N-C/T-3' (Hattman, 1980) to N6-(1-acetamido)-adenine (Swinton *et al.*, 1983), conferring phage DNA resistance to restriction from nucleases that target this region (Krüger & Bickle, 1983).

Another method bacteriophages nullify restriction-modification systems is by expressing proteins that bind to phage DNA in such a way that nuclease recognition sites are occluded, preventing restriction. This can be found in the bacteriophage P1 where the genes *darA* and *darB* encode proteins that are injected into the cell in parallel with phage DNA and are thought to inhibit *E. coli*'s type I restriction-modification systems via binding to the phage DNA (Iida *et al.*, 1987).

As well as DarA and DarB, other proteins can be expressed by phages to interfere with restriction-modification systems. The phage λ and hybrid phage λ*reverse* encode the proteins Ral and Lar, which promote methyltransferase activity whilst suppressing restriction activity of the *E. coli* type I EcoKI system (King & Murray, 1995; Loenen & Murray, 1986). The anti-restriction protein Ocr of the phage T7 mimics the structural properties of the phosphate backbone of 24 base pair DNA (Walkinshaw *et al.*, 2002) with greater affinity for binding to EcoKI than DNA (Atanasiu *et al.*, 2002) and so likely acts as a competitive inhibitor of type I restriction enzymes.

1.6. Synthetic Biology in *Clostridium*

The advent of synthetic biology has brought with it a vast potential for further exploitation of the microorganisms capable of gas fermentation. Through genetic engineering there is scope to increase the efficiency of existing metabolic pathways through over expressing enzymes in rate limiting steps or by knocking out genes that are detrimental to desired metabolite production. It is also possible to produce non-native products by designing synthetic pathways, altering the host organism's metabolism through the addition of heterologous genes, or introducing entire metabolic pathways from other organisms.

Whilst the genus *Clostridium* is typically associated with its more nefarious members such as *Clostridium botulinum*, *Clostridium tetanus*, *Clostridioides difficile* (formerly *Clostridium*), and *Clostridium perfringens* due to the serious risk they pose to human health, there are also several industrially useful strains. *Clostridium acetobutylicum*, *Clostridium saccharoperbutylacetonicum*, and *Clostridium beijerinckii* have all been used to ferment crops and molasses to form acetate, butanol, and ethanol on an industrial scale (Jones & Keis, 1995; Jones *et al.*, 2023) and *Clostridium butyricum* is a probiotic used as prophylaxis for opportunistic *C. difficile* infections (Woo *et al.*, 2011). As well as this, *Clostridium* has several acetogenic species including *Clostridium autoethanogenum* (Liew *et al.*, 2022). Furthermore, *Clostridium ljungdahlii* (Zhang *et al.*, 2020), *Clostridium carboxidivorans*, *Clostridium ragsdalei*, and *Clostridium muellerianum* (Thunuguntla *et al.*, 2024) with *C. autoethanogenum*

being used for the industrial production of acetate and ethanol from inorganic carbon (Peplow, 2015).

As such, there is great interest in applying synthetic biology techniques to engineer beneficial strains of *Clostridium* for applications in health and as platforms to produce high value chemicals. Despite *Clostridium* having transformation protocols published as early as 1988 (Oultram *et al.*, 1988), the development of genetic tools for this genus has typically lagged behind that of other bacteria. However, in recent years there has been significant advancements in genetic tools for researchers starting with the development of the pMTL80000 modular series of shuttle plasmids. The modules consist of a suite of antibiotic selectable markers, replicons, and promoters enabling recombinant DNA research in *Clostridium* (Heap *et al.*, 2009) laying the foundations for the tools that would follow.

1.6.1. CRISPR-Cas9

One of the most important additions to the synthetic biologist's toolkit came from the discovery of Clustered Regularly Spaced Short Palindromic Repeats (CRISPR) and associated proteins (cas), which would become be one of the most successful gene editing tools available (Hsu *et al.*, 2014). Much like restriction modification systems, the CRISPRcas system's natural function is to act as an immune system for the cell by cleaving invading foreign DNA. When foreign DNA enters the cell, Cas proteins incorporate short fragments of it into the CRISPR locus of the host cell's genome, each forming a spacer. The CRISPR locus consists of multiple spacers from previous infections in chronological order with

each spacer flanked by repeated sequences of ~30 bp creating the CRISPR array (Mojica *et al.*, 2005). The CRISPR array is constitutively transcribed into preCRISPR RNA that is divided into CRISPR RNAs (crRNA). Located upstream is a trans-activating crRNA (tracrRNA), which forms a duplex with the crRNA for incorporation into the endonuclease forming a cas-RNA complex (Deltcheva *et al.*, 2011). This complex can identify, bind to, and cause a double strand break (Garneau *et al.*, 2010) in specific sequences of foreign DNA that complement the crRNA alongside an adjacent 2-6 bp DNA sequence called a Protospacer Adjacent Motif (PAM) (Bolotin *et al.*, 2005).

By providing cas9 endonuclease-encoding gene and guide RNA (a fusion of crRNA and tracrRNA) on a plasmid it is possible to perform specific *in vivo* editing of the host cell's genome with the locus cleaved determined by the crRNA sequence (Jinek *et al.*, 2012). To delete a gene of interest it is also necessary to provide a donor region on the plasmid containing two lengths of DNA that are homologous to the sequences flanking the target gene (Jiang *et al.*, 2013). Since DNA double strand breaks encourages homologous recombination (Bibikova *et al.*, 2001), the cell then repairs the break caused by the cas-RNA complex using the donor region resulting in a clean deletion of the gene (Jiang *et al.*, 2013). Genetic cargo can also be delivered for integration into the cell's genome by adding a coding region between the two homology arms (**figure 4**).

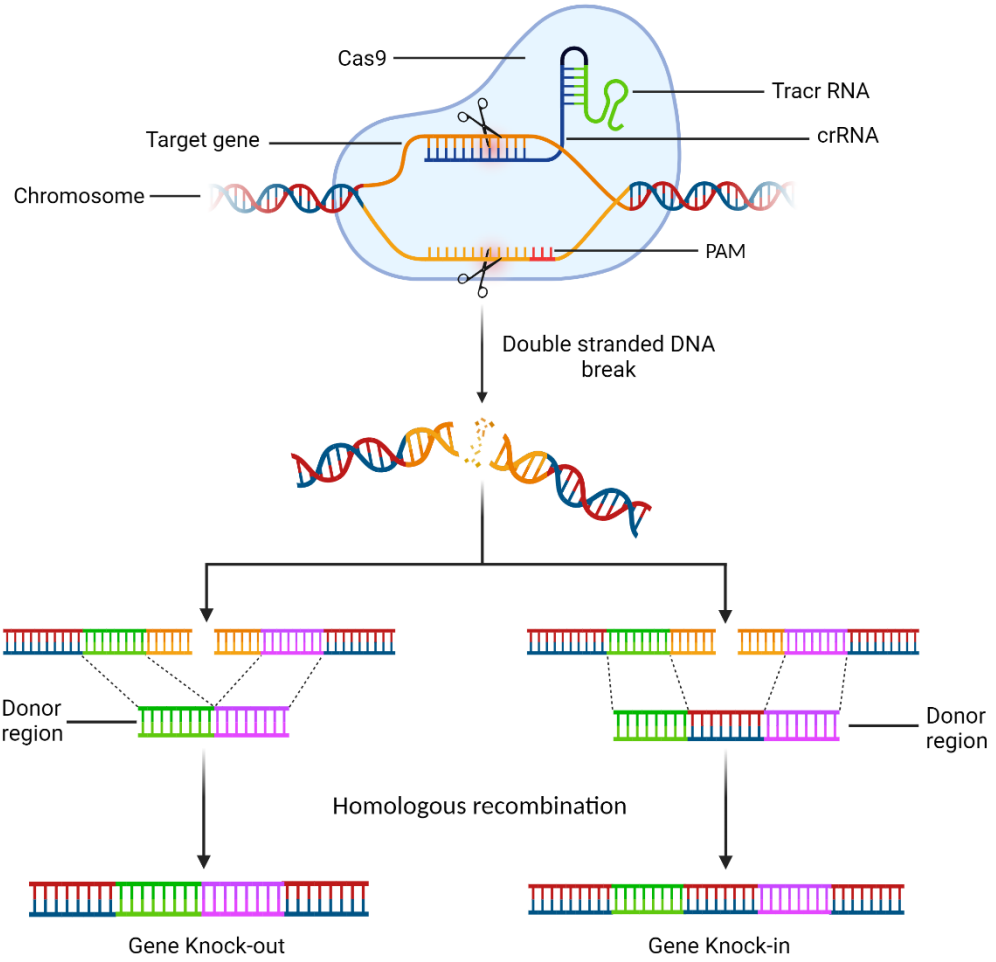

Figure 5 - Gene knockout and knock-in via cas9 mediated by homologous recombination. Created with BioRender.com

Since its discovery, CRSIPR-cas systems have been used successfully in the mutagenesis of several members of *Clostridium*, including *C. acetobutylicum* (Wasels *et al.*, 2017), *C. autoethanogenum* (Nagaraju *et al.*, 2016), *C. beijeinckii* (Wang *et al.*, 2015), *C. cellulolyticum* (Xu *et al.*, 2015), *C. cellulovorans* (Wen *et al.*, 2017), *C. ljungdahlii* (Huang *et al.*, 2016), *C. pasteurianum* (Bruder *et al.*, 2016), *C. saccharoperbutylacetonicum* (Wang *et al.*, 2017), and *C. tyrobutyricum* (J. Zhang, W. Zong, *et al.*, 2018).

CRISPR-cas gene editing technology has been developed for *Clostridium* initially expressing the *cas9* using a constitutive promoter (Wang *et al.*, 2018; Wang *et al.*, 2015; Wasels *et al.*, 2017). However, the Cas9 endonuclease is often toxic to the cell so it can be a beneficial to regulate *cas9* expression using an inducible promoter such as tetracycline (Wasels *et al.*, 2017), lactose (J. Zhang, W. Hong*, et al.*, 2018), and xylose (Muh *et al.*, 2019). This approach has since been refined by using a theophylline inducible riboswitch to control the promoter driving *cas9* expression. The RiboCas system allows for the tight regulation of *cas9* expression with induction only occurring when necessary, resulting in high levels of conjugation and knock-out efficiency (Cañadas *et al.*, 2019).

1.6.2. ClosTron

Before the availability of CRISPRcas based tools, mutant alleles were generated by mobile group II introns - a class of ribozymes that can self-splice from transcriptional RNA and insert themselves into a target region on the chromosome (**figure 5**). This action is facilitated by an Intron-Encoded Protein (IEP) with which the spliced RNA forms a lariat structure called a ribonuclear protein (RNP) complex that can recognise and bind to specific sequences in the chromosome. The IEP nicks the target DNA for insertion of the spliced RNA which is followed by reverse-transcriptase activity of the IEP to form a complementary strand of DNA. The RNA is then degraded by the host organism's nucleases and is replaced with an equivalent strand of DNA by the host's DNA polymerase. The DNA insert is then fully integrated into the chromosome by DNA ligase. Whilst not a true knockout, the insertion of the

intron disrupts the integrity of the target coding region resulting in the expression of a defective gene product, effectively inactivating the gene.

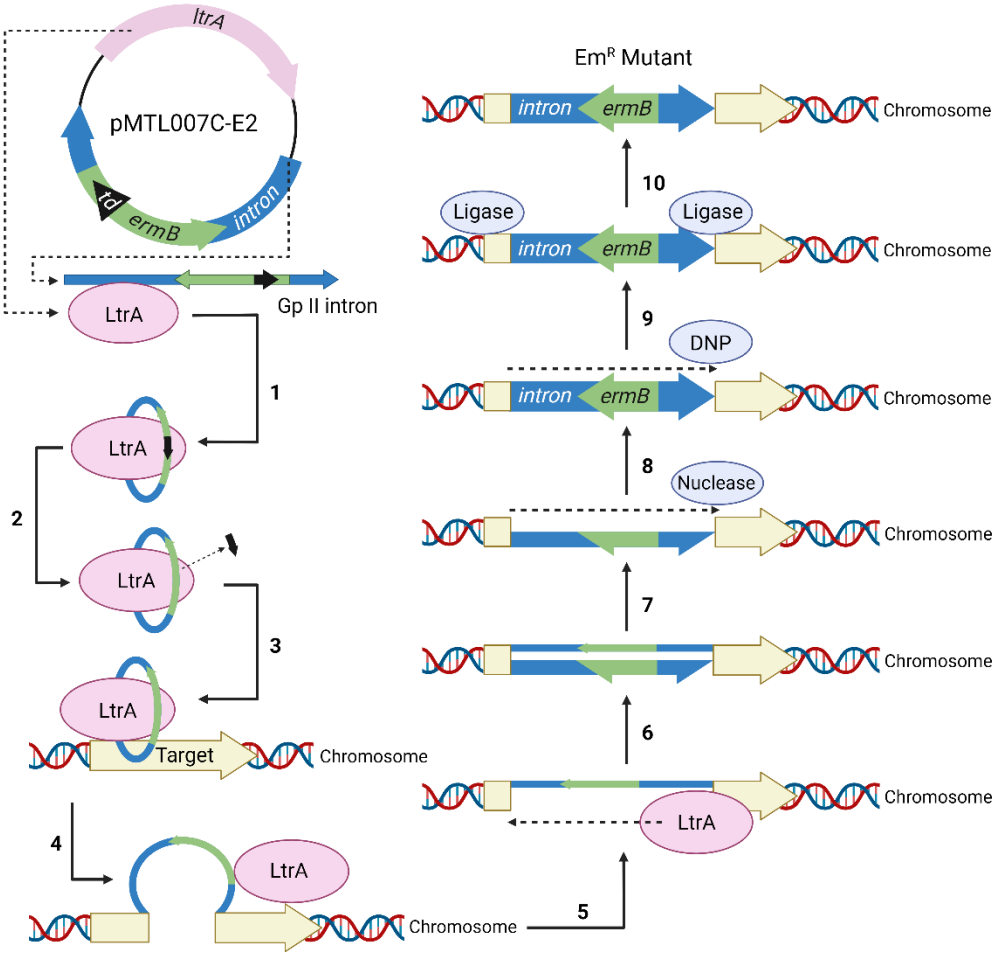

Figure 6 - The mechanism of mutagenesis using the ClosTron system.

The ClosTron plasmid "pMTL007C-E2" harbours a group II intron (blue) inside of which lays ermB (green), which is disrupted by the group I td intron (black) and so does not confer antibiotic resistance. (1) The RNA transcript of the group II intron containing ermB and td form a ribonuclear protein (RNP) complex with LtrA. (2) The td group I intron splices out of the complex. (3) The RNP recognises and binds to the target loci as determined by the group II intron sequence. (4) LtrA nicks the target DNA and 5. The RNA is inserted into the chromosome. (6) LtrA acts as a reverse transcriptase synthesising a complementary DNA strand to that of the inserted RNA. (7) Insert RNA is degraded by the host. (8) Host DNA polymerase (DNP) synthesises the opposite DNA strand. (9) Host DNA ligase seal gaps between the insert and

chromosome. (10) The target coding region is now disrupted by the group II intron and the host exhibits erythromycin resistance from the expression of the restored ermB gene, allowing for selection. Created with BioRender.com.

The target specificity of group II introns is determined primarily by interactions between the target DNA and the ~14 nucleotide bases in the excised RNA (Mohr *et al.*, 2000). Consequently, the insertion locus can be altered by introducing specific nucleotide changes to the RNA. This allowed for the development of the TargeTron technology where an algorithm was formulated that could reliably predict the nucleotide changes required to target specific loci in a known sequence (Perutka *et al.*, 2004) creating a tool for specific gene disruption when delivered on a vector (Chen *et al.*, 2007).

With no way to positively select for integrants, screening for mutants without an obvious phenotypic change can be a laborious process and, since integration frequencies can vary depending on the targeting region, low-efficiency targets can be prohibitively difficult to isolate. As such, there was a demand for the development of a built-in mutant selection system, which was met in the form of a Retrotransposition-Activated Marker (RAM).

The RAM is housed within the group II intron encoding region and consists of an antibiotic resistance gene interrupted by a group I intron from the *td* gene in bacteriophage T4. The *td* intron is spliced during the retrotransposition of the group II intron, restoring the integrity of the antibiotic resistance gene when inserted into the chromosome. The inclusion of a group I intron in the coding region of the antibiotic resistance gene ensures that resistance is only conferred after a successful integration event allowing for the selection of mutants on

agar plates supplemented with the appropriate antibiotic (Ichiyanagi *et al.*, 2002; Zhong *et al.*, 2003).

This technology was adapted to create a system for mutagenesis in *Clostridium* called ClosTron. This utilised the genetic components developed in the pMTL80000 series of vectors and follows the same philosophy that each part should consist of its own module separated by unique restriction sites. The ClosTron vector consists of a pCB102 or pCD6 Gram-positive replicon; *CatP, ErmB,* or *Aad9* antibiotic resistance marker; ColE1 gram negative replicon with *traJ* enabling conjugative transfer; and the ClosTron module. The ClosTron module consists of a constitutive P_{fdx} promoter from the ferredoxin gene of *Clostridium sporogenes* controlling the expression of the downstream group II intron followed by an ErmB based RAM, and finally the IEP *ltrA* (Heap *et al.*, 2010; Heap *et al.*, 2007).

ClosTron vectors have been used to successfully generate mutants in *Clostridioides difficile*, *Clostridium botulinum, C. sporogenes*, *C. beijerinckii*, *C. sordellii* (Heap *et al.*, 2010), *C. autoethanogenum* (Liew, Henstra*, et al.*, 2016), *C. acetobutylicum* (Cooksley *et al.*, 2012), and *C. cellulolyticum* (Cui *et al.*, 2014).

1.6.2. Genetic modification in *C. carboxidivorans*

Existing research into *C. carboxidivorans* is primarily focused on the adjustment of fermentation parameters and the effect those changes have on growth and metabolite production. These parameters include: pH (Fernández-Naveira *et al.*, 2016), fermentation temperatures (Ramió-Pujol *et al.*, 2015; Zhang *et al.*, 2016), salinity tolerance (Fernández-Naveira *et al.*, 2019), mixotrophic

fermentations (Vees *et al.*, 2022), culturing methods (Phillips *et al.*, 2015), trace metal optimisation (Han *et al.*, 2020), alternative bioreactor designs (Doll *et al.*, 2018; Shen *et al.*, 2014; Y. Shen *et al.*, 2017), synthetic co-cultures (Bäumler *et al.*, 2022), product toxicity (Kottenhahn *et al.*, 2021), CO_2 as a sole carbon source (Thunuguntla *et al.*, 2024), and tolerance to common syngas impurities (Rückel *et al.*, 2021).

Whilst progress has been made toward elucidating optimal fermentation processes for *C. carboxidivorans*, there is limited research in genetic modification and strain engineering. Only one study at present claims to have successfully conjugated a plasmid into *C. carboxidivorans* for the overexpression of the genes *adhE2*, *fnr*, and *aor* (Cheng *et al.,* 2019*)*. However, any attempts to consistently achieve conjugal transfer of DNA until now have been unsuccessful. This is largely due to the comprehensive restriction modification system found in *C. carboxidivorans* consisting of ten restriction-endonucleases of various types (Kottenhahn *et al.*, 2023) making the introduction of foreign DNA by conventional means impossible. Unsurprisingly is only one example of mutagenesis (Lakhssassi *et al.*, 2020) and adaptive laboratory evolution (Antonicelli *et al.*, 2023) in *C. carboxidivorans*. With no way to directly modify the genome to restore the mutated gene, it is impossible to prove if the mutations produced in these experiments are responsible for the observed changes in phenotype.

1.7. Aims of this project

The main aim of this study is to create a genetically domesticated strain of *C. carboxidivorans* that can be readily deployed for engineering to produce industrially relevant products. This will be achieved by applying the conjugal methylation donor strains and siteless Cas9 technology developed by Redfern (2021) to generate deletions of all RMS associated endonuclease encoding genes, enabling DNA transfer.

Once the genetically domesticated strain has been produced, a metabolic engineering approach will be adopted to generate a knockout in a gene that is hypothesised to improve growth and ethanol, butanol, and hexanol yield in CO_2 and H_2 gas fermentations. This will serve as proof of concept for the usefulness of this strain for industrial applications and for academic research.

ClosTron will also be explored as an alternative mutagenesis technique in *C. carboxidivorans* as well as the development of a new retrotransposition-activated marker for use in other important members of *Clostridium*.

Chapter 2 – Materials and Methods

2.1. Bacterial strains and culture conditions

2.2.1. Bacterial Strains

The following is a list of all bacterial strains utilised in this study.

Table 1 - Bacterial strains used in this study.

Strain	Function	Genotype/Description	Source
NEB® *E. coli* 10-beta	Cloning and plasmid storage	*Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	New England Biolabs, USA
E. coli sExpress	Donor strain during conjugal transfer of DNA.	R702 plasmid harbouring derivative of NEB *E. coli* Express strain. *fhuA2 [lon] ompT gal sulA11 R(mcr-73:miniTn10--TetS)2 [dcm] R(zgb-210:Tn10--TetS) endA1 Δ(mcrC-mrr)114:IS10* R702-Tc^R, Sm^R, Su^R, Hg^R Tra+, Mob	(Woods *et al.*, 2019)
E. coli sExpress_TII	Donor strain for conjugal DNA transfer to *C. carboxidivorans* .	Derivative of *E. coli* sExpress. Harbours recombinant R702 plasmid that contains Type II methyltransferase genes from *C. carboxidivorans P7.* Protects shuttle vector from Type II nucleases.	(Redfern, 2021)
E. coli sExpress_TIII	Donor strain for conjugal DNA transfer to *C. carboxidivorans*	Derivative of *E. coli* sExpress. Contains Type III methyltransferase genes.	(Redfern, 2021)

		Protects shuttle vector from Type III nucleases.	
Clostridium carboxidivorans P7	Compared to mutants in strain characterisation	Wild type strain	Purchased from the DSMZ. First published by Liou *et al.* (2005).
Clostridium carboxidivorans Δ2RM	Starting strain for genetic domestication.	A Type III and Type I RMS nuclease knockout mutant of *C. carboxidivorans C94*.	Christopher Humphreys
Clostridioides difficile R20291	Test species for alternative RAM development.	Hyper virulent strain of *C. difficile* isolated from an outbreak at Stoke Mendeville hospital, UK.	SBRC culture collection.
Clostridium sporogenes DSMZ 795	Test species for alternative RAM development.	Wild type strain.	SBRC culture collection.
Clostridium butyricum DSMZ 10702	Test species for alternative RAM development.	Wild type strain.	SBRC culture collection.

2.2.2. Routine bacterial culture conditions

All liquid cultures of *E. coli* strains were grown in Lysogeny Broth (LB) incubated at 37 °C in a shaking incubator with an agitation of 200 RPM. All culturing of *C. carboxidivorans*, *C. butyricum*, *C. sporogenes*, and *C. difficile* was performed in an anaerobic cabinet (MG1000 Mark II Anaerobic Workstation (Don Whitley, UK) at 37 °C with no agitation. When anaerobic conditions were required for bacterial growth in liquid media or agar plates these would be placed into the

MG1000 Mark II Anaerobic Workstation at least 72 hours and 16 hours prior to inoculation, respectively.

C. carboxidivorans was grown in YTAF MES media, a variation of the TYA media developed by Benevenuti *et al.* (2020) with the addition of fructose and MES buffer. The exact composition of the media was 14 g/L of tryptone, 9 g/L of yeast extract, 1.4 g/L of L-arginine, 10 g/L of 2-(*N*-morpholino) ethanesulfonic acid (MES) buffer, and 10 g/L of fructose. Once all components were in solution the pH of the media was adjusted 6.1.

C. butyricum and *C. difficile* were grown in BHIS media that consisted of 37 g/L of Brain Heart Infusion broth (Oxoid, UK), 5 g/L Yeast Extract, and 1 g/L of L-Cysteine.

C. sporogenes was grown in TYG media containing 30 g/L of Tryptone, 20g/L of Yeast Extract, and 1 g/L of Sodium Thioglycolate.

When solid media was required 15 g/L of agar was added to the above recipes prior to autoclave. The resulting mixture was melted using a microwave and 25 mL was poured into round agar plates.

2.2.3. Media supplements

Media supplements were prepared, filter sterilised, and stored as stock solutions as outlined in **table 2**.

Table 2 - Media supplements used in this study. **Thiamphenicol working concentration at 7.5 µg/mL for all species except when counter selecting C. sporogenes, which is at 20 µg/mL.*

Supplement	Diluent	Stock concentration (mg/mL)	Working concentration *E. coli* (µg/mL)	Working concentration Anaerobes (µg/mL)
Chloramphenicol	Ethanol	50	50	-
Clarithromycin	DMSO	6	-	6
D-Cycloserine	dH_2O	50	-	250
Erythromycin	Ethanol	50	500	-
Kanamycin	dH_2O	50	50	-
L-arabinose	dH_2O	150	-	1500
Lincomycin	dH_2O	20	-	20
Theophylline	DMSO	45	-	90
Thiamphenicol	1:1 Ethanol:dH_2O	15	-	7.5 or 20*
Uracil	dH_2O	1	-	10

2.2.4. Strain storage and revival

E. coli strains were stored at -80 °C in a 2 mL screw top cryotube with 500 µL of exponential culture and 500 µL of filter sterilised 50% v/v glycerol dH_2O solution. Anaerobic strains were also stored in a 2 mL screw top cryotube with 850 µL of exponential culture and 150 µL of DMSO.

Strains were revived by taking a small amount of the frozen stock using 1 µL inoculation loop to inoculate 3 mL of the relevant media used for the strain as per **2.2.2** in 5 mL culture tubes. This media would also be supplemented with the appropriate antibiotic if the strain was harbouring a plasmid.

2.3. Plasmid transformation and conjugation

2.3.1. Preparation of chemically competent *E. coli*

Chemically competent *E. coli* was prepared by creating a 5 mL overnight culture of the desired strain supplemented with any relevant antibiotic, in this study this would either be 10-beta, sExpress, sExpress T_II, or sExpress T_III. The following day, 1 mL of this culture would be used to inoculate two 500 mL conical flasks containing 100 mL of LB supplemented with antibiotic, if appropriate. This was incubated at 37 °C 200 RPM until an OD_{600} between 0.3 – 0.5 had been reached. The flasks were placed on ice for 15 minutes and were evenly aliquoted between six 50 mL falcon tubes. These were centrifuged at 2700 g for 10 minutes at 4 °C. The supernatant was removed the pellet in each falcon tube was resuspended in a chilled 10 mL solution of 100 mM $CaCl_2$ and 20% glycerol v/v dissolved in dH_2O. Once resuspended, the cells were placed on ice for a further 15 minutes and then centrifuged again at 2700 g for 10 minutes at 4 °C. The supernatant was then discarded, and each cell pellet was resuspended in 500 µL of the chilled 100mM $CaCl_2$ 20% v/v glycerol solution. The cells were aliquoted in volumes of 50 µL into 2 mL screw top cryotubes and stored at -80 °C until transformation.

2.3.2. Transformation of *E. coli*

To transform chemically competent *E. coli* created in **2.3.1** one tube containing 50 µL of competent cells per plasmid to be transformed was left on ice until defrosted, which took around 30 minutes. 1 µL of plasmid would be added to

the cell and mixed by gently flicking the tube and the resulting cell-DNA mixture would be kept on ice for 30 minutes. This would be heat shocked at 42 °C in a water bath for 30 seconds and placed on ice for 5 minutes. 950 µL of 10-beta/Stable Outgrowth Media (New England Biolabs, USA) would be added to the mixture and the cells were incubated at 37 °C 200 RPM for 60 minutes.

While the cells were incubating, LB agar selection plates were prepared supplemented with the appropriate antibiotic to maintain the transformed plasmid. In the case of sExpress and sExpress-derived strains of *E. coli*, the plates would be additionally supplemented with kanamycin to select for R702.

Once the cells had incubated for 60 minutes 100 µL of the culture would be spread onto the plates and incubated at 37 °C until pickable colonies had formed.

2.3.3. Conjugative DNA transfer

All plasmid DNA was transferred into non-*E. coli* species in this study via conjugation. Before conjugation the desired plasmid was transformed into chemically competent *E. coli* sExpress to act as a conjugal donor strain, this was saved as a stock as per **2.2.4** until required.

A stock of the desired recipient species was revived in 3 mL of species appropriate media, as outlined in **2.2.1**, in an anaerobic cabinet. Once the stock had revived and the culture had reached late exponential/early stationary phase, a 3 mL subculture was prepared using the same media and 30 µL of the revived stock as an inoculum as well as one agar plate per conjugation of the

species appropriate media type to act as a mating plate. On the same day a 5 mL LB culture, supplemented with kanamycin and shuttle plasmid specific antibiotic, of the *E. coli* sExpress conjugal donor strain was prepared and incubated at 37 °C 200 RPM.

The following morning, 50 µL of the *E. coli* sExpress conjugal donor strain was sub-cultured into another 5 mL of LB supplemented with the same antibiotics as the previous culture and incubated at 37 °C 200 RPM. Once this culture had reached an OD_{600} 0.2 - 0.4, 1 mL of it was aliquoted into a sterile 1.5 mL Eppendorf tube using a wide bore pipette and centrifuged at 3000 g for 3 minutes. The supernatant was carefully removed and discarded, and the pellet was gently resuspended in 500 µL of phosphate buffered saline (PBS) followed by a second centrifugation at 3000 g for 3 minutes after which the supernatant was removed. The tube containing the cell pellet was transported into the anaerobic cabinet and 200 µL of the recipient strain culture was added and the pellet was resuspended by flicking the tube. The resulting mixture of cells was spotted on to a mating plate using a wide bore pipette, once the spots had dried the plate was inverted and left in the cabinet for 16-18 hours. At the same time 2-5 species-specific agar plates are then placed in the anaerobic cabinet supplemented with an antibiotic selecting for the conjugated plasmid and d-cycloserine to counter-select *E. coli*.

Once this time had passed, the cells were harvested by flooding the mating plate with 1 mL of anaerobic PBS and using a wedge-shaped spreader to dislodge and resuspend the growth with the resulting slurry pipetted into a

sterile 1.5 mL Eppendorf tube. This was split across the selection plates with 100 µL per plate. Transconjugant colonies could be observed typically 1-4 days depending on the recipient species.

2.3.4. Conjugation efficiency of *C. carboxidivorans* genetically domesticated strains.

When comparing the conjugation efficiency between increasingly genetically domesticated strains of *C. carboxidivorans* the general conjugation protocol was followed as per **2.3.3** with the following changes:

- When the recipient and donor strain were mixed prior to spotting onto mating plates, the recipient strain of *C. carboxidivorans* was always in exponential phase and the OD_{600} was normalised to 1 across all repeats and strains analysed.
- The OD_{600} of the donor strain was normalised to 0.2 across all repeats and strains analysed.
- All mating plates were harvested after 18 hours of incubation.
- Variance of cell concentration in the harvested mating plate slurry was accounted for by assuming that the same number of recipient cells are harvested regardless of the total volume recovered from the plates. The volume of slurry collected in the 1.5 mL Eppendorf tubes was normalised by weighing them and diluting with PBS or transferring less volume to selection plates.

- A serial dilution of the slurry was performed to produce countable colonies on selection plates.

The conjugation efficiency of each strain was measured by counting the colonies formed on selection plates multiplied by dilution factor. This was repeated in triplicate for each strain and each repeat was performed in technical duplicate.

2.3.5. Plasmid Curing

Once mutagenesis using RiboCas or ClosTron is confirmed, it is then necessary to remove the associated plasmid. This was performed by reviving the plasmid containing strain and sub-culturing once per day for 3 days. On the fourth day a serial dilution of the culture was performed and 100 µL of each dilution was spread onto separate agar plates of species-appropriate media without selection. Once single colonies had formed, 50 were picked and transferred onto two agar plates, one with plasmid specific antibiotic selection and the other without. Plasmid loss was indicated in colonies that did not grow on the antibiotic supplemented plate but did on the plate without antibiotic. Up to three plasmid cured colonies were then used to inoculate appropriate liquid media to create bacterial stocks.

2.4. Molecular Biology techniques

2.4.1. Polymerase Chain Reaction

For the generation of DNA fragments for molecular cloning and sanger sequencing the high-fidelity polymerase Q5 (New England Biolabs, USA) was used as per the manufacturer's instructions.

Diagnostic colony PCRs for the screening of mutants was performed with the low-fidelity polymerase OneTaq (New England Biolabs, USA) as per the manufacturer's instructions. Template DNA for clostridial colony PCRs was prepared by picking the colonies to be screened from agar plates using a sterile toothpick, resuspending the cells in 20 µL of Monarch Plasmid Resuspension Buffer (New England Biolabs, USA) in a 0.2 mL PCR reaction tube, and boiling at 95 °C for 5 minutes in a thermocycler. Template DNA for *E. coli* colony PCRs was produced in the same way except cells were resuspended in dH_2O instead of the Plasmid Resuspension Buffer.

2.4.2. Restriction enzyme digestion of DNA

All restriction enzymes used in this study were manufactured by New England Biolabs (USA) and digestion reactions were performed as according to the manufacturer's instructions.

2.4.3. Plasmid DNA extraction

All plasmid DNA extracted from *E. coli* was performed using a Monarch® Plasmid Miniprep Kit (New England Biolabs, USA) as per the manufacturer's instructions.

2.4.4. Genomic DNA extraction

Genomic DNA was extracted from bacterial species for use as a DNA template or for Illumina full genome sequencing using a GenElute Bacterial Genomic DNA Kit (Sigma-Aldritch, USA) as per the manufacturer's instructions.

2.4.5. DNA visualisation and purification

DNA was visualised via agarose gel electrophoresis. Gels were made using a 1% w/v solution of agarose in TAE buffer (Tris 40 mM, acetic acid 20 mM, and EDTA 1 mM), which was heated in a microwave until the agarose had completely dissolved and kept in a 55 °C water bath until required. The molten gel was poured into casting tray with a gel comb to create wells followed by the addition of 0.001% v/v of SYBR Safe DNA Gel Stain (ThermoFisher Scientific, USA) and the gel was left to set. The solidified gels were placed into a gel electrophoresis tank and the DNA to be visualised, pre-mixed with 6x Purple Gel Loading Dye (New England Biolabs, USA), was loaded into the gel wells with one well loaded with Generuler 1 kb Plus DNA Ladder (ThermoFisher, USA). A 100 V current was run through the gel for 40-50 minutes until distinct, separate bands of DNA could be seen under UV light.

DNA was extracted and purified from agarose gels for use in molecular cloning and sanger sequencing using a Monarch DNA Gel Extraction Kit (New England Biolabs, USA).

2.4.6. Sanger sequencing

Constructed vectors and DNA fragments of loci of knocked out genes were screened via sanger sequencing, which was performed by Eurofins Genomics (Germany) and samples were prepared according to their guidelines.

2.4.7. Molecular Cloning

All molecular cloning in this study was performed by NEBuilder Hifi DNA Assembly (New England Biolabs, USA) using Hifi DNA Assembly Master Mix (New England Biolabs, USA) and DNA fragments generated from high fidelity PCR, restriction digest reactions of preexisting vectors, or ordered synthesised DNA fragments as per the manufacturer's instructions.

2.4.8. Construction of RiboCas knockout vectors

A knockout vector based on previous work (Redfern, 2021) created by Christopher Humphreys was purified from a revived *E. coli* stock (**2.4.3.**) and digested with the restriction enzymes *Sal*I and *Asc*I. Meanwhile, high fidelity PCR was used to generate DNA fragments for the left and right homology arms, using purified *C. carboxidivorans* genomic DNA as a template, as well as for the guide RNA (table 3), using the previous knockout plasmid as a template.

The plasmid backbone from the restriction digest reaction and all the PCRs were purified using gel electrophoresis followed by extraction. These 4 DNA

fragments were combined in a NEBuilder Hifi DNA Assembly reaction that, post incubation, was transformed into chemically competent *E. coli* 10-Beta (**2.3.2.**).

Once pickable colonies were formed on agar plates, these were screened by colony PCR using the forward primer for guide RNA generation and the reverse primer for right homology arm generation. Colonies that produced a band of 1.6 kb, indicating successful insertion of guide RNA and homology arms, were chosen for further screening via sanger sequencing by using these colonies to inoculate liquid cultures supplemented with chloramphenicol that were grown overnight with the vectors subsequently purified (**2.4.3.**) and sent for sanger sequencing (**2.4.6.**) along with the forward sequencing primer (Ccathlguide_F) which anneals upstream of the guide RNA, and the reverse sequencing primer (modpCB102_R) which anneals downstream of the right homology arm.

E. coli stocks were made at the same time as plasmid purification and stored at -80 °C until the sanger sequencing results were received where negatives were discarded.

Guide RNAs were chosen using Benchling's CRSIPR Guide RNA Design Tool, which ranks potential guide RNA based on an on-target efficiency score and off-target score. Three guides were tested in parallel per knockout.

Table 3 - Primer pairs used for the creation of C. carboxidivorans knockout vectors in this study.

Primer Name	Sequence
	pMTL8315y-v1_SynCas9_Ccar_03565

Left homology arm	*Forward:* CGAGTCGGTGCTTTTTTTTAGACGTCAATATATTAAGATACTAAATGATATAGTAGAAGGAAAGG *Reverse:* TTACATAATAAAATGCAGTCGCCATTTTATCACCCTTTTAAACCC
Right homology arm	*Forward:* GACTGCATTTTATTATGTAAGGTAAAGATGAATACACAAGAGATAGTAAGTAAACTTTGG *Reverse:* ATAACAAGTATTTTTTATTGGCGCGCCCTAGATCTTTCATCTTCTTACCTTCATTAGTTAGTG
Guide RNA	*Forward:* TTACCCCGTATCAAAATTTGTCGACTTAAAATCAGAAAACAGCATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC

	Pmtl8315y-v1_SynCas9_Ccar_24300
Left homology arm	*Forward:* CGAGTCGGTGCTTTTTTTTAGACGTCTAGCCACTTTCAACTTTTATAACATATTC *Reverse:* TCCAGTGACAGTAATATCAGAAATCATTTCTCACTTGCCTCC
Right homology arm	*Forward:* ATTTCTGATATTACTGTCACTGGATAAGAATGGTGAATTAAATGCACAG *Reverse:* TAACAAGTATTTTTTATTGGCGCGCCCTTGAAAAAGACTCCTATTTGAGC
Guide RNA	*Forward:* TTTATCTGTTACCCCGTATCAAAATTTGTCGACGTATATGGCGAATACTACGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC

	pMTL8315y-v1_SynCas9_Ccar_07120
Left homology arm	Forward: ACCGAGTCGGTGCTTTTTTTTAGACGTCTGGAGTTAATACAATAGACTGTGATG *Reverse:* TCAGTGATAGCCTGATCTTCACCATAAATATTACCTTTCTTTATGCTATATTTTCATAAC
Right homology arm	*Forward:* TGAAGATCAGGCTATCACTGAGGTAAAATAAAATAGCCTTGAATTTAATCTTGGC

	Reverse: ATAACAAGTATTTTTTATTGGCGCGCCCTATAACGTAGATTTAGATGGTCAATATAATTACAAATTAG
Guide RNA	*Forward:* TTACCCCGTATCAAAATTTGTCGACAGAATATTAGATCCTTGCATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC

	pMTL8315y-v1_SynCas9_Ccar_23805
Left homology arm	*Forward:* CGAGTCGGTGCTTTTTTTTAGACGTCGCTGTTCCTGCTACTACACTACAAATC
	Reverse: TTATTTCGTAAACGGTATCGGTCATATTAAACCTTCTTAATTTTATAAATTGTTTACAAACAC
Right homology arm	*Forward:* CGATACCGTTTACGAAATAAGGTAAAAGTTTAACTATTAAAGGATTTACTGATGAATATATTTGC
	Reverse: ATACATAACAAGTATTTTTTATTGGCGCGCCCTATACAAGTAGCCCCAAATACTGAAGC
Guide RNA	*Forward:* TTTATCTGTTACCCCGTATCAAAATTTGTCGACATATTGTAGGAAATCCCCCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC

	pMTL8315y-v1_SynCas9_Ccar_08700
Left homology arm	*Forward:* CGAGTCGGTGCTTTTTTTTAGACGTCGAAAAATTTATAATTCACAATTTAAGCACATGG
	Reverse: TCTTTCTACAGCCGACTGCATCCATTTTATCAAGCTTTAGTATTTATGTTATAGC
Right homology arm	*Forward:* TGCAGTCGGCTGTAGAAAGAGGTAGCTACCATGACTTTAGTGCC
	Reverse: AAGTATTTTTTATTGGCGCGCCAAAATATACAAAAAGGAGAAGCTTGG
Guide RNA	*Forward:* TTACCCCGTATCAAAATTTGTCGACTCTTCTATATCCTCTCCTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC

	pMTL8315y-v1_SynCas9_Ccar_16055

Left homology arm	*Forward:* CGGTGCTTTTTTTTAGACGTCTGCTGAACTTTTACAAATAGGC *Reverse:* CCGGAGCTCCGATAAAAAATGGTAAAGAAATATGCCTAATTGCTTTGTAG
Right homology arm	*Forward:* ATTTTTTATCGGAGCTCCGGACCATTATAACATCCCTTCCTTTAATG *Reverse:* AAGTATTTTTTATTGGCGCGCCAAGTAAAAGAAATATATGAAAAAGTTTGATG
Guide RNA	*Forward:* TTACCCCGTATCAAAATTTGTCGACGCACCTGTATACATTCCTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC

	Other primers
Guide RNA	*Universal Reverse*: AAAAAAAGCACCGACTCG
Ccathlguide_F	*Sanger Sequencing Forward:* GTTGTTAGAGAAAACGTATAAATTAG
modpCB102_R	*Sanger Sequencing Reverse:* GTCGGTACATTTGAAATATTG

2.4.9. Construction of ClosTron vectors

The pMTL007C-E2::Cbut_PyrE-381;382s plasmid was synthesised by ATUM Bio (USA) and was used in the construction of all other ClosTron vectors in this study.

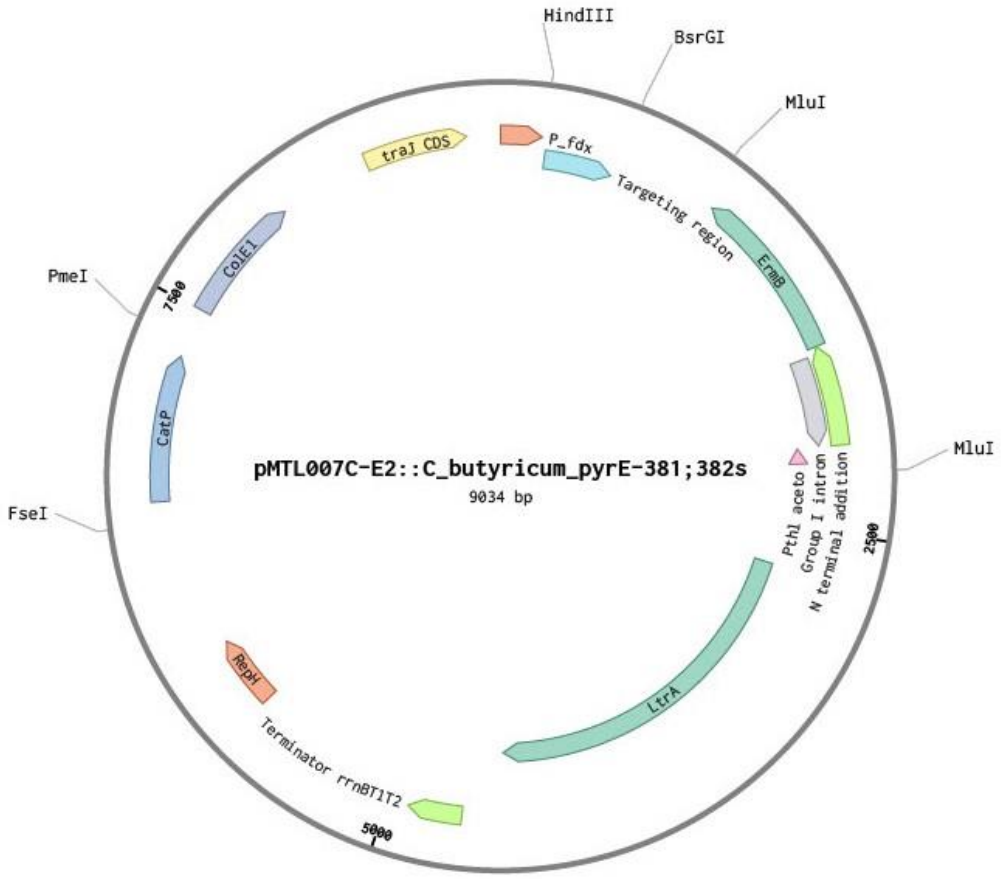

Figure 7 - Plasmid map of ermB::RAM ClosTron vector.

The ClosTron vector pMTL007E-C::Cbut_PyrE-381;382s was created by first swapping the *catP* antibiotic marker from pMTL007C-E2::C_but_PyrE-381;382s an *ermB* marker by performing a restriction digest with the enzymes *Fse*I and *Pme*I to create a linear fragment of DNA without the *catP* marker. A DNA fragment containing an *ermB* marker was created by PCR using pMTL83251 as a template with the primers ErmB_ClosBB_F and ErmB_ClosBB_R and these were purified from an agarose gel and used in a NEBuilder Hifi Assembly reaction. After incubation the reaction mixture was used to transform *E. coli* 10-beta, which was plated on LB agar plates supplemented with erythromycin. Once pickable colonies had formed eight of these were picked to inoculate liquid cultures. The following day the plasmid was purified and sent for sanger

sequencing with the forward sequencing primer pCB102-F2 and the reverse sequencing primer ColE1-R1 to ensure no SNPs were created during cloning. Remaining purified plasmid was stored at -20 °C for further cloning from when the sequencing results arrived.

Once the sequencing was completed a positive plasmid was taken out of storage and another restriction digest was performed to remove the DNA sequencing containing the *ermB*::RAM using the enzyme *Mlu*I. The backbone was again purified from an agarose gel and a NEBuilder hifi reaction was performed on this DNA fragment along with the *catA*::RAM ordered as a synthesised DNA fragment. The reaction mixture was transformed into *E. coli* 10-β after incubation and plated onto clarithromycin supplemented LB agar plates. Once colonies had formed these were screened for successful insertions via colony PCR using the primers CatARAM_R. Positives would be screened further via sanger sequencing as previously described using the primer CatARAM_R.

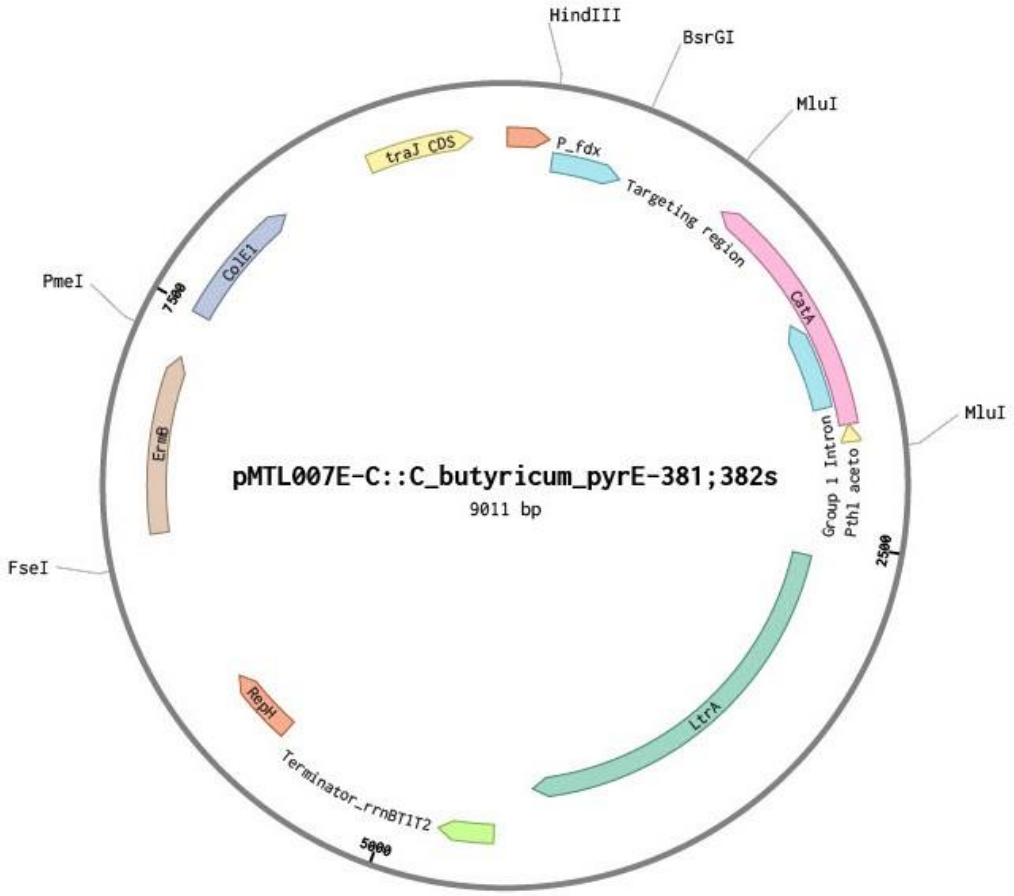

Figure 8 - Plasmid map of catA::RAM ClosTron vector.

To retarget *catA*::RAM ClosTron vector a digestion reaction of using enzymes *Hind*III and *Bsr*GI to remove the targeting region from the vector and inserting the new one via Hifi assembly reaction.

Table 4 - Primers and gene fragments used in the construction of ClosTron vectors.

Primer Name	Sequence
	ermB Fragment
ErmB_ClosBB_F	ATTGTTATGGATTATAAGCGGCCGGCCGAAGCAAACTTAAGAGTG
ErmB_ClosBB_R	TCATGAGATTATCAAAAAGGAGTTTAAACACATTCCCTTTAGTAACG
	Targeting regions

C. sporogenes PyrE 470;472s	ATAAAGTTGTGTAATTTTTAAGCTTATAATTATCCTTAGAAGTCGTTT CGGTGCGCCCAGATAGGGTGTTAAGTCAAGTAGTTTAAGGTACTA CTCTGTAAGATAACACAGAAAACAGCCAACCTAACCGAAAAGCGA AAGCTGATACGGGAACAGAGCACGGTTGGAAAGCGATGAGTTAC CTAAAGACAATCGGGTACGACTGAGTCGCAATGTTAATCAGATATA AGGTATAAGTTGTGTTTACTGAACGCAAGTTTCTAATTTCGATTACT TCTCGATAGAGGAAAGTGTCTGAAACCTCTAGTACAAAGAAAGGT AAGTTATGCGAAACGACTTATCTGTTATCACCACATTTGTACAATCT GTAGGAGAACCTATG
C. carboxidivorans PyrE 123;124s	ATAAAGTTGTGTAATTTTTAAGCTTATAATTATCCTTAGCAGACGGG GTAGTGCGCCCAGATAGGGTGTTAAGTCAAGTAGTTTAAGGTACT ACTCTGTAAGATAACACAGAAAACAGCCAACCTAACCGAAAAGCG AAAGCTGATACGGGAACAGAGCACGGTTGGAAAGCGATGAGTTA CCTAAAGACAATCGGGTACGACTGAGTCGCAATGTTAATCAGATAT AAGGTATAAGTTGTGTTTACTGAACGCAAGTTTCTAATTTCGATTTC TGCTCGATAGAGGAAAGTGTCTGAAACCTCTAGTACAAAGAAAGG TAAGTTATCTACCCCGACTTATCTGTTATCACCACATTTGTACAATCT GTAGGAGAACCTATG
C. difficile R20291 *PyrE* 1 414;515s	ATAAAGTTGTGTAATTTTTAAGCTTATAATTATCCTTAGGAGACGTT GTAGTGCGCCCAGATAGGGTGTTAAGTCAAGTAGTTTAAGGTACT ACTCTGTAAGATAACACAGAAAACAGCCAACCTAACCGAAAAGCG AAAGCTGATACGGGAACAGAGCACGGTTGGAAAGCGATGAGTTA CCTAAAGACAATCGGGTACGACTGAGTCGCAATGTTAATCAGATAT AAGGTATAAGTTGTGTTTACTGAACGCAAGTTTCTAATTTCGATTTC TCCTCGATAGAGGAAAGTGTCTGAAACCTCTAGTACAAAGAAAGG TAAGTTACCTACAACGACTTATCTGTTATCACCACATTTGTACAATCT GTAGGAGAACCTATG
C. difficile R20291 *PyrE* 2 213;214s	ATAAAGTTGTGTAATTTTTAAGCTTATAATTATCCTTAGCTATCGGTG GAGTGCGCCCAGATAGGGTGTTAAGTCAAGTAGTTTAAGGTACTA CTCTGTAAGATAACACAGAAAACAGCCAACCTAACCGAAAAGCGA AAGCTGATACGGGAACAGAGCACGGTTGGAAAGCGATGAGTTAC CTAAAGACAATCGGGTACGACTGAGTCGCAATGTTAATCAGATATA AGGTATAAGTTGTGTTTACTGAACGCAAGTTTCTAATTTCGATTATA GCTCGATAGAGGAAAGTGTCTGAAACCTCTAGTACAAAGAAAGGT AAGTTAACTCCACCGACTTATCTGTTATCACCACATTTGTACAATCT GTAGGAGAACCTATG

catA with group I intron

catA with group I intron	CAGATATTTATTACGTGGCGACGCGTGAAGTTCCTATACTTTCTAGA GAATAGGAACTTCCTAAAAAATACAGCTGTTTGGGTGGTGAGCTA ACTCTTGAACAGAGTTCATAAATAGACCTGCGTGGTAACCGTCGCA TACTGCATGATGCAATTGTAGGCTAATAGGAAGGTAAATCAAATTT CCTTGGTGTATCAATCTACCAGCAGTTATTATAGGAAGCAAATATCT TGAATTGTTATTTATATTTAAGTTAAAAGATGTGAATGAAGTCCATG GTATAACGCTGAATGAAAAAGTATTCTCAGGAACAGGTGTCTTAG GAAATAGAGATCCTGATCCGTTGTATTTTTCAACATCTGCTGTGTAT GCTCTATGGAAAGTCTCGAAGTCATTAGTTACACTAGTCCAAACTG CGCTAAAACTCTCTGATCTCCTGTCAAAAATAGTGTATAGAGGCTCT AATCTGTCCCAGTATCCTAAGTCTCCCTCACAATTATAACTAGTTCTA AAAGCAGGGTGGCTGTTTACTACAGTAGTAATCAAGAATATAAGTG CAGGAGTGAACCTATACTTATGTTGTTTTATGAATCTGTATAACACA GAAACATCTATCTCTTGGGTTAATTGAGGCCTGAGTATAAGGTGAC TTATACTTGTAATCTATCTAAACGGGGAACCTCTCTAGTAGACAATC CCGTGCTAAATTGTAGGACTGCCCTTTAATAAATACTTCTATATTTAA AGAGGTATTTATGAAAAGCGGAATTTATCAGATTAAAAATACTTTCT CTAGAGAAAATTTCGTCTGGATTAGTTACTTATCGTGTAAAATCTGA TAAATGGAATTGGTTCTACATAAATGCCTAACGACTATCCCTTTGGG GAGTAGGGTCAAGTGACTCGAAACGATAGACAACTTGCTTTAACA AGTTGGAGATATAGTCTGCTCTGCATGGTGACATGCAGCTGGATAT AATTCCGGGGTAAGATTAACGACCTTATCTGAACATAATGCAGGCT AAATGAAGTGTTCTGATTCATATAGTGATTAAACACCTCAGTCCTAT TCCAAACATCAAAGTCAATCCTGTTGAACTTCATATGAATCCCTCCT AATTTATACGTTTTCTCTAACAACTTAATTATACCCACTATTATTATTT TTATCAATATAGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCAC GCGTTGGGAAATGGCAATGATAGC
	Sanger sequencing
pCB102_F2	GTCAAGTATGAAATCATAAATAAAG
ColE1_R1	CTTTTCTACGGGGTCTGAC
Clostron_screen_R	TTTTCAAGCTCTAGTGCTATAGC
ColE1_+tra_F2	CCATCAAGAAGAGCGAC
CatARAM_R	ACGATCTAGCCATATATGTATATCTCC

2.5. RiboCas gene knockout generation in *C. carboxidivorans*

Knockouts were generated using the RiboCas system (Cañadas *et al.*, 2019) modified for use in *C. carboxidivorans* by Redfern (2021).

First, the knockout vector was constructed as per **2.4.8.** and transformed into both *E. coli* sExpress T_II and *E. coli* sExpress_TIII.

These strains were then used to conjugate the knockout vector into *C. carboxidivorans* (**2.3.3.**) with the addition of 10mM of L-arabinose to the mating plates and all liquid cultures of sExpress T_II or sExpress_TIII derived strains.

Once individual colonies had formed on the selection plates, these were streaked to single colonies onto new YTAF plates supplemented with d-cycloserine, thiamphenicol, and theophylline.

After 72 hours pickable colonies would form and these were screened via colony PCR using primers that annealed up and down stream of the target gene using wild type *C. carboxidivorans* gDNA as a negative control. Colonies with a DNA band lower than the control on the visualised gel, indicating a successful knockout, were picked to inoculate 3 mL of YTAF MES, which was used to create bacterial stocks and genomic DNA extraction.

A high-fidelity PCR was performed using the genomic DNA as a template with the same primers as in the initial screening to amplify the knockout region. The resultant DNA bands were excised and purified for sanger sequencing to check

for complete deletion of the gene and bookmark insertion using the same primers.

Once sanger sequencing was complete one of the positive mutant strains was chosen for plasmid curing (**2.3.5**). Once plasmid loss was confirmed these strains were stored as bacterial stocks.

Table 5 - Primers for cPCR mutant screening and sanger sequencing when generating knockouts.

Primer Name	Sequence
	Ccar_23805
Ccar_23805_F	TCATTTTGAGAAATACCTGTGG
Ccar_23805_R	CTACAAGTGCTACAAACAATCC
	Ccar_24305
Ccar_24300_F	AGCATAAGTGAAAAAATAACAATAAGG
Ccar_24300_R	AATTAACTTACGCTGATGAAGG
	Ccar_08700
Ccar_08700_F	TAATGATATACAAAATACCTTCTCCATC
Ccar_08700_R	TAATAGAAGGTACTAATGACTATAGAGG
	Ccar_07120
Ccar_07120_F	ATTATAAGCTGTGTATTAAGGATAATCTTG
Ccar_07120_R	ATGGTGCATTACTGATTGG
	Ccar_03565

Ccar_03565_F	TGGCAAAGAACTTAAAGAGTTAGG
Ccar_03565_R	TGCAAGAAATTTAACTGCTTATTTG
	Ccar_16055
Ccar_16055_F	TGAAGATGATGAAACTGTGC
Ccar_16055_R	ATGCGGAGAAACTTTTGC

2.6. ClosTron mutagenesis

ClosTron vectors created as per **2.4.8.** were transformed into chemically competent *E. coli* sExpress and conjugated into their target organism. Once colonies had formed on selection plates, eight were picked to inoculate 1 mL of the species appropriate media supplemented with antibiotic selecting for the plasmid backbone in sterile 2 mL screw top tubes. Once these cultures were in late exponential-stationary phase, 100 µL of each culture was spread onto separate agar plates supplemented with the RAM specific antibiotic to select for successful integrants. Once colonies had formed on the RAM-selection plates these were screened for successful intron insertion at *PyrE* via colony PCR using primers that flank the *PyrE* locus. Colonies that produced a DNA band ~2 kb higher than the wild-type control, indicating intron insertion, after running on agarose gel were used to inoculate 3 mL of liquid media to create stocks for storage at -80 °C. These stocks were revived for plasmid curing (**2.3.5**) and the subsequent plasmid cured strains were stored as bacterial stocks.

Table 6 - Primers for screening Clostron mutants via cPCR.

Primer Name	Sequence
	C. butyricum PyrE
Cbut_pyrE_F	atgattgaatgtggagtgttg
Cbut_pyrE_R	attgagaatagtaatcatcaattcttgc
	C. sporogenes PyrE
Cspo_pyrE_F	caaaattaaatctattttgagacactcc
Cspo_pyrE_R	atgtaatagagtttattatggcagg
	C. carboxidivorans PyrE
Ccar_pyrE_F	aaagatatcctgttgtcaataacc
Ccar_pyrE_R	aagatttatgtgaaataagaggaatcg
	C. difficile* R20291 *PyrE
Cdif_pyrE_F	tttagacgaaataagaggaattatataagg
Cdif_pyrE_R	gatttttcatacttacccctcc

2.6.1. ClosTron *PyrE* mutant characterisation

PyrE inactivation from intron insertion was validated by testing mutants for uracil auxotrophy. This was performed by patching wild type and plasmid-cured mutants onto minimal media agar plates (**table 5**) supplemented with or without 0.25 mg of uracil.

Table 7 - Minimal media agar for uracil auxotrophy assay.

Component	Per Litre
Minimal media agar	
Amino acid solution	400 mL
Salt Solution	200 mL
20 % w/v Glucose solution	100 mL
Trace salt solution	40 mL
$FeSO_4 \cdot 7H_2O$	20 mL
Vitamin solution	20 mL
Agar No. 1 (Oxoid, UK)	15 g
dH_2O	220 mL
Amino acid solution	
Cas-amino acids	50 g
Tryptophan	2.5 g
L-Cysteine	2.5 g
Salt solution	
KH_2PO_4	9.0 g
NaCl	9.0 g
Na_2HPO_4	50 g
$NaHCO_3$	50 g
Trace salt solution	
$CaCl2 \cdot 2H_2O$	1.30 g
$MnCl2 \cdot 4H_2O$	0.50 g
$CoCl2 \cdot 6H_2O$	0.05 g
$MgCl_2 \cdot 6H_2O$	1.00 g
$(NH_4)_2SO_4$	2.00 g
$FeSO_4 \cdot 7H_2O$ Solution	
$FeSO_4 \cdot 7H_2O$	0.4 g
Vitamin solution	
Ca-D-panthothenate	0.1 g
Pyridoxine	0.1 g

Biotin	0.1 g

2.7. Characterisation of *C. carboxidivorans* CO_2 and H_2 fermentations

2.7.1. Semi-defined medium and serum bottle preparation.

The characterisation of strains of *C. carboxidivorans* produced in this study grown on CO_2 and H_2 was carried out using a semi-defined medium. This was P11 medium as described in Thunuguntla *et al.* (2024). The components of which were as follows:

Table 8 - P11 media recipe.

Component	**Per Litre**
P11 Media	
NH_4Cl	2.50 g
KCl	0.25 g
KH_2PO_4	0.25 g
$MgSO_4 \cdot 7H_2O$	0.50 g
$CaCl_2 \cdot 2H_2O$	0.10 g
Yeast Extract	0.5 g
MES	10 g
Vitamin solution	10 mL
Trace metal solution	10 mL
Reducing solution	10 mL
dH_2O	970 mL

Vitamin Solution 100x	
Pyridoxine	0.010 g
Thiamine	0.005 g
Riboflavin	0.005g
Calcium pantothenate	0.005 g
Thioctic acid	0.005 g
p-(4)-Aminobenzoic Acid	0.005 g
Nicotinic acid	0.005 g
Vitamin B12	0.005 g
Biotin	0.002 g
Folic Acid	0.002 g
2-Mercaptoethanesulfonic acid	0.010 g

Trace metal solution 100x	
Nitrilotriacetic acid	2.00 g
$MnSO_4 \cdot H_2O$	1.00 g
$Fe(NH_4)_2(SO_4)_2.6H_2O$	0.80 g
$CoCl_2 \cdot 6H_2O$	0.20 g
$ZnSO_4 \cdot 7H_2O$	1.00 g
$NiCl_2 \cdot 6H_2O$	0.20 g
$Na_2MoO_4 \cdot 2H_2O$	0.20 g
Na_2SeO_4	0.10 g
$Na_2WO_4 \cdot 2H_2O$	0.20 g

Reducing Solution 100x	
L-cysteine	20 g
$Na_2S \cdot 9H_2O$	20 g

The base media was first made without yeast extract, reducing solution, vitamin solution, and trace metal solution, and adjusted to pH 6.1. The resazurin stock solution were added and the resulting mixture was left in the anaerobic cabinet for at least five days for most of the oxygen to diffuse out. 250 mL serum bottles were then left in the same anaerobic cabinet for at least one day to remove any adsorbed oxygen and 25 mL of media was decanted into them. Once the media

was completely to near completely anoxic, as indicated by the colour change from resazurin, a rubber stopper was then inserted into the serum bottle that were then sealed with an aluminium crimp. The bottles were immediately autoclaved after sealing to destroy any potential contaminating microorganisms. After autoclaving, the vitamin, trace metal, yeast extract, and reducing stock solutions were added using a 1 mL syringe fitted with a 0.22 µM filter and a 0.5 mm outer diameter (25 gauge) needle. Addition of reducing agents was performed the day prior to inoculation to allow time for full reduction of the media.

2.7.2. Serum bottle inoculation, culture conditions, and sampling.

Before inoculation the headspace of the serum bottles was replaced with a gas mixture of 80% H_2 and 20% CO_2 using a gas exchange unit (Anaerobic Lap Gaschange System, GR Instruments BV, Netherlands). 1.6 bar of H_2 was added followed by CO_2 to a final pressure of 2 bar.

To inoculate the serum bottles, a stock of the strain of the *C. carboxidivorans* to be characterised was revived in 3 mL of YTAF media. Once this had reached exponential phase, 30 µL of this was used to inoculate 3 mL of P11 media with 10 g/L of fructose added as a carbon source. This culture was used to inoculate the serum bottles to OD 0.05 once it had reached exponential phase (~1 OD).

The purpose of this first round of growth on gas was to adapt the strain from heterotrophic growth on fructose to autotrophic growth. Once gas was being

utilised by the strain for growth, indicated by pressure drop and OD_{600} increase, the culture from the serum bottles was used to inoculate a second round of serum bottles.

The second serum bottles were again incubated at 25 °C and agitated at 200 RPM with samples taken regularly to record OD_{600} and pH. These samples were also used to analyse production of acetate, butyrate, hexanoate, ethanol, butanol, and hexanol production via HPLC-MS. The internal pressure of the bottles was measured using a digital pressure gauge (DPG 110, Omega Engineering, USA) alongside a "blank" serum bottle pressurised at 2 bar containing 25 mL of water that had the same volume drawn from it as in the culture sampling process. The purpose of this was to determine if drop in pressure was attributed to the increased volume of the headspace after each sampling.

2.8. Plasmids and other chemically synthesised DNA used in this study

2.8.1. Chapter 3

Table 9 - Plasmids used in chapter 3 for the genetic domestication of C. carboxidivorans.

Name	Description	Source
pMTL8315y-v1	Siteless pMTL80000 plasmid developed by Redfern (2021)	Elizabeth Redfern

pMTL8315y-v1_SynCas9_Ccar_03565	Siteless synthetic Cas9 vector developed by Redfern (2021). Altered to create in-frame deletion of Type I restriction nuclease Ccar_03565.	This study
pMTL8315y-v1_SynCas9_Ccar_24300	Siteless synthetic Cas9 vector developed by Redfern (2021). Altered to create in-frame deletion of Type I restriction nuclease Ccar_24300.	This study
pMTL8315y-v1_SynCas9_Ccar_07120	Siteless synthetic Cas9 vector developed by Redfern (2021). Altered to create in-frame deletion of Type II restriction nuclease Ccar_07120.	This study
pMTL8315y-v1_SynCas9_Ccar_23805	Siteless synthetic Cas9 vector developed by Redfern (2021). Altered to create in-frame deletion of Type II restriction nuclease Ccar_23805.	This study
pMTL8315y-v1_SynCas9_Ccar_08700	Siteless synthetic Cas9 vector developed by Redfern (2021). Altered to create in-frame deletion of Type III restriction nuclease Ccar_08700.	This study
pMTL83151	Standard pMTL modular vector. pCB102, ColE1+traJ, c*atP,* MCS.	SBRC culture collection

2.8.2. Chapter 4

Table 10 - Plasmids used in chapter 4 for the metabolic engineering of C. carboxidivorans Δ7RM.

Name	Description	Source
pMTL8315y-v1_SynCas9_Ccar_16055	Siteless synthetic Cas9 vector developed by Redfern (2021). Altered to create in-frame deletion of hydrogenase *HytA* Ccar_03565.	This study

2.8.3. Chapter 5

Table 11 - Plasmids used in chapter 5 for the development of the catA::RAM.

Name	Description	Source
pMTL83251	Standard pMTL80000 modular vector. pCB102, ColE1+traJ, ErmB, MCS.	SBRC culture collection
pMTL83251-Pthl-CatA	pMTL shuttle vector constitutively expressing CatA from *Paenibacillus silvae*	This study
pMTL83251-Pthl-TetM-Tn916	pMTL shuttle vector constitutively expressing TetM from Tn916	This study
pMTL83251-Pthl-TetM-Sau	pMTL shuttle vector constitutively expressing TetM from multiple *Staphylococcus aureus* strains (WP_025642386)	This study
pMTL007C-E2::Cbut_PyrE-381;382s	*ermB*::RAM ClosTron vector targeting *PyrE* in *C. butyricum*	Atum Bio, US
pMTL007E-C::Cbut_PyrE-381;382s	*catA*::RAM ClosTron vector targeting *PyrE* in *C. butyricum*	This study
pMTL007E-C::Cspo_PyrE-470;472s	*catA*::RAM ClosTron vector targeting *PyrE* in *C. sporogenes*	This study

pMTL007E-C::Ccar_PyrE-123;124s	*catA*::RAM ClosTron vector targeting *PyrE* in *C. carboxidivorans*	This study
pMTL007E-C::Cdif_PyrE-414;515s	*catA*::RAM ClosTron vector targeting *PyrE* in *C. difficile*	This study
pMTL007E-C::Cdif_PyrE-213;214s	*catA*::RAM ClosTron vector targeting *PyrE* in *C. difficile*	This study
pMTL007E-C::Cdif_Spo0A-177;178a	*catA*::RAM ClosTron vector targeting *Spo0A* in *C. difficile*	This study

Chapter 3 – Genetic Domestication of *C. carboxidivorans*

3.1. Introduction

As outlined in **1.6.**, *Clostridium carboxidivorans* is an acetogenic organism capable of converting inorganic carbon in the form of carbon dioxide and carbon monoxide into ethanol, butanol, and hexanol. This means *C. carboxidivorans* has great potential to act as a biological chassis for gas fermentation, producing biofuels from waste carbon as part of a circular bioeconomy. Moreover, *C. carboxidivorans'* innate ability for chain elongation could be exploited to synthesise organic molecules of higher carbon content than what is currently possible in gas fermentations. As such, establishing synthetic biology tools for the organism and engineering a genetically domesticated strain would further develop gas fermentation as a technology for commodity chemical production.

The main obstacle to achieving these goals is that *C. carboxidivorans* possesses an extensive Restriction Modification System (RMS), consisting of three type I, two type II, two type III, and three type IV systems (Kottenhahn *et al.*, 2023). Their existence makes the delivery of heterologous DNA challenging by conventional means, impeding any direct metabolic engineering.

To eliminate this barrier, a strategy was devised to evade the RMS of *C. carboxidivorans* to deliver CRISPR-cas9 based knockout vectors targeting RMS associated nuclease-encoding genes and thereby generate a genetically

domesticated strain. The key steps of this strategy are discussed in **1.3.** and comprise: (i) determining the nucleotide sequences and methylation patterns targeted by the RMS; (ii) developing an "RMS silent" vector that does not contain these nucleotide motifs; (iii) devising a means to deliver this vector into the cell (i.e. via conjugation or electroporation); (iv) creating a methylation donor strain to aid in vector delivery, and; (v) developing a suitable knockout method for nuclease-encoding gene deletion.

The genetic domestication of *C. carboxidivorans* in this project is the continuation of Dr. Elizabeth Redfern's work as part of a PhD thesis (Redfern, 2021). In Dr. Elizabeth Redfern's work, the key milestones of creating a RMS motif silent vector, establishing DNA transfer to *C. carboxidivorans* with this vector, creating a methylation donor to aid in DNA transfer, and knocking out a key nuclease were achieved.

3.1.1. Characterising the RMS of *C. carboxidivorans*

To evade an RMS, it is first necessary to know which restriction recognition sites are targeted for removal from the vector being introduced into the cell. These were revealed by performing PacBio SMRT sequencing (SMRTseq) of *C. carboxidivorans'* genomic DNA and then supplying the data to Dr Rich Roberts at NEB, who provided a list of probable and possible RMS genes and their methylation specificity. This service was provided at a private webpage at the Restriction Enzyme Database (REBASE) hosted by NEB. Analysis revealed that *C. carboxidivorans* possesses three type I, three type II, two type III, and three type IV systems (**figure 9**).

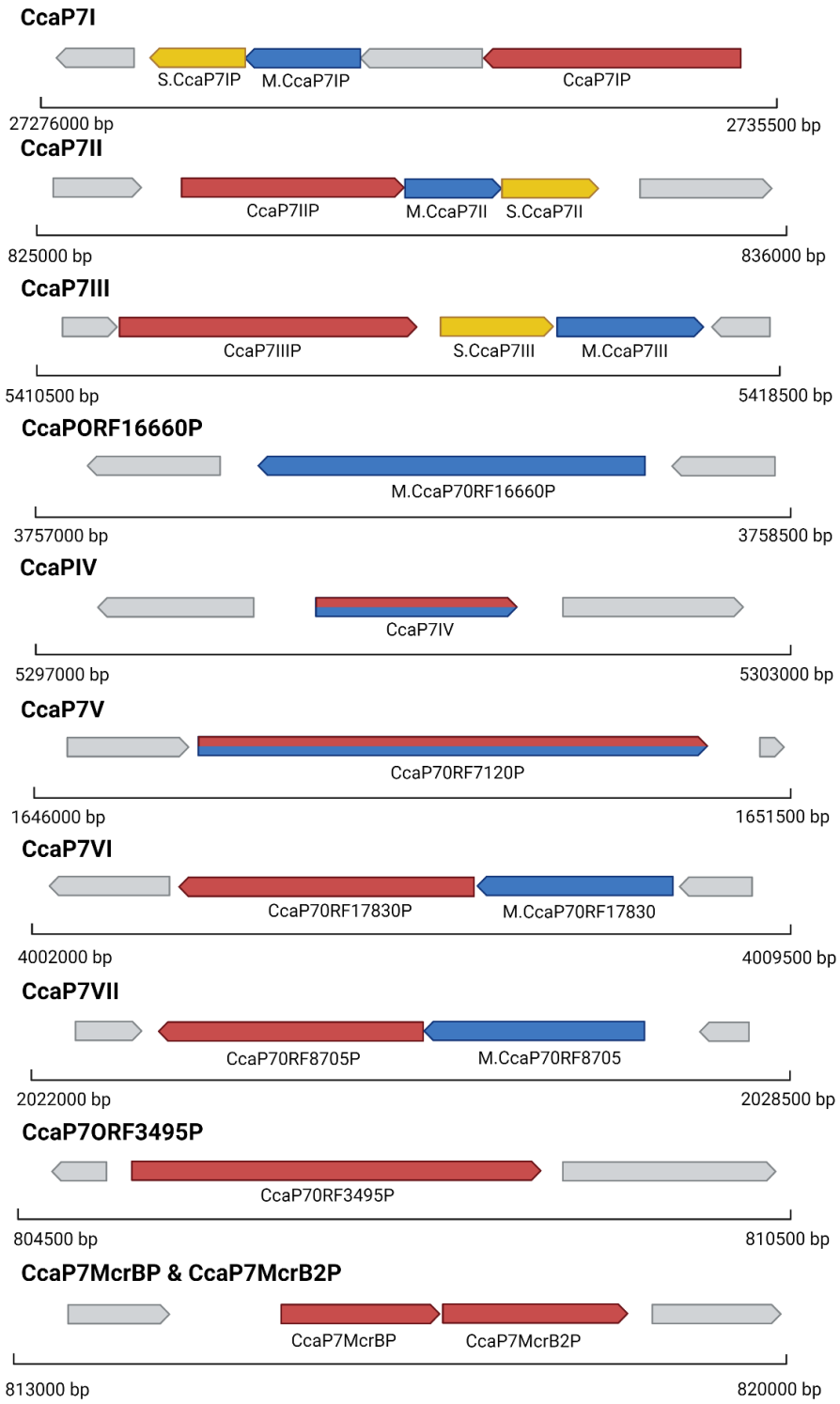

Figure 9 - Genetic layout of the Nucleases (red), Methyltransferases (blue), and specificity subunits (yellow) of C. carboxidivorans' RMSs. Created with BioRender.com.

Further analysis was performed with the PacBio SMRTseq data that, as described in **1.3.1**, allows for mapping of methylated DNA in parallel with

nucleotide sequence to determine the methylome of an organism by identifying recurring motifs of methylation in DNA sequences. Six different methylation motifs were identified as part of *C. carboxidivorans'* genome (**table 12**), all of which featured an adenosine residue methylated to N6-methyldeoxyadenosine (6mA).

Table 12 - Methylation motifs found in the genome of C. carboxidivorans via PacBio SMRT sequencing. *Underlined bases indicate methylation.*

Methylated sequence	RMS type	Methylation Type
C<u>A</u>YNNNNNCTGC	I	6mA
CC<u>A</u>NNNNNNNNTCG	I	6mA
GC<u>A</u>NNNNNNNTNNCG	I	6mA
AGA<u>A</u>GC	II/III	6mA
CRAAA<u>A</u>R	II/III	6mA
GATA<u>A</u>T	II/III	6mA

The protein sequences identified by the REBASE pipeline were then aligned against the REBASE Gold Standard Database, a curated collection of RMS enzymes with experimentally verified recognition sequences. If no recognition sequence could be assigned to an RMS using this method, further analysis was performed against the full REBASE database through manual alignments to infer possible recognition sequences (**table 13**).

Table 13 - The different RMSs of C. carboxidivorans and their assigned PacBio SMRTseq derived recognition sequences determined through REBASE analysis. *(*) Indicates recognition sequence allocations from Kottenhahn et al. (2023). Underlined bases indicate methylation.*

RMS Name	Type	Recognition Sequence	Locus Tags
CcaP7I	I gamma	CAYNNNNNCTGC	Nuclease: Ccar_11975 Methyltransferase: Ccar_11965 Specificity: Ccar_11960
CcaP7II	I gamma	CCANNNNNNNNTCG	Nuclease: Ccar_03565 Methyltransferase: Ccar_03570 Specificity: Ccar_03575
CcaP7III	I gamma	GCANNNNNNNTNNCG	Nuclease: Ccar_24300, Methyltransferase: Ccar_ Ccar_24305, Specificity: Ccar_ Ccar_24310
CcaP7ORF16660P	II	CTSAG*	Methyltransferase: Ccar_16660
CcaP7IV	II-G Alpha	GATAAT	Fused system: Ccar_23805
CcaP7V	II-G Alpha	CRAAAAR*	Fused system: Ccar_07120
CcaP7VI	III beta	AGAAGC*	Nuclease: Ccar_17825, Methyltransferase: Ccar_17830
CcaP7VII	III beta	GAAAT*	Nuclease: Ccar_08700, Methyltransferase: Ccar_08705
CcaP7McrBP	IV		Nuclease: Ccar_03525
CcaMcrB2P	IV		Nuclease: Ccar_03520

CcaP7ORF03495	IV	Nuclease: Ccar_03495

All type I systems were confidently assigned a methylation motif identified through PacBio SMRT sequencing. It was found that the target recognition domain in the specificity subunit of CcarP7III had a good similarity to that of S.Spy322I and S.spy8198I, whose recognition sequences are GCANNNNNNRTTG and GCANNNNNNTTAA, making it a strong candidate for the recognition sequence GCANNNNNNNTNNCG. The same analysis was performed on the specificity subunit of CcarP7II revealing a strong homology to S.BceSVI and S.BmeD34I, which have the recognition sequences CCANNNNNNNCTTA and CCANNNNNNNCTC, respectively. Indicating that the most likely target sequence for this RMS is CCANNNNNNNNTCG. Due to a process of elimination, the remaining unassigned Type I recognition sequence CAYNNNNNCTGC was assigned to CcaP7I.

Out of the remaining methylation patterns only GATAAT could be confidently assigned to CcaPIV due to strong homology of the target recognition domain to M.CboAf650IV and M.CboBDIII, which also recognise GATAAT.

Since this work was completed, a separate study was published by Kottenhahn *et al.* (2023) who carried out further characterisation of the RMSs of *C. carboxidivorans* in a similar manner. In addition to the PacBio SMRTseq and REBASE pipeline, Kottenhahn *et al.* (2023) inserted each of the methyltransferase encoding genes of *C. carboxidivorans* into a pCDFDuet-1

expression vector for the *in vivo* methylation of *E. coli*. Following expression, each strain of *E. coli* was analysed via PacBio SMRT sequencing to determine the methylation pattern each methyltransferase is responsible for.

The recognition site allocations are consistent with that of Redfern (2021) with some additional findings. The first of which is attributing the sequence CRAAAAR to the RMS CcarP7V, with the associated methyltransferase of this system producing this methylation pattern when expressed in *E. coli*.

Another was allocating the recognition site AGAAGC to CcaP7VI. In the previous work carried out by Redfern (2021), this sequence was allocated to the same RMS based on homology to the methyltransferases MXne.XORFEP (which target AGACC) M.XneXORFF, and M.Sba6781 (which both target AAGACC). Since homology was only found in methyltransferases that recognise similar sequences, this allocation was made tentatively. Kottenhahn *et al.* (2023) attributed this sequence to CcaP7VI due to all other methyltransferases being accounted for, despite its associated methyltransferase not producing any DNA methylation when expressed in *E. coli*.

The methylation motif GAAAT was found to be produced by the methyltransferase from the RMS CcaP7VII when heterologously expressed in *E. coli* despite this motif not appearing in the genomic DNA of *C. carboxidivorans* when analysed via PacBio SMRT sequencing.

Finally, the orphan (lacking associated nuclease) methyltransferase CcaP7ORF16660P was assigned the target site CTSAG due to several entries in REBASE that share homology with this protein recognising this site. This

methylation motif was not found in the genome of *C. carboxidivorans* nor in *E. coli* when expressed on a plasmid. The authors postulate that the methyltransferase was rendered non-functional at some point after its associated nuclease was lost, removing the selection pressure for CTSAG methylation.

Neither body of work determined the recognition sequences of the type IV systems.

3.1.2. Creation of RM silent vectors

With most of the recognition sites of *C. carboxidivorans'* RM systems determined, it was possible for Redfern (2021) to design a minimal vector that lacked these motifs in a bid to bypass them.

The starting point of the vector design was based on the pMTL80000 series of shuttle plasmids that are specifically designed for research in *Clostridium* (Heap *et al.*, 2009). These vectors feature a four-part modular design consisting of a Gram-positive replicon, a Gram-negative replicon, a selection marker, and an application specific module (such as a multiple cloning site). Each module was assessed for the frequency at which recognition sites occurred within their sequences and a complete vector was decided considering the number of sites they contained and desired function.

Final modules selected were: The Gram-positive replicon pCB102 because it contained only three recognition sites, the Gram-negative replicon conjugative element containing ColE1 + traJ, *catP* as a selection module, and a multiple

cloning site as the application specific module. Under the pMTL80000 nomenclature system each module is assigned a number forming the name of the plasmid with the chosen vector therefore being pMTL83151.

pMTL83151 contains sixteen recognition sites in total. Two counts of CAYNNNNNCTGC, two of GCANNNNNNNTNNCG, three of AGAAGC, seven of CRAAAAR, and two of GATAAT. Twelve of these sites were removed by PCR amplification of DNA fragments from pMTL83151 followed by Hifi DNA assembly to create the vector pMTL8315y-v1.

The four remaining sites were contained within the ColE1 + traJ module, one of which was found in the ColE1 RNA II pre-primer. These four sites were removed in the same manner as before to create two new vectors, pMTL8315y-v2 and pMTL8315y-v3. The difference between these two vectors was the nucleotide base change made to remove the recognition site within the ColE1 RNA II pre-primer region. In pMTL8315y-v2 the 16th base thymine was replaced with adenine, and in pMTL8315y-v3 the 14th base adenine replaced with a thymine.

To ensure that these alterations, particularly to the Gram-positive replicon, did not render the vectors non-functional, conjugal transfer of each was attempted into *C. sporogenes*. All produced transconjugant colonies with pMTL8315y-v2 displaying almost three-fold higher conjugation efficiency when compared to the unmodified vector.

3.1.3. Conjugative transfer of pMTL8315y-v1 into *C. carboxidivorans*

Prior to the creation of pMTL83151-v2 and pMTL8315y-v3, conjugative transfer of pMTL8315y-v1 into *C. carboxidivorans* was attempted using *E. coli* sExpress as a conjugal donor. This yielded fourteen transconjugant colonies, however, when this was repeated using all three vectors no transconjugant colonies were formed after multiple attempts.

3.1.4. Isolation of the highly competent strain C94

Due to the repeated failure in reproducing conjugative transfer of any of the RM silent vectors, an alternative strategy was by Redfern (2021). Given mutant sub-populations can exist within wild type cultures, provided that their phenotype does not have any deleterious effects, it was hypothesised one such mutant sub-population amenable to DNA transfer existed within the initial culture used and was responsible for the observed successful DNA transfer. With this in mind, a stock of one of the transconjugant strains obtained was revived, and the plasmid was cured via repeated re-culturing without antibiotic pressure until thiamphenicol sensitivity was restored.

The receptivity to conjugal transfer of DNA into this strain, dubbed C94, was tested with all three variants of the RM silent vectors constructed. Interestingly, only DNA transfer experiments with pMTL8315y-v2 produced transconjugant colonies. Importantly, this result could be reliably repeated producing on

average 282.3 transconjugant colonies. This is perhaps due to the differences in ColE1 RNA II pre-primer region.

3.1.5. Mimicking the native methylation profile

As discussed in **1.3.1.**, the mimicking of the methylation patterns recognised by an organism's RMS can be a highly effective method to deliver DNA to otherwise genetically recalcitrant organisms by methylating vectors *in vivo*. This is particularly important in cases where the modification of bases to avoid recognition sites is impossible, an example of which being the donor region of a CRISPR-cas based knock-out vector. Donor regions contain sequences of DNA that must be homologous to the regions flanking the target gene and whilst some sites can be avoided by truncating these regions (called homology arms) this comes at the cost of reduction in knock-out efficiency. Moreover, recognition sites can be contained within the extreme 3' end of the upstream homology arm or in the 5' end of the downstream homology arm, making truncation impossible if a clean in-frame deletion is desired. Mimicking the native methylation profile of an RMS allows these sites to remain in the plasmid whilst not impeding DNA transfer.

The strategy employed by Redfern (2021) was to modify R702, a plasmid found in our conjugal donor strains that facilitates horizontal transfer of co-resident plasmids whilst being defective in transferring itself, to contain the methyltransferase encoding genes of *C. carboxidivorans* under the control of the L-arabinose inducible promoter P_{BAD}. This allows for the *in vivo* methylation of shuttle vectors during the standard conjugation protocol with the only extra

step being the addition of L-arabinose to the growth media of the donor strain. Additionally, this would avoid any complications during conjugative DNA transfer that could arise under a three-plasmid system where *in vivo* methylation is mediated by a third, separate, plasmid.

The modification of R702 was performed by first assembling two small vectors. The first of which contained the methyltransferase genes of the type II systems and their respective ribosomal binding sites whose expression was controlled by the P_{BAD} promoter, and the second plasmid was the same except for the type III methyltransferases of *C. carboxidivorans* replacing the type II. Downstream of the methyltransferase encoding genes was an apramycin resistance cassette (*aadA1*). These were flanked upstream and downstream by 500 bp homology arms that would act as a template for homologous recombination at the *aadA1* locus of the R702 plasmid.

These plasmids were used as a template for the amplification of a single linear DNA fragment of the methylation cassette and homology arms via PCR for the type II and type III methyltransferases. This DNA fragment was transformed into a competent strain of *E. coli* called InterStellar, a R702 containing derivative of the ClonTech *E. coli* strain, Stellar. This strain expresses the λ-red recombineering machinery genes *exo, beta,* and *gam,* which allows for *in vivo* modification of plasmids within the cell via homologous recombination with transformed fragments of DNA. Successful recombination events of the methylation cassette were selected for by apramycin supplementation of media and were further confirmed by Sanger sequencing. The now methyltransferase

encoding modified R702 plasmids were subsequently extracted and transformed into NEB Express to create two different conjugal donor strains, one containing type II methyltransferases and the other type III, called sExpress_TII and sExpress_TIII. This approach was also attempted with the type I methyltransferases and associated specificity subunits without success.

The conjugal efficiency of these strains was assessed under the hypothesis that, prior to conjugation, *in vivo* methylation of the shuttle vector would occur following the induction of expression of the methyltransferase cassette that would allow for the shuttle vector to bypass the relevant *C. carboxidivorans* RM systems and therefore an increase in conjugation efficiency would be observed (**figure 10**). Conjugation of the RM silent vectors pMTL8315y-v1, pMTL8315y-v2, and pMTL8315y-v3 as well as the standard pMTL83151 plasmid into the C94 strain was attempted. Interestingly, when using the methylating donor strains only pMTL8315y-v1 resulted in transconjugant colonies whereas before only pMTL8315y-v2 could be successfully conjugated using the standard sExpress conjugal donor strain. When comparing the conjugation efficiency of the methylating strains to sExpress, sExpress_TII resulted in a 4-fold increase and sExpress_TIII produced a 9-fold improvement suggesting that *in vivo* methylation of shuttle vectors allowed for some evasion of *C. carboxidivoran's* RM systems.

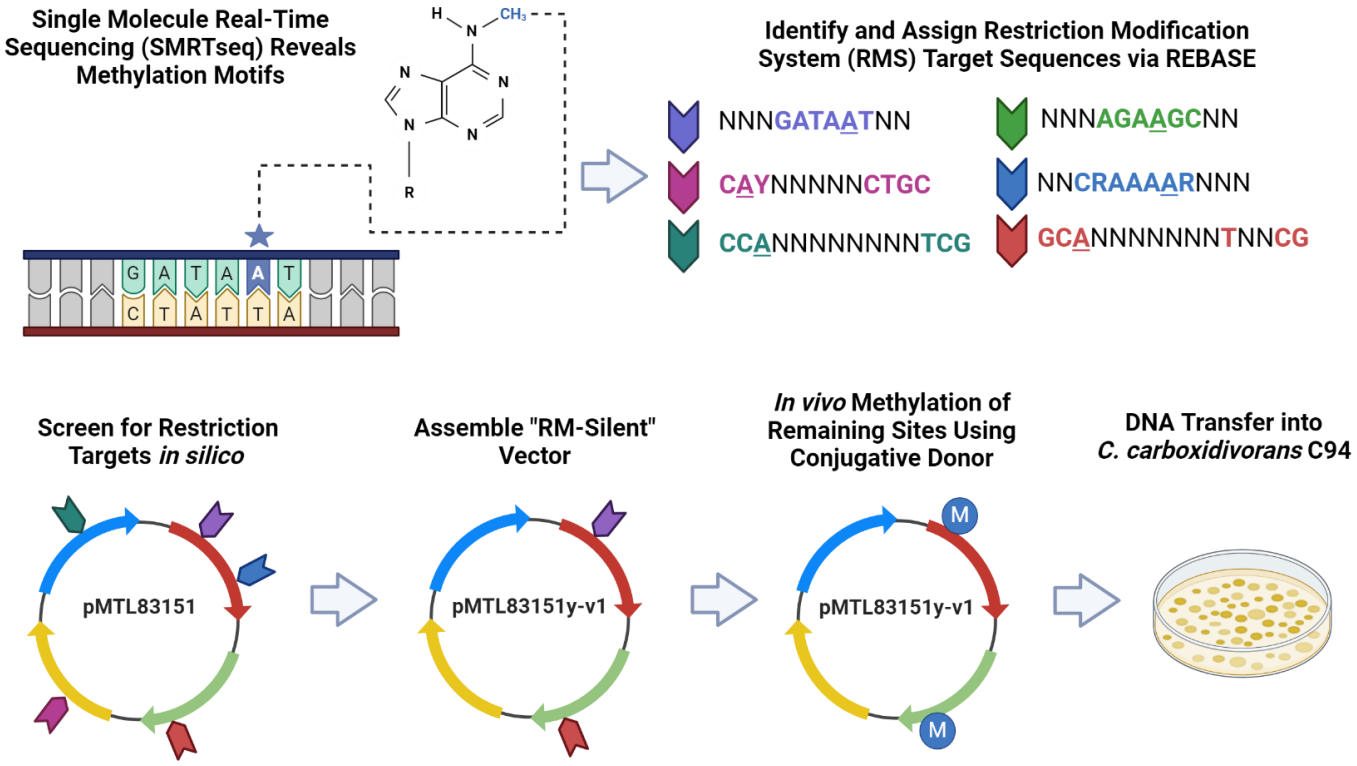

Figure 10 - The strategy employed to enable DNA transfer into C. carboxidivorans.

The methylome of C. carboxidivorans is determined by PacBio SMRT sequencing. These methylation motifs correspond to recognition targets of nucleases in C. carboxidivorans, assigned through comparison with the NEB REBASE database. This is followed by the screening and removal of the same sites within the vector that is to be delivered. Recognition sites that could not be removed are then appropriately methylated by a methyltransferase harbouring conjugal donor strain. This same strain is used to deliver the RM silent vector to the highly competent C. carboxidivorans C94 via conjugation. Created with BioRender.com.

3.1.6. Development of genome modification tools for *C. carboxidivorans.*

With the successful creation of an RM silent vector, methylation donor strains, and means of DNA delivery, the next step in the genetic domestication of *C. carboxidivorans* was to develop a tool that would allow for genomic modification with the intent to delete RMS associated nuclease encoding genes.

The CRISPR-Cas9 based RiboCas system developed by Cañadas *et al.* (2019) (**1.5.1.**) was chosen as a knock-out tool due its efficacy in the genetic modification of other members of *Clostridium*. The key modifications made for compatibility with *C. carboxidivorans* was to alter the Cas9 encoding gene to remove any RM recognition sites through silent mutations, exchanging the plasmid backbone with pMTL8315y-v1, and replacing the P_{araE} promoter that controls transcription of the guide RNA with the constitutive thiolase promoter sourced from *Clostridium acetobutylicum* ATCC 824. This created the plasmid pRECas(pthI)-v1-TIII.2R.

Using the methylating conjugal donor strains sExpress_TII and sExpress_TIII, it was possible to deliver the modified RiboCas CRISPR-Cas9 vectors to the C94 strain to create an in-frame deletion of the RMS associated nuclease encoding gene of the CcaP7VI Ccar_17825 (creating strain *C. carboxidivorans* Δ1RM) by Elizabeth Redfern and the subsequent deletion of the Type I system CcaP7I

Ccar_11975 (creating strain *C. carboxidivorans* Δ2RM) by Christopher Humphries.

3.1.7. Aims

Having established that a CRISPR-based KO vector could be delivered into *C. carboxidivorans* and its functionality established through the deletion of a type I and type III restriction enzyme gene. The creation of a genetically domesticated strain, in which all the restriction genes of all the type I, II and III *C. carboxidivorans* RMSs were deleted, was now feasible. The aim of this chapter was therefore to: (i) sequentially delete the five remaining restriction genes in the existing double mutant, and (ii) to assess the effects on the frequency of conjugative plasmid transfer.

Where possible, only the nuclease element of an RMS was targeted for deletion. This is because methyltransferases can perform a regulatory function in gene expression (Zhou *et al.*, 2021). Their deletion, therefore, could have unintended deleterious effects. In the event of a RMS nuclease being part of a Type II fused system where methylation and nuclease activity is governed by a protein encoded by a single gene, as is the case in CcaP7IV and CcaP7V, the entire gene was targeted for knock out. Type IV systems were not targeted as these only inhibit the uptake of incorrectly methylated DNA and as such can be bypassed by using conjugal *E. coli* donor strains deficient in the methyltransferases *dcm.*

3.2. Results

CRISPR-Cas vectors designed for the deletion of RMS associated nucleases were constructed as described in **2.4.8.** and knockouts were generated using these vectors as described in **2.5.** Each knockout vector used in this study contained a unique 24 nucleotide "bookmark" sequence as cargo to be delivered to the target locus. These bookmark sequences consist of a 20 nucleotide protospacer followed by a 3 nucleotide protospacer adjacent motif (PAM) as developed by Seys *et al.* (2020). This provides a targeting region for the reinsertion of nuclease encoding genes at the same locus as they were removed, allowing for future complementation studies whereby the knocked-out gene is re-inserted to restore the phenotype and thereby allow further characterisation of these nucleases. Additionally, these bookmarks provide convenient insertion points for genes that could be beneficial for alcohol production via gas fermentation in future research with this strain. The bookmarks inserted for each knockout vector as well as the guide RNAs that generated successful knockouts are shown in **table 14**. The specific sequences of these bookmarks can be found in Seys *et al.* (2020). The nuclease-encoding genes associated with CcarP7VI and CcarP7I were knocked out by Elizabeth Redfern and Christopher Humphreys, respectively, the former was replaced with no bookmark sequence and the latter was replaced with the bookmark sequence "BM4" (GAGGGTTGTGGGTTGTACGGAAGG).

Table 14 - Guide RNA sequences used for targeting RMS associated nuclease with modified RiboCas vectors. Target genes are listed in order of deletion in this study.

Vector Name	RMS Name	Target Gene	gRNA Sequence	Bookmark (BM)
pMTL8315y-v1_SynCas9_Ccar_23805	CcaP7IV	Ccar_23805	atattgtaggaaatccccct	BM6
pMTL8315y-v1_SynCas9_Ccar_24300	CcaP7III	Ccar_24300	gtatatggcgaatactacgc	BM5
pMTL8315y-v1_SynCas9_Ccar_08700	CcaP7VII	Ccar_08700	tcttctatatcctctcctga	BM11
pMTL8315y-v1_SynCas9_Ccar_07120	CcaP7V	Ccar_07120	agaatattagatccttgcat	BM7
pMTL8315y-v1_SynCas9_Ccar_03565	CcaP7II	Ccar_03565	ttaaaatcagaaaacagcat	BM12

3.2.1 CcaP7IV (Ccar_23805) Knockout

The first nuclease encoding genes knocked out belonged to the fused type-II RMS CcaP7IV with the recognition sequence GATAAT **(figure 11**) creating the strain *C. carboxidivorans* Δ3RM. The guide RNA that produced nuclease deficient mutants had a knockout efficiency of 56.25%.

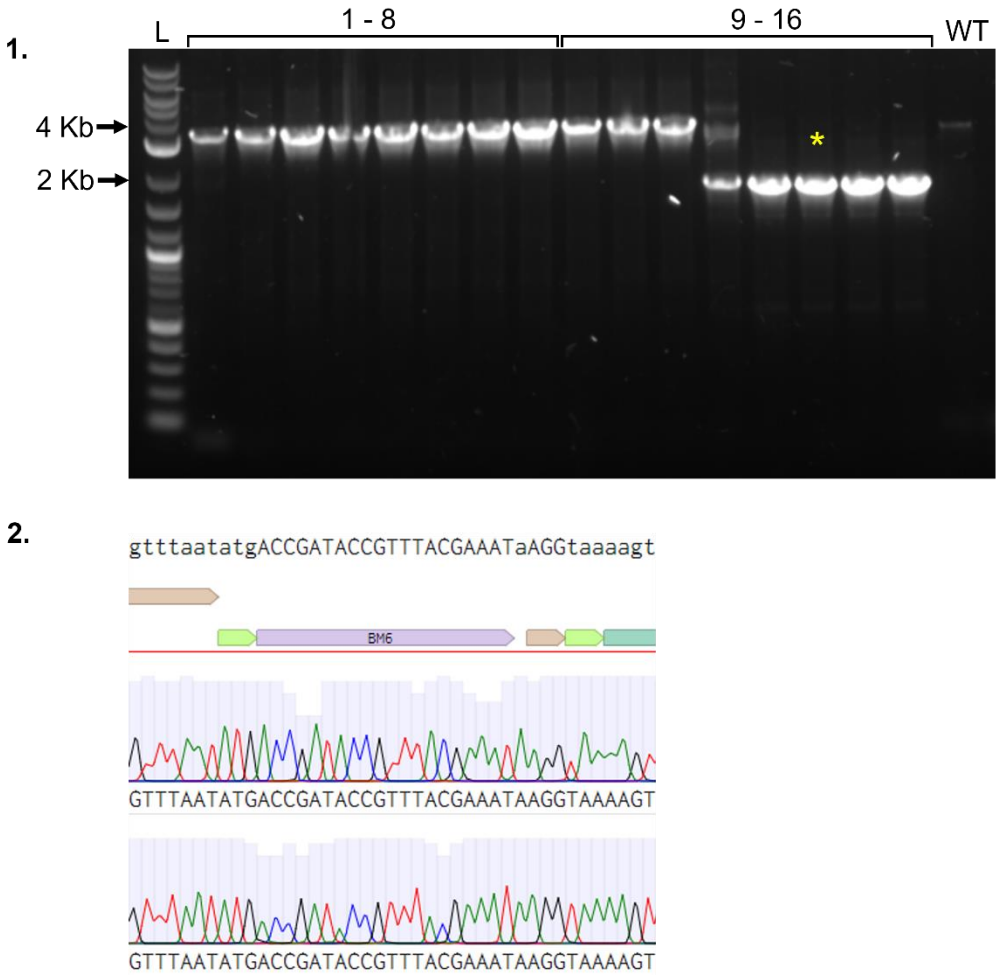

Figure 11 - 1: Agarose gel electrophoresis cPCR of colonies formed on theophylline induction plates using primers Ccar_23805_F and Ccar_23805_R to screen for Ccar_23805 deficient mutants.

1-8 – guide RNA 2. 9-16 guide RNA 3. L – DNA ladder Quick-Load 1 kb Plus (New England Biolabs, USA). WT - wild type. Mutant selected for further modification indicated by (*). 2: Sanger sequencing of knockout locus confirming bookmark insertion.

3.2.2. CcaP7III (Ccar_24300) Knockout

The second nuclease encoding gene deleted belonged to the type-I RMS CcaP7III with the recognition sequence GCANNNNNNNTNNCG **(figure 12)** creating the strain *C. carboxidivorans* Δ4RM. The guide RNA that produced nuclease deficient mutants had a knockout efficiency of 37.5%.

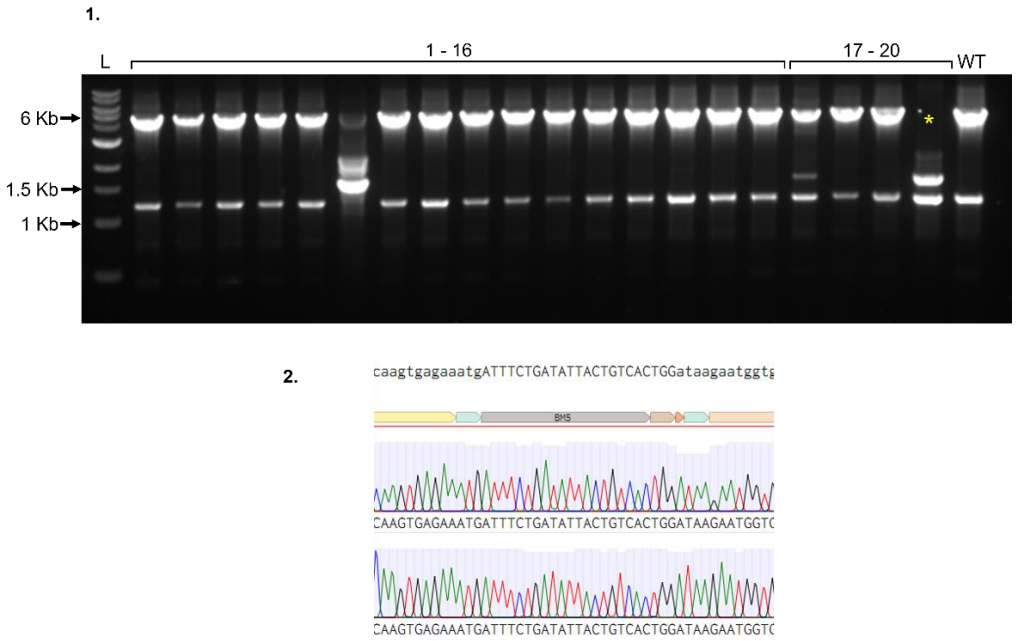

Figure 12 - 1: Agarose gel electrophoresis cPCR of colonies formed on theophylline induction plates using primers Ccar_24300_F and Ccar_24300_R to screen for Ccar_24300 deficient mutants.

1-16 – guide RNA 2. 17-20 guide RNA 3. L – 1kb DNA ladder (New England Biolabs, USA). WT - wild type. Mutant selected for further modification indicated by (*). 2: Sanger sequencing of knockout locus confirming bookmark insertion.

3.2.3. CcaP7VII (Ccar_08700) Knockout

The third nuclease encoding gene knocked out belonged to the type-III RMS CcaP7VII with the recognition sequence GAAAT **(figure 13)** creating the strain *C. carboxidivorans* Δ5RM. The guide RNA that produced nuclease deficient mutants had a knockout efficiency of 17.4%

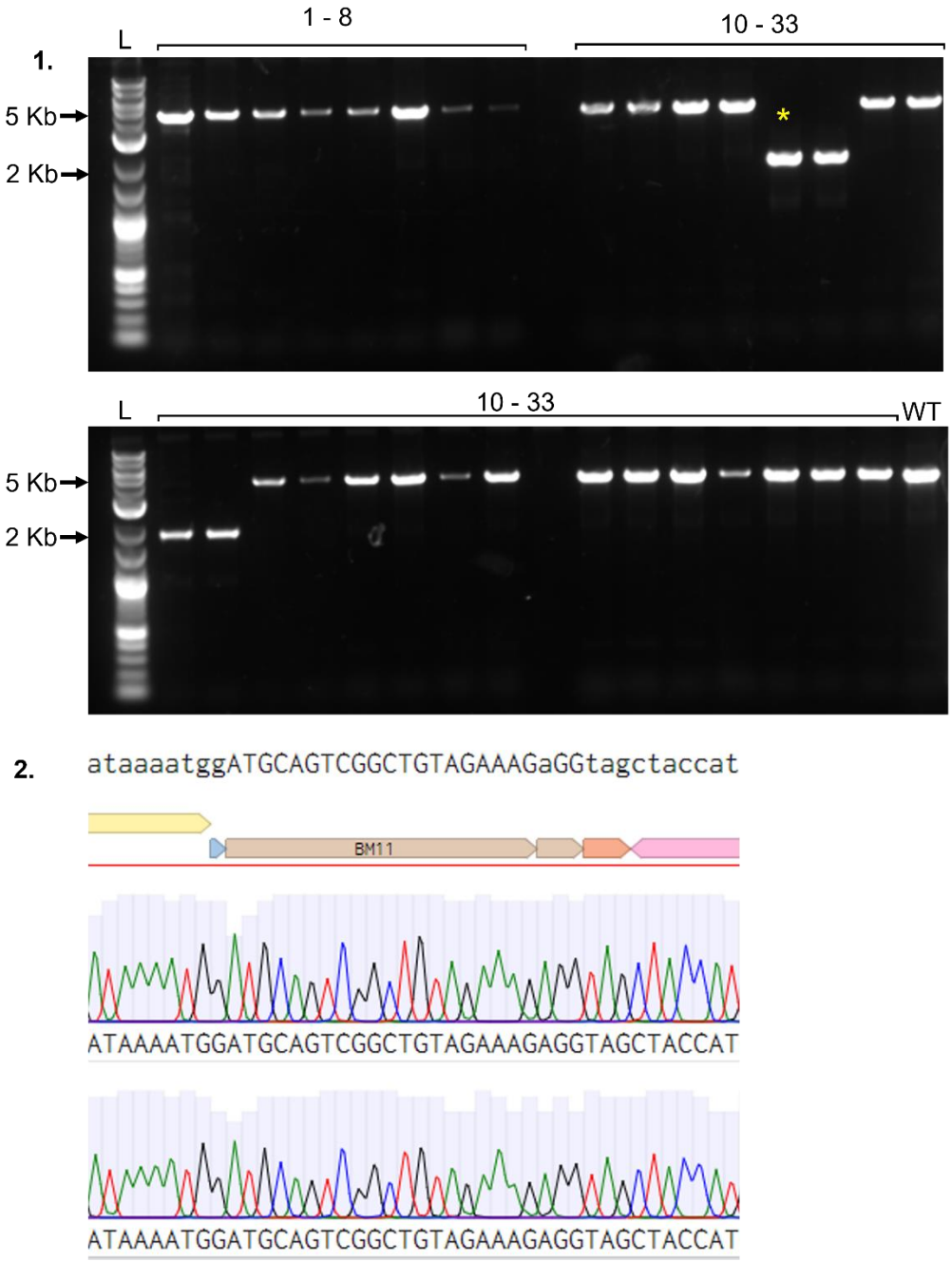

Figure 13 - 1: Agarose gel electrophoresis cPCR of colonies formed on theophylline induction plates using primers Ccar_08700_F and Ccar_08700_R to screen for deficient Ccar_08700 mutants.

1-8 – guide RNA 1. 10-33 guide RNA 3. L – Quick Load 1kb plus DNA ladder (New England Biolabs, USA). WT - wild type. Mutant selected for further modification indicated by (*). 2: Sanger sequencing of knockout locus confirming the insertion of the bookmark.

3.2.4. CcaP7V (Ccar_07120) Knockout

The fourth nuclease encoding gene deleted out belonged to the fused type-II RMS CcaP7V with the recognition sequence CRAAAAR **(figure 14)** creating the strain *C. carboxidivorans* Δ6RM. Of the colonies screened via cPCR only one produced banding on an agarose gel. This colony happened to be a nuclease deficient mutant and so this vector appears to have a knockout efficiency of 100% but this unlikely to be the case.

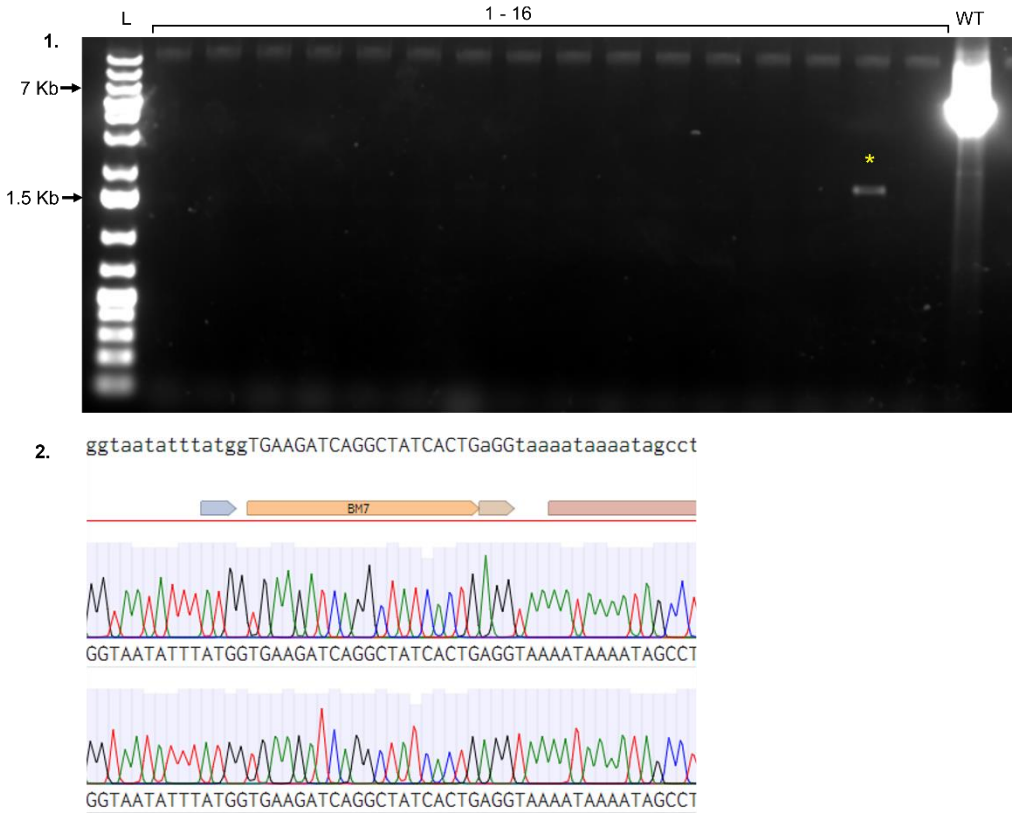

Figure 14 - 1: Agarose gel electrophoresis cPCR of colonies formed on theophylline induction plates using primers Ccar_07120_F and Ccar_07120_R to screen for deficient Ccar_07120 mutants.

1-16 – guide RNA 2. L – GeneRuler 1kb plus DNA ladder (Thermo Scientific, USA). WT - wild type. Mutant selected for further modification indicated by (*). 2: Sanger sequencing of knockout locus confirming the insertion of the bookmark.

3.2.5. CcaP7II (Ccar_03565) Knockout

The fifth and final nuclease encoding gene knocked out belonged to the fused type-I RMS CcaP7II with the recognition sequence CCANNNNNNNNTCG **(figure 15)** creating the final strain *C. carboxidivorans* Δ7RM. The guide RNA that produced nuclease deficient mutants had a knockout efficiency of 100%.

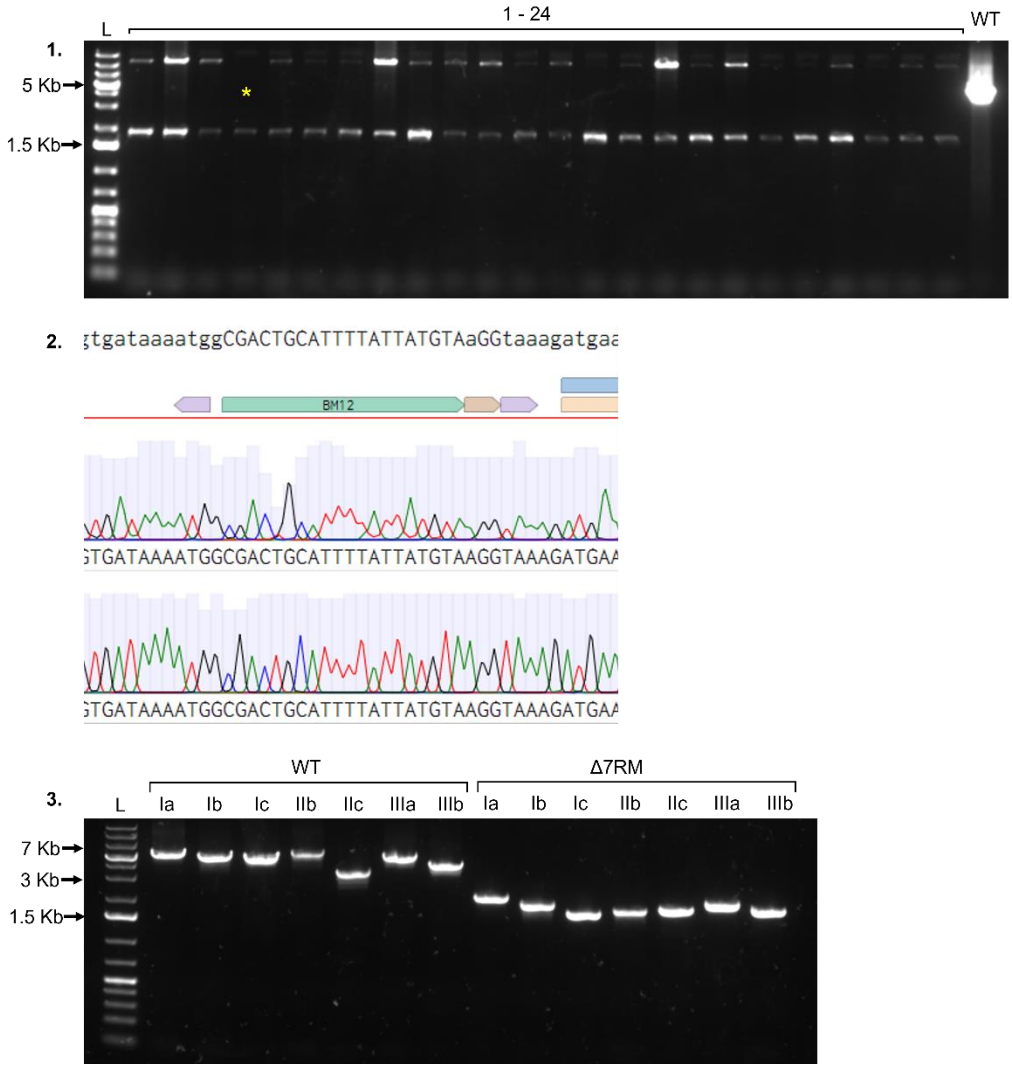

Figure 15 - Agarose gel electrophoresis cPCR of of colonies formed on theophylline induction plates screen for Ccar_07120 mutants and PCR of knockout loci in C. carboxidivorans Δ7RM. *1-16 – guide RNA 2. L – GeneRuler 1kb plus DNA ladder (Thermo Scientific, USA). WT - wild type. Mutant selected for further modification indicated by (*). 2: Sanger sequencing of knockout*

locus confirming the insertion of the bookmark. 3: Agarose gel electrophoresis gDNA PCR using the same primers as when screening for nuclease deficient mutants. The amplified locus of each nuclease knocked out in C. carboxidivorans Δ7RM (right) compared with the nuclease loci of the wildtype (left). L = Generuler 1 Kb plus DNA ladder (Thermo Scientific, USA) Ia = CcaP7I, Ib = CcaP7II, Ic = CcaP7III, IIb = CcaP7V, IIc = CcaP7IV, IIIa = CcaP7VI, IIIb = CcaP7VII.

3.2.6. Conjugation efficiency of the Domesticated Strain

With all type I, II, and II RMS associated nucleases deleted, the amenability of DNA transfer via conjugation was assessed for the final strain as well as each intermediate knockout stage. This was performed as described in **2.3.4. using** sExpress as a conjugal donor strain to deliver pMTL83151, a standard plasmid part of the pMTL80000 series of vectors. Each conjugation was performed in technical duplicate and biological triplicate and the results of which are shown in **figure 16**.

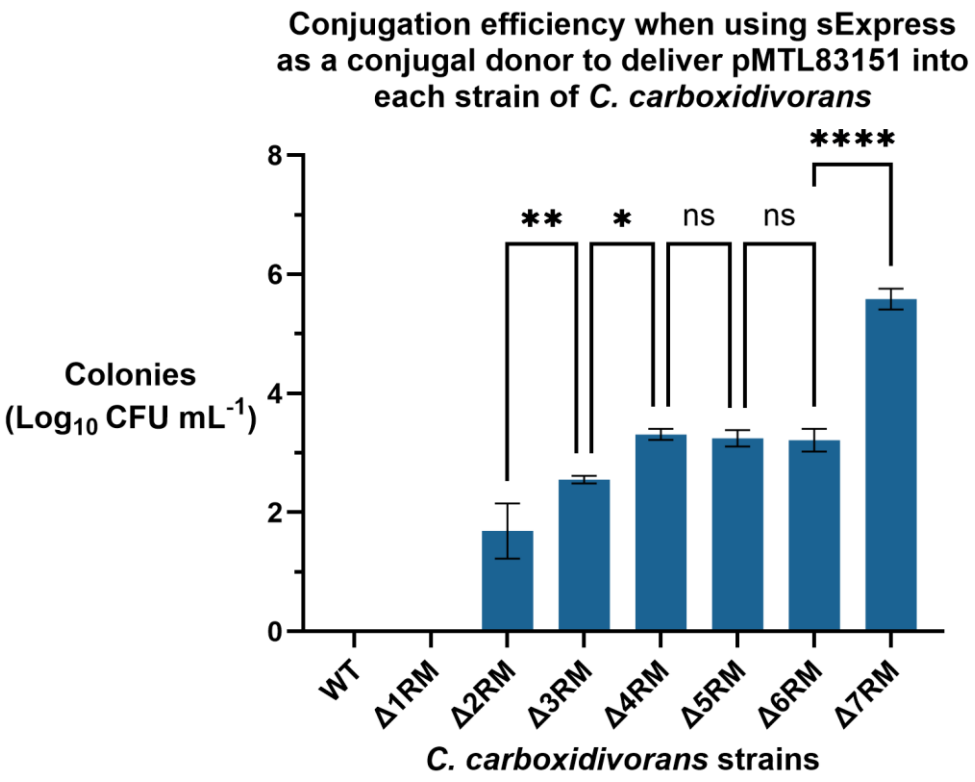

Figure 16 - The mean conjugation efficiencies of each intermediate strain and the final domesticated strain (C. carboxidivorans Δ7RM).

The data shows a $\log_{10}$ transformation of the number of transconjugant colonies formed per millilitre of mating plate slurry transferred to selection plates. sExpress was used as a donor strain to deliver the vector pMTL83151. Conjugations were performed in technical duplicate and biological triplicate. * = $P \leq 0.05$, ** = $P \leq 0.01$, **** = $P \leq 0.0001$, ns = $P > 0.05$. CFU = Colony Forming Units. Error bars represent standard deviation.

For statistical analysis a log transformation of the data was performed followed by a one-way ANOVA and a *post hoc* Tukey's test. Statistically significant differences were found between strains Δ2RM and Δ3RM (P = 0.0059), strains Δ3RM and Δ4RM (P = 0.0158), and strains Δ6RM and Δ7RM (P = <0.0001). No statistically significant differences were found between strains Δ4RM, Δ5RM, and Δ6RM.

The strain Δ1RM is a strain deficient in the gene encoding the nuclease associated with the RMS CcaP7VI and was produced in previous work by Redfern (2021). The strain Δ2RM was created by Christopher Humphreys and is a further modified version of Δ1RM with the nuclease encoding gene associated the type-I RMS CcaP7I deleted.

3.2.7. Illumina Sequencing of *C. carboxidivorans* Δ7RM Genome

The development of the *C. carboxidivorans* Δ7RM strain required repeated sub-culturing, which can result in unintended changes to the genome through the accumulation of Single Nucleotide Polymorphisms (SNPs) acquired by random mutation events. Moreover, the Cas9 knockout system has the potential to produce off-target effects altering the host cell's genome. As such, it was important to perform full genome sequencing of the final strain to ensure that no deleterious mutations exist. This is particularly important in an organism that is intended for an industrial process as key enzymes in relevant product forming pathways may be affected. In the case of *C. carboxidivorans*, these mutations could occur in the genes encoding enzymes governing the Wood-Ljungdahl pathway and solvent synthesis, rendering the final strain unfit for purpose.

Extracted and purified genomic DNA (**2.4.4.**) of *C. carboxidivorans* Δ7RM was subjected to Illumina full genome sequencing by DeepSeq (Nottingham, UK). The resulting sequencing data was mapped against the genome sequence of the starting strain C. *carboxidivorans* C94 created by Redfern (2021) to ensure

that any mutations identified could be attributed to the domestication process conducted in the study.

Table 15 – Summary of variants in open reading frames of the Illumina genomic sequence data of C. carboxidivorans Δ7RM when mapped against the genomic sequence of C. carboxidivorans C94.

Locus Tag	Position	Base change	Coverage (%)	Frequency (%)	Average Quality	Predicted Amino Acid Change	Gene Product
Ccar_00380	71463	G>T	57	100	32.89473684	p.Arg696Ile	Alanyl-tRNA synthetase
Ccar_00445	84711	G>T	81	100	33.83950617	p.Gly313Val	Twitching motility protein PilT
Ccar_00510	98186	G>T	82	100	35.54878049	p.Glu223*	Phosphate transport system permease protein
Ccar_04580	1062580	G>T	44	100	35.63636364	p.Glu43Asp	Glycine reductase complex component C subunit alpha
Ccar_06000	1393366	G>T	51	100	31.23529412	p.Ala65Ser	Membrane protein
Ccar_06535	1517454	C>T	41	100	35.87804878	p.Ala268Val	Shikimate dehydrogenase
Ccar_08360	1943570	C>A	40	100	34.675	p.Ala94Ser	Histidine Kinase
Ccar_09455	2185545	C>A	51	100	36.56862745	p.Met1?	Hypothetical protein
Ccar_13870	3182963	A>G	75	78.66666667	31.3220339	-	Hypothetical protein
Ccar_16450	3728977	A>T	64	76.5625	34.95918367	-	Pseudogene
Ccar_17395	3916288	C>A	78	94.87179487	35.56756757	p.Ala201Ser	Hypothetical protein
Ccar_20250	4527710	C>A	63	100	36.49206349	p.Leu232Ile	23S rRNA (cytosine1962-C5)-methyltransferase
Ccar_23215	5150107	C>A	64	98.4375	34.85714286	p.Val374Leu	Membrane protein
Ccar_24055	5353493	G>T	81	100	32.45679012	p.Tyr379*	MATE family multidrug efflux transporter

Analysis revealed 14 SNPs with the predicted protein level changes consisting of 9 missense, 2 nonsense, 2 silent, and 1 non-stop mutation. Affected open reading frames and predicted gene products are summarised in **table 15.**

Nonsense mutations are characterised by a stop codon gain or start codon loss resulting in either the expression of a nonfunctional truncated protein or no expression at all. Consequently, nonsense mutations are the most likely to result in a change in phenotype. In *C. carboxidivorans* Δ7RM, three such mutations exist. The coding regions of Ccar_00510, a phosphate transport system permease protein, and Ccar_24055, a Multidrug and Toxic Compound Extrusion (MATE) family multidrug efflux pump, both feature a premature stop codon and the coding region of Ccar_09455, a hypothetical protein, has its start codon removed. Blast analysis of these revealed no gene redundancy i.e. no duplicate or similar genes existing within the genome of *C. carboxidivorans*. The viability of *C. carboxidivorans* Δ7RM confirms these proteins are not essential for cell survival under standard culture conditions. Further blast analysis did not reveal more information on the function of the hypothetical protein Ccar_09455, homologous genes (>97% identity) exist in the genomes of the closely related organisms *Clostridium drakei* and *Clostridium scatologenes* but are similarly annotated as hypothetical proteins. The genome of *Clostridium sp. JS66* also has a homologous gene annotated as a “tetratricopeptide repeat protein” however, tetratricopeptide repeat motifs are found in a wide variety of proteins with disparate functions so no inferences can be made on purpose of the encoded protein (Cerveny *et al.*, 2013). Since Ccar_00510 and Ccar_24055 encode for proteins that are found in the cell membrane it is

possible that the loss of these could affect conjugation efficiency in some manner. Therefore, it would be prudent to sequence the genomes of all intermediate strains in this study to see if the appearance of these mutations coincides with an increase in conjugation efficiency.

Apart from the silent mutations found in Ccar_13870 (hypothetical protein) and Ccar_16450 (pseudogene), nine other coding regions feature a deviation from the C94 strain all containing missense (non-synonymous) mutations. One such mutation that may alter the ability of the cell to uptake DNA is that in the coding region of the twitching mobility protein PilT (Ccar_00445), a hexameric ATPase that mediates disassembly and retraction of type IV pilli (Merz *et al.*, 2000). The *pilT* gene deficient mutants of *Neisseria gonorrhoeae* (Wolfgang *et al.*, 1998) and *Pseudomonas stutzeri* (Graupner *et al.*, 2001) have been shown to lose their competency for natural transformation. In contrast to this, *pilT* mutants of *Thermus thermophilus* were shown not to be defective in natural transformation but the inactivation of a related gene, *pilF*, resulted in the cell losing the ability to be transformed (Salzer *et al.*, 2014). Whilst the method of DNA delivery in this study is conjugation rather than transformation, there is clearly a relationship between type IV pili and DNA uptake. Therefore, it would be worth restoring the genotype of *pilT* in *C. carboxidivorans* Δ7RM to that of the wild type to ensure that any potential effects this mutation has on conjugation efficiency is accounted for. Moreover, this could reveal an alternative route for increasing competency in genetically incalcitrant strains to RMS deletion. Additionally, the genes Ccar_06000 and Ccar_23215, annotated simply as “membrane proteins”, both feature a missense mutation that could

impact conjugation efficiency if they happen to interact with the pilus of the donor strain. However, BLAST analysis of their respective nucleotide sequences reveals no clues as to their function, so this is merely speculative.

Importantly, no genes involved in acidogenesis, solventogenesis, the Wood-Ljungdahl pathway show any mutations, suggesting that, at least on a genetic level, the ability for carbon fixation and alcohol production has been retained. Further characterisation of *C. carboxidivorans* Δ7RM's fermentation capabilities will follow in Chapter 4.

3.3. Discussion

3.3.1 Conjugation efficiency of genetically domesticated *C. carboxidivorans* strains

As shown by **figure 7** in **3.2.6.**, there is a general trend of an increase in conjugation efficiency as RMS associated nuclease-encoding genes are deleted from the genome of *C. carboxidivorans*.

The first nuclease-encoding gene knocked out was that found in the RMS CcaP7VI, which has been attributed with the recognition sequence AGAAGC (Kottenhahn *et al.*, 2023). Given that the sequence AGAAGC occurs within the vector pMTL83151 three times **(table 16)**, it would be expected that the deletion of the gene encoding the nuclease that targets this sequence would be accompanied by an increase in conjugation efficiency. However, this was not the case with zero colonies formed on selective plates post conjugation. This could be explained by the cumulative effects of the remaining RMSs negating

the *C. carboxidivorans* Δ1RM strain's ability to uptake DNA. If this effect is strong enough, it could mask any increase in conjugation efficiency producing data that falsely suggests that the nuclease associated with CcaP7VI does not actively inhibit the uptake of foreign DNA. This hypothesis could be proved through a complementation study using the final *C. carboxidivorans* Δ7RM and a cas9 vector designed to reinsert the nuclease encoding gene of CcaP7VI at the same locus from which it was deleted, taking advantage of the bookmarking system implemented in this study (Seys *et al.*, 2020). Except for Type-IV systems, no other RMSs remain in *C. carboxidivorans* Δ7RM and so the true effect the RMS CcaP7VI has on conjugation efficiency can be inferred. Alternatively, when Kottenhahn *et al.* (2023) characterised the RMSs of *C. carboxidivorans* P7 the methylation motif AGAAGC could not be confidently assigned to any methyltransferase and it was only by process of elimination that it could be attributed to CcaP7VI. When considering that in the same study the methyltransferase of this RMS showed no methylation activity when recombinantly expressed in *E. coli*, it is possible that this recognition sequence has been assigned to CcaP7VI erroneously and the occurrence of these methylation motifs in the genome of *C. carboxidivorans* P7 is a result of another unidentified methyltransferase.

Table 16 - The RMSs of C. carboxidivorans P7, their recognition sequences, and the frequency at which those sequences occur in the vector pMTL83151.

RMS Name	Type	Recognition Sequence	Frequency in pMTL83151
CcaP7I	I gamma	CAYNNNNNCTGC	3

CcaP7II	I gamma	CCANNNNNNNNTCG	1
CcaP7III	I gamma	GCANNNNNNNNTNNCG	1
CcaP7IV	II-G Alpha	GATAAT	2
CcaP7V	II-G Alpha	CRAAAAR	7
CcaP7VI	III beta	AGAAGC	3
CcaP7VII	III beta	GAAAT	9

The second nuclease encoding gene knocked out was that found in the type-I RMS CcaP7I with the recognition sequence CAYNNNNNCTGC creating the strain *C. carboxidivorans* Δ2RM. The recognition sequence assigned to CcaP7I by both Redfern (2021) and Kottenhahn *et al.* (2023) appears within the vector pMTL83151 three times **(table 16)**, so the observed increase in conjugation efficiency is to be expected **(figure 16)**. This strain marks the first instance where a standard vector can be delivered to *C. carboxidivorans* via conjugation without the aid of a methyltransferase harbouring donor strain, strongly suggesting that CcaP7I is an active RMS. The degree at which CcaP7I suppresses DNA uptake, however, is hard to establish with the results of the experiment in **3.2.6.** because of the potential for interference by the remaining RMSs. This could be more accurately determined by conducting a complementation study in a similar way as discussed with CcaP7VI.

The third nuclease encoding gene knocked out was that associated with the type-II fused RMS CcaP7IV creating the strain *C. carboxidivorans* Δ3RM. The knockout of this RMS was accompanied by an increase in conjugation efficiency

similarly to that found with the removal of the nuclease component of CcaP7I. The recognition sequence GATAAT was assigned to the RMS CcaP7IV by both Redfern (2021) and Kottenhahn *et al.* (2023) and occurs in the vector pMTL83151 twice, which would explain this statistically significant increase in conjugation efficiency. This increase is greater than that exhibited by its parent strain despite the frequency of its assigned recognition sequence in pMTL83151 being one less than that of CcaP7I. However, when considering that the remaining RMSs are not the same in both strains it is not a fair comparison to make, and so it cannot be concluded that CcaP7IV presents a greater barrier to DNA uptake than CcaP7I. More conclusive comparisons could be made with a complementation study as describe previously whilst delivering a vector that contains each recognition sequence with an equal frequency.

The fourth nuclease encoding gene deleted was the type-I nuclease CcaP7III to create the strain *C. carboxidivorans* Δ4RM. Again, the deletion of this nuclease encoding gene resulted in a statistically significant increase in conjugation efficiency compared to its parent strain *C. carboxidivorans* Δ3RM **(figure 7)**. This is an expected result as the recognition sequence assigned to this RMS in studies by both Redfern (2021) and Kottenhahn *et al.* (2023), GCANNNNNNNTNNCG, appears in the shuttle vector, only once. Despite the low occurrence of this recognition sequence, this deletion resulted in an almost a 6-fold increase in conjugation efficiency compared to *C. carboxidivorans* Δ3RM.

Following CcaP7III, the fifth nuclease-encoding gene, belonging to the type-III RMS, CcaP7VII, was deleted creating the strain *C. carboxidivorans* Δ5RM. A recognition sequence could not be assigned by Redfern (2021) to CcaP7VII however by expressing the methyltransferase component of this RMS in *E. coli* Kottenhahn *et al.* (2023) determined this sequence to be GAAAT. This sequence appears nine times in the vector pMTL83151, the most prevalent of the RMS recognition sites. Despite this, deleting the nuclease of CcaP7VII resulted in no significant difference in conjugation efficiency. When determining the methylome of *C. carboxidivorans* P7 via PacBio SMRTseq Kottenhahn *et al.* (2023) could not find the methylation pattern GAAAT, suggesting that the RMS Ccap7VII is not actively expressed. Since *C. carboxidivorans P7* contains numerous RMSs, it is possible that the fitness cost of maintaining these systems without bacteriophage-driven selection pressure may be causing the P7 strain to shed redundant RMSs, such as CcaP7VII. Another explanation is that the RMS CcaP7VII has an additional cellular function (Vasu & Nagaraja, 2013), it has been shown that the type II methyltransferases *Eco*RII and *Sso*II have been shown to act as transcription regulators that recognise operator sequences different from their respective methylation recognition sequences in addition to their DNA methylation activity (Karyagina *et al.*, 1997; Som & Friedman, 1994). Therefore, CcaP7VII may well have gained an additional cellular function that is necessary for cell survival, which would maintain its presence within the genome of *C. carboxidivorans* over time despite the addition of other RMSs that may have rendered it superfluous for cellular defence. It would be interesting to determine if the nuclease component of CcaP7VII actively cleaves DNA by

overexpressing in *E. coli*, extracting, and purifying the protein for the treatment of a sample of DNA that is known to contain the sequence GAAAT. Lack of nuclease activity would suggest that CcaP7VII could have an alternative function to RMS activity. Additionally, the expression of the gene could be monitored at various stages of growth through reverse transcription PCR.

The penultimate RMS associated nuclease encoding gene knocked out was that belonging to the fused Type-II RMS CcaP7V to create the strain *C. carboxidivorans* Δ6RM. Like the previous RMS, a recognition sequence could not be assigned to this RMS by Redfern (2021) but Kottenhahn *et al.* (2023) assigned it the recognition sequence CRAAAAR. This sequence appears in the vector pMTL83151 seven times and as such it is expected that an increase in conjugation efficiency would be observed after knocking out the relevant nuclease encoding gene. However, like in the previous strain, there was no statistically significant difference in the number of transconjugant colonies formed between *C. carboxidorans* Δ6RM and its parent strain. Kottenhahn *et al.* (2023) noted that the specific methylation motif CAAAAAR can be found in several different species of *Clostridium* including *C. botulinum*, *C. sporogenes*, *C. autoethanogenum*, *C. pasteurianum*, *C. tetani,* and *C. ljungdahlii* as well as in *Eubacterium limosum* (Roberts, Vincze*, et al.*, 2003). This same motif is also ubiquitous in 36 different strains of *C. difficile* and deactivating the methyltransferase responsible for this methylation pattern (*camA*) resulted in sporulation defects, suggesting a role in gene regulation (Oliveira *et al.*, 2020). Due to the similarities CRAAAAR has to this sequence, it is possible that Ccap7V

also serves a regulatory purpose that has since superseded its function as a RMS.

The final nuclease encoding gene knocked out belonged to the type-I RMS CcaP7II, creating the strain *C. carboxidivorans* Δ7RM. This RMS recognises the sequence CCANNNNNNNNTCG, which appears in the vector pMTL83151 once. The deletion of this nuclease-encoding gene marks the highest level of amenability to DNA uptake observed by a substantial margin. *C. carboxidivorans* Δ7RM displayed a 230-fold increase in conjugation efficiency when compared to its progenitor (*C. carboxidivorans* Δ6RM) with a mean 40,5000 CFU/mL. This is a marked improvement to the conjugation efficiency shown when delivering the RMS silent vector pMTL83151y-v1 using the methylation donor strains sExpress_TII and sExpressT_III, which produced a mean CFU of 1,237.65 and 30,038.8 (Redfern, 2021), respectively.

Plasmid curing has been associated with an increase in transformation efficiency in *C. pasteurianum* (Grosse-Honebrink *et al.*, 2017) and the process for generating subsequent knockouts requires curing the Cas9 harbouring plasmid from the cell. Therefore, it would have been prudent to include a control that was a plasmid-cured variant of each strain analysed in **3.2.6.** to ensure that repeated plasmid curing was not responsible for the increase in conjugation efficiency observed. Whilst this is unlikely to be the case given that increase in conjugation efficiency is not uniform across each strain and that repeated plasmid curing hasn't been shown to increase conjugation efficiency, however it is an omission worth mentioning when interpreting the results.

3.4. Conclusions

In conclusion, full genetic domestication of *C. carboxidivorans* has been achieved. Additionally, targeted removal of RMS associated nuclease encoding genes using with an RM silent vector in tandem with methylation donor strains has been shown to be an effective method to achieve this.

There has also been found strong evidence that 2 of *C. carboxidivorans* RMSs are inactive, consistent with the characterisation performed by Kottenhahn *et al.* (2023). This means that in future genetic domestication endeavours, particularly in organisms with a high number of RMSs, researchers should perform RMS characterisation beforehand to better prioritise knockout targets.

This work stands as a case study for the successful genetic domestication of an incalcitrant microorganism and provides a workflow to achieve this in other high-potential non-model organisms.

Chapter 4 – Metabolic Engineering and Characterisation of *C. carboxidivorans* Δ7RM

4.1. Introduction

As discussed in **3.3.2.**, the process of genetic domestication of *C. carboxidivorans* required extensive modifications, which may have inadvertently resulted in a strain that had lost the ability to grow on gas or produce organic alcohols. Whilst sequence and analysis of the genome of the final strain showed no mutations in the Wood-Ljungdahl pathway associated genes or in genes governing alcohol production or chain elongation, there could be alterations on a regulatory level that causes the new strain to differ in the balance of fermentation products from that of the wild type. Therefore, it was important to conduct gas-fermentation experiments with *C. carboxidivorans* Δ7RM to ensure that this was not the case.

With a fully genetically amenable strain now available, there is also an opportunity to perform metabolic engineering in parallel with characterisation to demonstrate the potential of *C. carboxidivorans* Δ7RM as a platform for sustainable chemical production and as a research strain.

For simplicity, a single gene would be selected in *C. carboxidorans* Δ7RM that is hypothesised improve to alcohol production when grown on CO_2 and H_2. With no previous genetic modification studies in *C. carboxidivorans*, the existing literature for the closely related model ethanol-producing gas-fermenting

bacteria *C. autoethanogenum* was searched for potential knockout targets. One such study demonstrated that separate in-frame deletions of two bi-functional aldehyde/alcohol dehydrogenase encoding genes (*adhE1* and *adhE2*) via allelic exchange resulted in a three-fold increase in ethanol for both mutant strains (Liew *et al.*, 2017). A BLASTP analysis of *AdhE1* (CAETHG_3747) and *AdhE2* (CAETHG_3748) revealed that the *C. carboxidivorans* genome contains two highly homologous (>80%) AdhE encoding genes that are contiguous in the genome in a similar manner to Adhe1 and Adhe2. The resemblance of these genes between the two acetogens suggests they conduct the same metabolic function, and that their inactivation in *C. carboxidivorans* is likely to produce similar solventogenesis-enhancing effects to those seen in *C. autoethanogenum*.

Knockout targets could also be found in a study published by Antonicelli *et al.* (2023) where the growth and product formation of *C. carboxidivorans* was improved when growing on CO_2 and H_2 via Adaptive Laboratory Evolution (ALE). By repeated propagation of *C. carboxidivorans* in batch fermentations growing on CO_2 and H_2 an adapted strain was produced that displayed a significant increase in growth and alcohol production. Genomic analysis of the final strain revealed a frameshift mutation in the A subunit (Ccar_16055) of the electron bifurcating hexameric [FeFe]-hydrogenase HytABCE_1E_2 (Di Leonardo *et al.*, 2022) creating a premature stop codon. The resulting truncated protein is 380 amino acids shorter than the original protein, and only 80 amino acids long and is therefore very likely non-functional. This hydrogenase forms a complex with formate dehydrogenase (FdhA) (Wang *et al.*, 2013) and provides the cell with

the electrons for the reduction of NADP and ferredoxin whilst, in tandem with FdhA, also providing the electrons for the reduction of CO_2 to formate in the first step of the methyl branch of the Wood-Ljungdahl Pathway **(figure 17)**. An example of the deletion of any subunit-encoding genes of this complex resulting in an increase in alcohol production under autotrophic growth in acetogens does not exist in the literature. However, a patent has been filed from Nagaraju and Koepke (2019) who explored group-II intron mutagenesis of the hydrogenase-associated subunits of $HytABCE_1E_2$ complex in *C. autoethanogenum* and demonstrated that mutations of *hytA* (CAETHG_2798) and *hytE1* (CAETHG_2797) resulted in significant increase in cell growth and ethanol production whilst reducing acetate formation when growing on CO_2 and H_2. Interestingly, the equivalent disruption of all other hydrogenase-encoding genes were deleterious to growth and ethanol production when grown on CO_2 and H_2. These results suggest that disruption of the hydrogenase components of the $HytABCE_1E_2$ complex does not affect its CO_2 reduction capabilities. It is postulated that other hydrogenases in *C. autoethanogenum* and *C. carboxidivorans* are either more efficient or are upregulated to compensate.

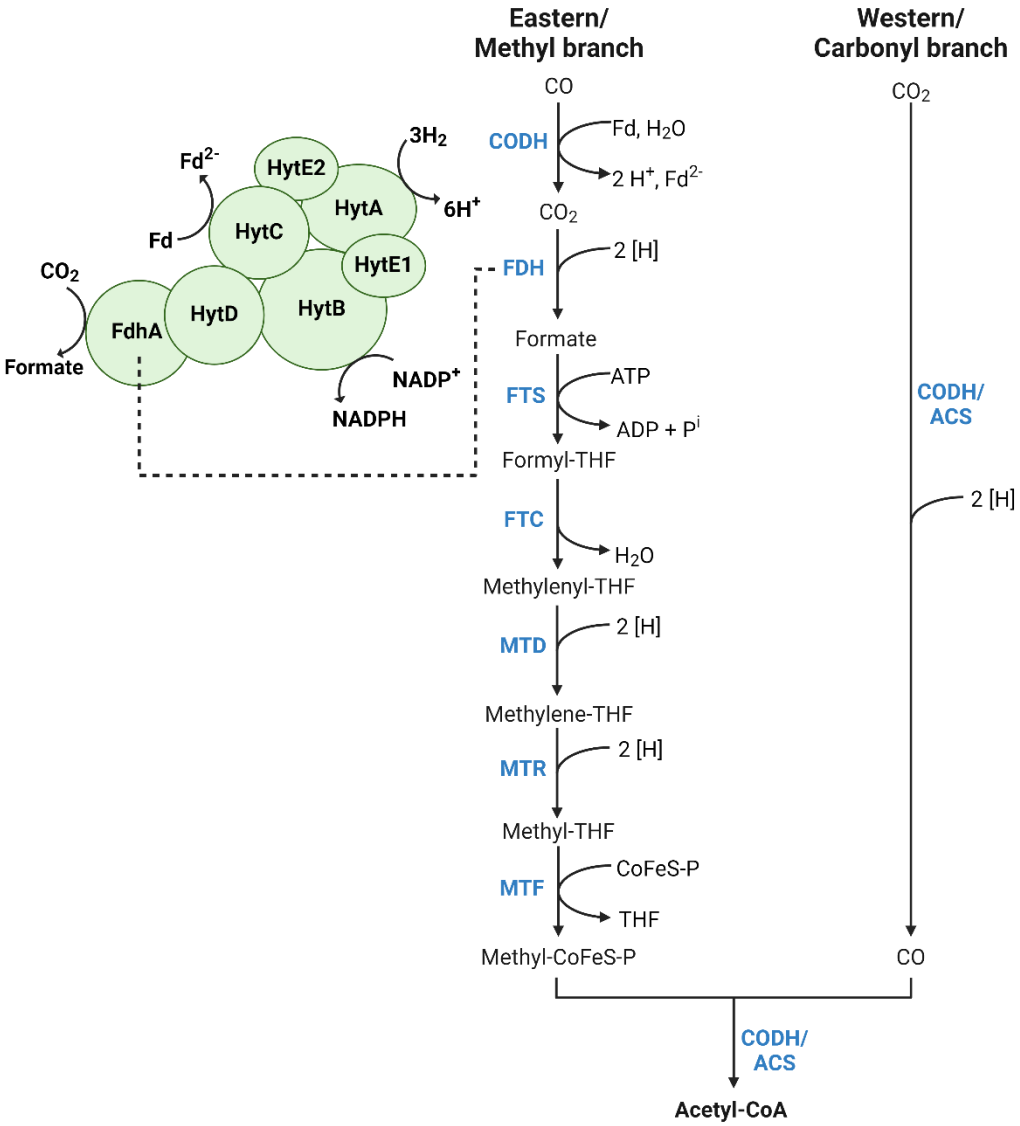

Figure 17 - The proposed protein interactions of the electron bifurcating hexameric [FeFe]-hydrogenase HytABCE1E2 (Poudel et al., 2018) and its function within the Wood-Ljungdahl Pathway.

CODH = Carbon Monoxide Dehydrogenase, FDH = Formate Dehydrogenase, THF = Tetrahydrofolate, FTS = Formyl-THF Synthase, FTC = Formyl-THF Cyclohydrase, MTD = Methylene-THF Dehydrogenase, MTR = Methylene-THF Reductase, MTF = Methyltransferase, ACS = Acetyl-CoA Synthase. Created with BioRender.com.

It was decided that *hytA* would be the most suitable target for deletion because there is strong evidence that it would result in improved growth and

solventogenesis in *C. carboxidivorans* and presents an opportunity to validate the findings from Antonicelli *et al.* (2023). Moreover, *adhE1* and *adhE2* deletions have already been demonstrated to have a favourable effect on alcohol production in *C. autoethanogenum* and a *hytA* mutant has not been published outside of a patent making it a more novel contribution to the literature.

4.1.2. Aims

The aim of this chapter is to characterise the growth and fermentation products of the domesticated strain *C. carboxidivorans* Δ7RM through batch gas fermentations on CO_2 and H_2. It is hypothesised that there will be no significant difference between the domesticated strain and the wild type because genomic analyses showed no mutations in key gas fermentation metabolic pathways and the deleted RMSs have no role in *C. carboxidivorans'* core metabolism.

The second aim of this chapter is to demonstrate *C. carboxidivorans* Δ7RM's potential as a platform for chemical production from inorganic carbon and as a platform strain for research into chain elongating acetogens. This will be achieved by exemplifying metabolic engineering in *C. carboxidivorans* Δ7RM through the deletion of *hytA,* which is hypothesised to improve biomass formation and alcohol production in gas fermentations on CO_2 and H_2. If fermentations are successful and the *hytA* mutant strain exhibits superior growth and ethanol, butanol, and hexanol production when grown autotrophically, proteomic analysis of the strain will be performed to determine if an alternative hydrogenase is upregulated to compensate for the loss of *hytA*.

4.2. Results and Discussion

4.2.1. *hytA* (Ccar_16055) knockout

CRISPR-Cas vectors designed for the deletion of *hytA* were constructed as described in **2.4.8.** and the knockout was performed using these vectors as described in **2.5.** As with the deletion of the nuclease-encoding genes (**3.2.**), a bookmark sequence (bookmark 8 in Seys *et al.* (2020)) was delivered as cargo on the knockout vector to the locus of *hytA*. Three different guide RNAs were used in parallel when attempting to knock out *hytA*, with the second guide RNA (ATAGGATGTGTTAGTGGCGG) successfully producing an in-frame deletion (**figure 18**) resulting in strain *C. carboxidivorans* Δ7RM Δ*hytA*.

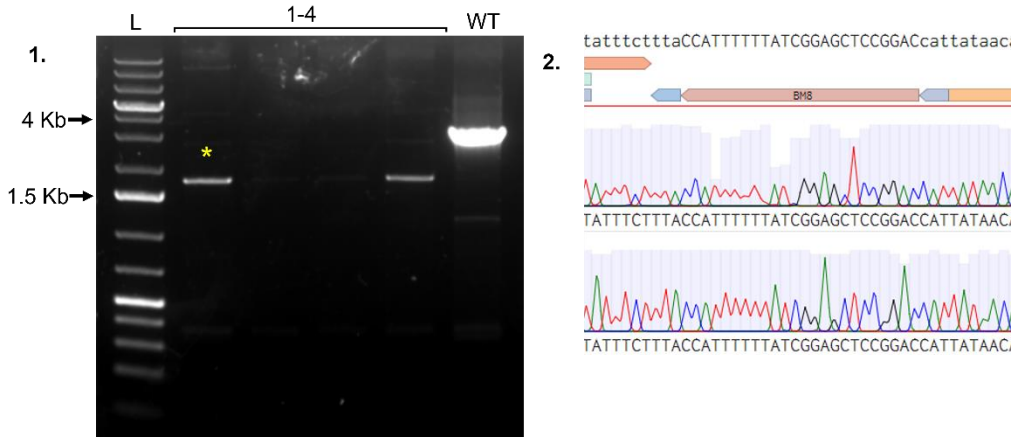

Figure 18 - 1: Agarose gel electrophoresis cPCR of colonies formed on theophylline induction plates using primers Ccar_16055_F and Ccar_16055_R to screen for hytA deficient mutants.

1-4 guide RNA 2. L - GeneRuler 1kb plus DNA ladder (Thermo Scientific, USA) WT – wild type. 2: Sanger sequencing of knockout locus confirming insertion of bookmark 8. Mutant selected for Sanger sequencing and further characterisation indicated by (*).

The four colonies screened via cPCR on induction plates all produced amplicons of expected size indicating a successful knockout, however 2 (3 and 4 **figure 18**) produced very faint bands. This could be due to insufficient amount of template when picking colonies or too much biomass interfering with reaction components. This indicates that the guide RNA tested has a 100% knockout efficiency.

4.2.2. Illumina Sequencing of *C. carboxidivorans* Δ7RM Δ*hytA* genome.

The genomic DNA of the newly created *C. carboxidivorans* Δ7RM Δ*hytA* strain was extracted (**2.4.4.**) and a sample was prepared for Illumina sequencing by DeepSeq (Nottingham, UK). Sequencing data were mapped against the genome sequence of *C. carboxidivorans* Δ7RM to identify any unwanted mutations that could have occurred during the deletion of *hytA*.

The only mutation identified was a single SNP (C>T) at position 56 of the non-coding pseudogene Ccar_16450.

4.2.3. Characterisation of *C. carboxidivorans* Strains on CO_2 and H_2

With the *hytA* mutant created, the next step was to characterise its growth and fermentation products along with *C. carboxidivorans* Δ7RM and the wild-type strain (acquired from the DSMZ) when growing on CO_2 and H_2.

Antonicelli *et al.* (2023) previously established a fermentation protocol that produced the full suite of metabolic products of interest (i.e. hexanol, butanol, ethanol, hexanoate, butyrate, and acetate) in high enough concentrations where any changes between strains should be detectable via HPLC. It was found that for batch fermentations in serum bottles, a 1:4 CO_2:H_2 gas ratio in the headspace, an incubation temperature of 25 °C, and a media:gas volumetric ratio of 1:9 (25 mL : 225 mL) were the optimal conditions for alcohol production and growth. Therefore, these conditions and the same modified 1754 PETC modified media (**table 17**) used by Antonicelli *et al.* (2023) was chosen for the characterisation of the strains created in this study.

Each strain was initially revived from a frozen stock in 3 mL YTAF MES media that was then used to inoculate 5 mL of 1754 PETC modified media with 10% fructose to minimise the transfer of media components from the YTAF MES into the serum bottles as these may be used as an alternative carbon source. These cultures were used to inoculate serum bottles containing 25 mL of 1754 PETC modified media, which were previously filled with a mixture of H_2 and CO_2 in a 1:4 ratio under 2 bars of pressure. These serum bottles were incubated at 25 °C and agitated at 200 RPM.

Table 17 - 1754 PETC modified media used by Antonicelli et al. (2023) for autotrophic growth of C. carboxidivorans.

Component	Per Litre
1754 PETC modified media	
NH_4Cl	1.00 g

NaCl	0.80 g
KCl	0.10 g
KH_2PO_4	0.10 g
$MgCl_2 \cdot 6H_2O$	0.20 g
$CaCl_2 \cdot 2H_2O$	0.04 g
Yeast Extract	0.50 g
MES	5 g
Vitamin solution	10 mL
Trace metal solution	10 mL
Reducing solution	10 mL
dH_2O	970 mL

---

141 DSMZ Vitamins solution 100x

Pyridoxine	10.0 mg
Thiamine	5.00 mg
Riboflavin	5.00 mg
Calcium pantothenate	5.00 mg
Thioctic acid	5.00 mg
p-(4)-Aminobenzoic Acid	5.00 mg
Nicotinic acid	5.00 mg
Vitamin B12	0.10 mg
Biotin	2.00 mg
Folic Acid	2.00 mg

1754 PETC Trace Elements Solution 100x

Nitrilotriacetic acid	1.50 g
$MgSO_4 \cdot 7H_2O$	3.00 g
$MnSO_4 \cdot H_2O$	0.50 g
NaCl	1.00 g
$FeSO_4 \cdot 7H_2O$	0.10 g
$CoCl_2 \cdot 6H_2O$	0.10 g
$CaCl_2$	0.10 g
$ZnSO_4 \cdot 7H_2O$	0.10 g
$CuSO_4 \cdot 5H_2O$	0.01 g
$AlK(SO) \cdot 12H_2O$	0.01 g
H_3BO_3	0.01 g
$NiCl_2 \cdot 6H_2O$	0.01 g
$Na_2WO_4 \cdot 2H_2O$	0.01 g
Na_2SeO_4	0.01 g
$Na_2MoO_4 \cdot 2H_2O$	0.01 g

Reducing Solution 100x

L-cysteine	20 g
$Na_2S·9H_2O$	20 g

These cultures were meant to be used to inoculate a second and final batch of serum bottles after each strain had adapted to autotrophic growth, indicated by an increase in optical density and decrease of internal pressure. However, after two weeks no drop in pressure was observed and only an OD_{600} of <0.2 was achieved.

After a second attempt with no success, the literature was consulted revealing a recently published study comparing the fermentation products of the newly discovered species *C. muellerianum* P21 with other medium-chain alcohol-producing acetogenic Clostridium, including *C. carboxidivorans*, after growing on CO_2 and H_2 (Thunuguntla *et al.*, 2024). This study used a media called P11 (**2.7.1.**), which contains higher concentrations of minerals as well as the addition of the vitamin 2-Mercaptoethanesulfonic acid. Adaptation was repeated in the same manner as before with the new media in the hope that the additional media components would aid in adaptation from heterotrophic to autotrophic growth. However, no growth was observed with the strains *C. carboxidivorans* Δ7RM and *C. carboxidivorans* Δ7RM Δ*hytA*. One culture of the wild-type strain did successfully adapt to growth on CO_2, but sub-culturing failed.

After contacting the corresponding author of the paper (Thunuguntla *et al.*, 2024), it was suggested to alter the adaptation process by preparing the

inoculum on a syngas mixture (CO, CO_2, and H_2) rather than CO_2 and H_2. The inoculation preparation process was repeated except the CO_2 and H_2 was replaced with a syngas mixture consisting of 40:30:30 of CO:CO_2:H_2 at 1.5 bar during the adaptation step. The wild-type strain successfully adapted to growth on the syngas mixture and, once in exponential phase, was used to inoculate new serum bottles containing CO2 and H_2 but these also failed to grow. The domesticated and *hytA* mutant strains again grew to an OD_{600} of ~0.2 with a small decrease in internal pressure but when these were used to inoculate serum bottles for growth on CO_2 and H_2, no further growth was observed.

Bacterial flocculation (large aggregation of particles) could be observed during attempts to grow each strain of *C. carboxidivorans* on CO_2 and H_2. This phenomenon was also noted by S. Shen *et al.* (2017) and Shen *et al.* (2020) where it was found that syngas batch fermentations in serum bottles incubated at 37 °C resulted in auto aggregation within 48 hours after stationary phase. It was shown that flocculation could be avoided by reducing the incubation temperature to 25 °C or by supplementing the growth media with the surfactants saponin and Tween 80 (Shen *et al.*, 2020). This differs from the observations in this study where bacterial flocculation was observed at low levels of growth within 48 hours after inoculation.

It is possible that the bacterial auto aggregation observed is a stress response due to the lack of carbon available for growth. Initial growth of cultures could be attributed to the yeast extract being used as a carbon source that, when depleted, requires the cell to switch to autotrophic growth. A failure to make

this transition could activate a stress response in the cell resulting in agglomeration. Additionally, when viewing the samples of the flocculated cultures under a microscope, the cells show an elongated morphology (**figure 19**). This is a sign of bacterial filamentation that occurs when cells continue to elongate but fail to divide and can be induced by stress related factors such as nutrient deficiency or oxidative stress (Rizzo *et al.*, 2020).

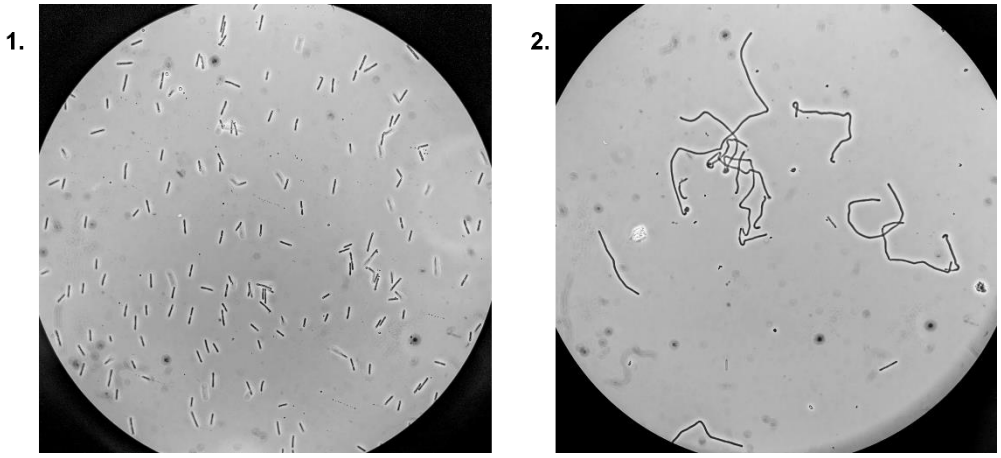

Figure 19 - 1: A sample of C. carboxidivorans grown heterotrophically under 100x magnification. 2: A sample of flocculated C. carboxidivorans P7 culture from attempted adaptation to growth on CO_2 and H_2 under 100x magnification.

An alternative explanation is of the product toxicity inducing a flocculation response in the cells. It was observed by Kottenhahn *et al.* (2021) that supplementing autotrophic *C. carboxidivorans* cultures with 40 mM of hexanol would cause cells to agglomerate into flocs. Additionally, these cultures would produce significantly less biomass compared to the control group indicating hexanol is toxic to *C. carboxidivorans* at certain concentrations and that flocculation is a response to product toxicity. It is possible that a similar phenomenon may be occurring in the autotrophic adaptations attempted in

this study. Since accurate optical density readings require a homogenous culture, flocculation would give unreliable measurements that could suggest a lower growth than what is occurring. This could mean there is enough biomass to produce toxic levels of product but since there was no observed drop in pressure, indicating no fixation of CO_2 for biomass production, this is unlikely to be the case. It is most likely that the issue lies in the cultivation protocols followed, and a refinement of which would lead to reliable autotrophic adaption.

4.2.4. Genomic analysis of C94 and starting strain

To determine if any mutations were acquired prior to development of *C. carboxidivorans* Δ7RM that could explain the lack of autotrophic growth, Illumina sequencing data of C94 by Redfern (2021) and of the starting starting laboratory strains were mapped against the published wild-type genome sequence (**table 18**).

Table 18 - Variants in open reading frames of C94 (white) and the starting laboratory strain (grey) Illumina sequencing data when mapped against the wild-type genome sequence of C. carboxidivorans P7. *Silent mutations have been omitted.*

Locus Tag	Position	Base change (%)	Coverage (%)	Frequency	Average Quality	Predicted Amino Acid Change	Gene Product
Ccar_00825	160189	G>T	71	97.2	35.8	p.Gly200Val	GTP-sensing pleiotropic transcriptional regulator CodY
Ccar_00825	160221	C>G	62	100	36.5	p.Leu211Val	GTP-sensing pleiotropic transcriptional regulator CodY
Ccar_01050	205543	C>A	62	98.4	33.9	p.Ala378Asp	MFS transporter
Ccar_01705	367599	G>T	73	100	34.2	p.Ala173Ser	Hypothetical protein
Ccar_03545	823150	A>Del.	55	98.2	34.5	-	Pseudogene
Ccar_04190	970007	G>T	65	100	36.9	p.Glu229*	Hypothetical Protein
Ccar_07510	1736431	T>A	25	100	35.4	p.Gln178His	N-acetylmannosamine kinase
Ccar_10985	2513827	C>A	57	100	34.1	p.Gly396Cys	Cell wall-binding repeat-containing protein
Ccar_17250	3878707	C>A	87	98.9	36.6	p.Ser432Ile	Murein biosynthesis integral membrane protein MurJ
Ccar_17445	3925819	A>C	78	97.4	33.0	p.Asp598Glu	threonine--tRNA ligase
Ccar_17445	3926219	C>T	75	98.7	34.7	p.Gly465Asp	threonine--tRNA ligase
Ccar_20565	4582631	C>A	73	100	37.4	p.Asp305Tyr	phosphoenolpyruvate carboxykinase
Ccar_25335	5629616	G>T	77	100	34.5	p.Glu15*	ATPase
Ccar_25825	5726044	G>T	61	100	33.8	p.Ala188Ser	acetylglutamate kinase

The creation of the C94 strain consisted entirely of growth on fructose with no characterisation of autotrophic growth, so it is possible that deleterious mutations had accumulated in the Wood-Ljungdahl Pathway that have gone undetected. However, no such mutations could be found in the Illumina sequencing data. Moreover, there is no other obvious mutations that could lead to the inability of the domesticated strain to grow on gas.

Two missense mutations can be found in *codY* (Ccar_00825), a global regulatory protein that monitors nutrient levels by detecting Guanosine-5'-triphosphate (GTP), which has been shown to play an important role in virulence gene expression in *C. difficile* (Dineen *et al.*, 2007)*, Staphylococcus aureus* (Majerczyk *et al.*, 2008)*, Streptococcus pneumonia* (Hendriksen *et al.*, 2008)*, Streptococcus pyogenes* (Malke & Ferretti, 2007)*, Streptococcus mutans* (Lemos *et al.*, 2008)*, Listeria monocytogenes* (Bennett *et al.*, 2007)*, Bacillus cereus* (Hsueh *et al.*, 2008), and *Bacillus anthracis* (van Schaik *et al.*, 2009). It has been implicated in the direct or indirect expression of over 160 different genes in *C. difficile* (Dineen *et al.*, 2010). Additionally, CodY has been shown to regulate biofilm formation in the species listed above, so the mutations found in Ccar_00825 may have some role in the flocculation observed when attempting to adapt the strains in this study to autotrophic growth.

Since *C. carboxidivorans* is not a pathogen and so would not have virulence related genes to regulate, it could be that CodY plays a role in the expression of genes that enable the switch from heterotrophic to autotrophic growth. However, in a study by Woods *et al.* (2022) the essential gene set for

autotrophic growth of *C. autoethanogenum* was determined by generating a library of random mutants created by a *mariner*-based transposon system (Zhang *et al.*, 2015) in a technique known as Transposon Insertion Sequencing (TIS). Illumina sequencing of the sites of transposon insertion using a sequencing primer specific to the junction between the transposon and chromosome in the mutant library can elucidate the essentiality of genes under specific growth conditions. If enough insertions can be found within a coding region it can be deduced that this gene is not essential for survival under the growth conditions from which the sample was harvested. The data set produced in this study showed that *codY* is essential for growth under heterotrophic conditions, and so could not be validated autotrophically. Since the strains produced in the study grow on fructose as a carbon source, this suggests that the mutations in *codY* identified in **table 18** do not have a deleterious effect on the function of the translated protein.

This dataset of essential genes provides an opportunity to infer if the mutations found in coding regions annotated as hypothetical proteins are inhibiting adaptation to autotrophic growth. In the C94 strain there are two SNPs found in coding regions annotated as hypothetical proteins, a missense mutation in Ccar_01705 and a premature stop codon in Ccar_04190 producing a truncated protein of 228 amino acids in length compared to the native 433 amino acids. Additionally, *C. carboxidivorans* Δ7RM features a lost stop codon in the hypothetical protein Ccar_09455 and a missense mutation in Ccar_17395. However, BLAST analysis of the nucleotide sequences of these coding regions

against the genome of *C. autoethanogenum* produces no matches, so the essentiality of these genes to autotrophic growth cannot be inferred.

When performing the same analysis for all other missense and nonsense mutation-containing coding regions for all strains in this study, no genes are marked as essential for autotrophic growth.

4.3. Conclusion

In conclusion, the *hytA* gene encoding the HytA component of the $HytABCE_1E_2$ complex was successfully knocked out in *C. carboxidivorans* Δ7RM. However, no cultures of any strain of *C. carboxidivorans*, including the wild type, were successfully grown autotrophically on a gas mixture of CO_2 and H_2.

Genomic analysis revealed no obvious reason on a genetic level why the strains created were incapable of autotrophic growth. When considering this and the fact that the wild-type was similarly resistant to adapting to growth on CO_2 and H_2, the difficulties faced in culturing these strains is likely to be solved through a refinement of the cultivation protocols employed in this study. This could take the form of a gentler adaptation phase where a series of cultures with increasing headspace pressure of syngas before transferring to CO_2 and H_2.

Chapter 5 – Alternative Retrotransposition-Activated Markers

5.1. Introduction

As described in **1.5.2.**, the ClosTron is an insertional mutagen based on a mobile group II intron that is redirected to the chosen gene target by appropriate alteration of its RNA encoding sequence. Insertion leads to the inactivation of the target gene, most commonly through the interruption of the coding sequence and the consequent production of a non-functional protein. An important part of this system is the utilisation of a Retrotransposition-Activated Marker (RAM), which is crucial for the rapid selection of clones carrying the inserted intron within a population. The RAM consists of a gene encoding an antibiotic resistance gene in which the integrity of the coding region has been interrupted through the presence of a small segment of DNA encoding a group I intron. As group I introns are self-catalytic, it is spliced during the retrotransposition and insertion of the group II intron into the chromosome. Consequently, the antibiotic resistance gene is re-activated allowing for the direct selection of insertional mutants on agar medium supplemented with the requisite antibiotic.

The group I intron used is derived from the *td* gene of the bacteriophage T4. The exon sequence required for its splicing correspond to G/ACCCAAGAGA, where the intron resides between the first and second nucleotide. The ClosTron RAM comprises an *ermB* erythromycin resistance gene that is under the control

of a constitutive P_{thl} promoter, derived from the thiolase gene of *C. acetobutylicum*. Since *ermB* does not natively contain the correct exon sequence, a linker sequence encoding an additional 12 codons incorporating the sequence GACCCAAGAGA was added to the 5'-end of the gene. The additional 12 amino acids at the N-terminus of ErmB had no discernible effect on function (Heap *et al.*, 2007). This allowed the subsequent inclusion of the *td* group I intron and demonstration that the resultant *ermB*::RAM could be used to directly select Group II intron insertions on agar medium supplemented with erythromycin (Em).

Once the ClosTron has generated a mutant in a target organism, it cannot be used again the make a second mutant as the cell as it is now Em-resistant (^{R}). The creation of multiple mutants is instead possible either by using a ClosTron plasmid lack a RAM or through a process of marker recycling in which FRT sites that flank the *ermB*::RAM allow the excision of the insertionally activated *ermB* gene from the chromosome by the action of FLP recombinase. This enzyme is provided through the transformation or conjugation of a plasmid encoding a yeast-derived FLP recombinase gene (Heap *et al.*, 2010). The excision of the *ermB* gene from the chromosome restores erythromycin sensitivity to the strain and so a new *ermB*::RAM bearing ClosTron plasmid targeting a different locus can be used. Although the latter method was used to generate three sequential knockouts in *C. acetobutylicum* (Steiner *et al.*, 2011), attempts to excise *ermB* using FLP recombinase in other members of the *Clostridium* were unsuccessful (Kuehne & Minton, 2012).

One obvious way to make multiple mutations using TargeTrons would be to develop RAMs based on different antibiotic resistance genes. Indeed, in *E. coli* at least, such alternative RAMs do exist that confer resistance to kanamycin (Saldanha *et al.*, 2013) and trimethroprim (Zhong *et al.*, 2003). Unfortunately, these markers do not function in clostridia. Indeed, the range of antibiotic resistance markers available for use in these anaerobes are extremely restrictive, essentially being limited to genes conferring resistance to Em, tetracycline (Tc), and thiamphenicol (Tm). The latter antibiotic is used with those genes (*cat*) encoding chloramphenicol acetyltransferase. Over and above enabling the generation of multiple mutants, the availability of an alternative RAM to that based on *erm*, would expand the potential application of the ClosTron technology to those organisms who already exhibit resistance to erythromycin, such as *C. difficile* 630 (Farrow *et al.*, 2001) , where the *ermB*::RAM cannot be used.

5.2. Results

5.2.1. Screening for progenitor *catA*::RAM sequences

The two most widely applicable antibiotic resistance markers in clostridial species are the *C. perfringens catP* gene and the *ermB* gene of *Enterococcus faecalis* encoding resistance to chloramphenicol/thiamphenicol (Cm/Tm) and Em, respectively. To date, only a RAM for the latter has been described. At the time that *ermB*::RAM was devised, it was established that the addition of the

same 12 amino acid encoding extension to the *C. perfringens catP* did not result in an active gene conferring resistance to antibiotic on the bacterial host (N.P. Minton, personal communication). The simplest alternative would be to incorporate the requisite group I exon sequence GACCCAAGAGA into the coding region of the gene such that the functionality of the encoding protein is unaffected. This is most simply achieved by using codon redundancy to ensure that no change is made to the native amino acid sequence of the encoded antibiotic resistance protein. In practice, the sequence of the protein needs to encompass either one of the many variant 5-aa sequences shown in frame 1 (**table 19**), or one of the two 4-aa sequences illustrated in frames 2 and 3.

Table 19 - The possible translated amino acid sequences of each reading frame of the td group I exon sequence GACCCAAGAGA.

Frame	AA/Nucleotide Sequence of three possible frames				
1	L/M/V/S/P/T/A/Q/K/E/W/R/G	T	Q	E	I/M/T/N/K/S/R
	G	ACC	CAA	GAG	A
2	R/G	P	K	R	
	GA	CCC	AAG	AGA	
3	D	P	R	D/E	
	GAC	CCA	AGA	GA	

Accordingly, a database of all translated chloramphenicol acetyltransferase genes in GenBank were downloaded (on 22 Feb 2022) and screened for the presence of those amino acid sequences that would allow the incorporation of the *td* exon sequence into the coding region. Of the 1,386 proteins screened only one (GenBank Accession No. WP_111269875) was found to possess an

amino acid sequence (LTQEI, frame 1 in Table 19) capable of being encoded by the *td* exon sequence GACCCAAGAGA. This was a 216 aa protein sequence designated as a type A chloramphenicol O-acetyltransferase (EC. 2.3.1.28) translated from a gene identified in a genome assembly of a *Paenibacillus silvae* strain isolated in China in 2018. The identified LTQEI motif was ideally located within the protein, beginning at amino acid position 29 of 216, where the insertion of the group I intron would be expected to inactivate the protein (**figure 20**). The occurrence of the sequence towards the 3'-end of the gene, for instance, could result in a protein truncated at its COOH-terminus that could conceivably retain some activity when in the RAM form.

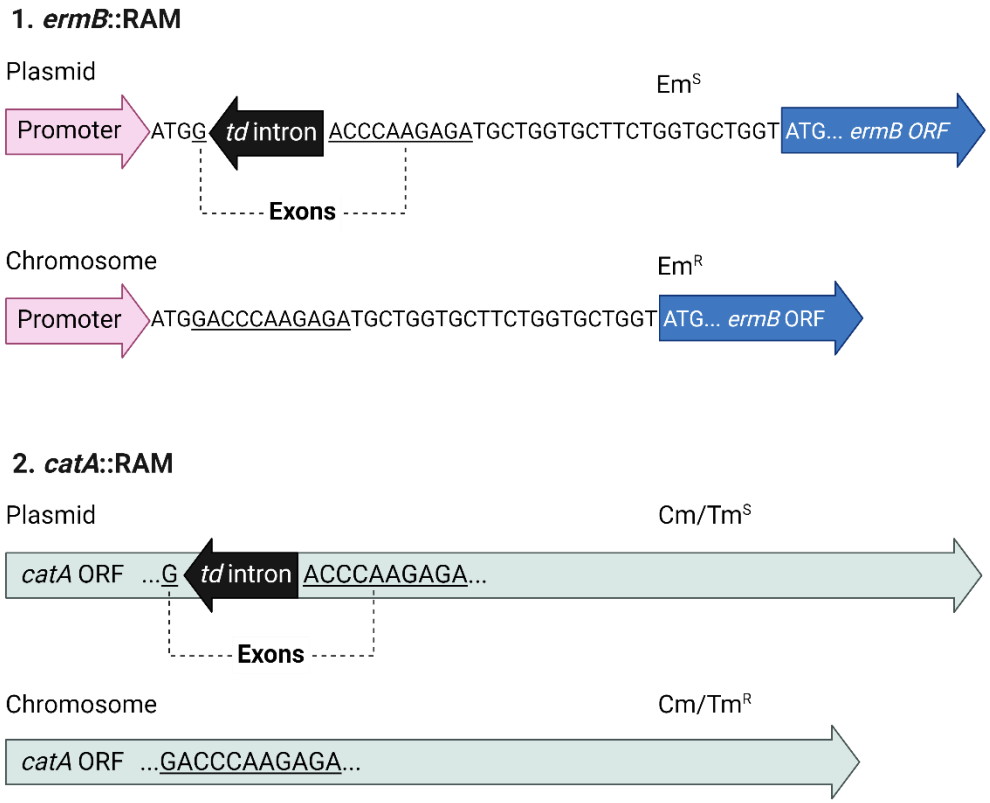

Figure 20 - Illustration of the ermB::RAM (1) and putative catA::RAM (2) genetic layout when on a plasmid and after integration into the chromosome.

Excision of the td intron restores functionality of each gene upon integration into the chromosome conferring antibiotic resistance. Created with BioRender.com.

5.2.2. Screening for progenitor *tetM*::RAM sequences

All 4,632 translated *tetM* genes from GenBank were downloaded (on 22 Feb 2022) and screened for the appropriate amino acid sequences that would enable insertion of the *td* group I intron. This revealed thirteen potential *tetM*::RAM candidates. These were further filtered based on the position of the amino acid motif within the sequence, a *tetM* gene with the *td* intron inserted toward the 3' end of the coding region results in the expression of a protein truncated toward its C-terminus and would likely retain function. Three of the remaining TetM proteins were chosen based on their homology and similarity in length to TetM from the transposon Tn916 in *Enterococcus faecalis*.

The first from genome sequencing data of *Lactobacillus johnsonii* strain G2A isolated in the USA featuring the amino acid motif QTQEI (Accession No. QIA88364), the second from *Streptomyces* sp. S3 isolated in Brazil with the amino acid motif DPRE (Accession No. NNN30134), and the third found in multiple genome entries but features most prominently in strains of *Staphylococcus aureus* features the amino acid motif RPKR (Accession No. WP_025642386).

5.2.3. Establishing CatA is functional

While some of the genes/protein of the 1386 members of the database screened have been experimentally shown to confer chloramphenicol resistance on their host organisms, many of the representatives were derived from genome sequencing projects. This was particularly true of those 63 CatA genes/proteins found in *Paenibacillus* genomes, the only class of Cat protein that contained examples in possession of the requisite LTQEI peptide sequence required to build a RAM. It follows that despite the apparent homology, there was no direct evidence that the predicted type A chloramphenicol O-acetyltransferase had activity and could confer the requisite chloramphenicol/thiamphenicol resistance to the host. Accordingly, the initial step was to confirm that the identified CatA protein was active.

As the *Paenibacillus silvae* strain was not available in public culture collections, the *catA* gene was synthesised using clostridial codon usage and incorporating the requisite *td* group I intron sequence into the LTQEI-encoding region. Additionally, the gene was synthesised together with the strong, constitutive P_{thl} promoter of *C. acetobutylicum* immediately 5' to the translated started codon of the *catA* gene. Following synthesis, the gene and promoter were cloned into the multiple cloning site of the *ermB*-based vector pMTL83251 and the resultant plasmid introduced into the *E. coli* donor sExpress. This strain was then used to transfer the plasmid into a *Clostridium butyricum* recipient, selecting for Em^R resistant transconjugants on agar media supplemented with Em. Following incubation of the mating plates, three independent Em^R

transconjugant colonies were selected and inoculated into liquid BHIS media supplemented with clarithromycin and incubated overnight at 37 °C. The following day an aliquot of each of these cultures was normalised to an OD_{600} of 1 and 10 µL of the normalised culture was used to inoculate 1 mL of BHIS media supplemented with increasing concentrations of antibiotic relevant to selection of the *catA* gene. This was thiamphenicol in the case of *C. butyricum* (**figure 21**) and chloramphenicol in the case of *E. coli* (**figure 22**). After 72 hrs the OD_{600} of the cultures were measured.

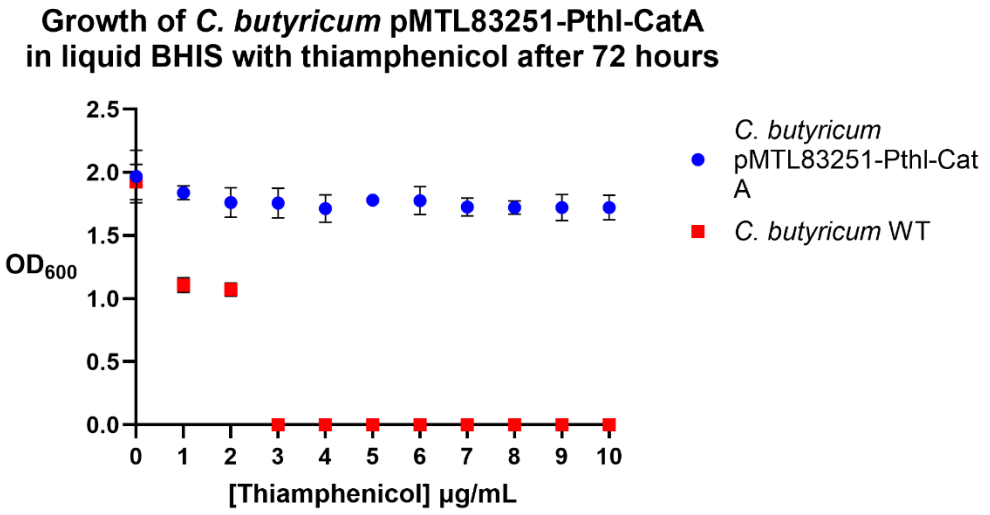

Figure 21 - Minimum inhibitory concentration of thiamphenicol in C. butyricum harbouring catA (blue circles) and the wild type (red squares). N = 3.

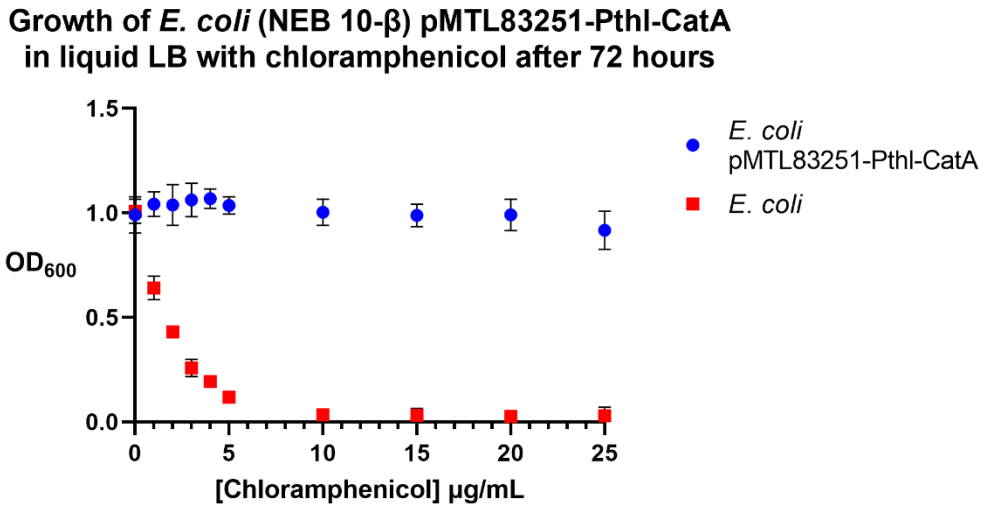

Figure 22 - Minimum inhibitory concentration of chloramphenicol in E. coli NEB 10β harbouring catA (blue circles) and E. coli NEB 10-β (red squares). N = 3.

5.2.4. Establishing TetM RAM candidates are functional

As with when screening for *catA*, many of the TetM proteins screened were sourced from genome sequencing projects. Consequently, there was only circumstantial evidence that the RAM candidate TetM proteins discovered in **5.2.2.** could confer resistance to tetracycline due to their shared homology with other experimentally validated *tetM* genes. As such, it was necessary to gather direct evidence that the *tetM*::RAM candidates could confer antibiotic resistance when expressed.

This was approached in the same manner as when validating functionality of the CatA protein. Each *tetM* gene was individually synthesised with the same constitutive *C. acetobutylicum* P_{thl} promoter immediately to the 5' end of their respective start codons as single DNA fragments. These fragments were then

cloned into the multiple the cloning site of separate pMTL83251 vectors via HiFi.

Only the third *tetM* from *Staphylococcus aureus* was able to be successfully cloned into a completed vector, named pMTL83251-Pthl-TetM-Sau, which was subsequently transformed into a sExpress *E. coli* donor. This was used to conjugatively transfer the vector to *C. butyricum* but no resistance to tetracycline was observed in transconjugant colonies, even at concentrations as low as 1 µg/mL.

5.2.5. Testing of a *catA*::RAM in *Clostridium*

Having established that the *Paenibacillus silvae catA* gene did indeed confer resistance to chloramphenicol or thiamphenicol in *E. coli* and *C. butyricum,* respectively, experiments were undertaken to establish whether it could be modified to become a RAM and be used in ClosTron mutagenesis. To progress, in the first instance, a control ClosTron plasmid was built (as described in **2.4.7.2**) based on the tried and tested plasmid pMTL007C-E2 that targeted the *pyrE* gene of *C. butyricum.* The retargeted plasmid was designated pMTL007C-E2::Cbut_PyrE-381;382s. This pMTL007 plasmid makes use of the *C. perfringens catP* gene (**C**) as its selectable marker and uses the *ermB*::RAM (**E2**) as its intron targeting system. The predicted changes made to retarget the intron were designed to insert between position 381 and 382 of the sense strand (**s**) of the *C. butyricum pyrE* gene. Hence the plasmid designation, pMTL007**C**-**E2**::Cbut_PyrE-381;382**s**. In parallel, an equivalent plasmid was made using the *catA*::RAM in place of the *ermB*::RAM and in which the *catP* selectable plasmid

marker was replaced with *ermB*. The new plasmid was designated pMTL007E-C::Cbut_PyrE-381;382s.

The two retargeted plasmids were transferred into in *C. butyricum* with initial selection for the plasmid-borne selectable marker, either *ermB* (pMTL007C-E2) or *catA* (pMTL007E-C). Thereafter, putative transconjugants colonies were restreaked onto media supplemented with the antibiotic corresponding to the RAM being employed. Those colonies that developed were then screened by an appropriate cPCR to establish whether intron insertion at the *pyrE* locus had taken place. In total 24 colonies from each RAM were screened, all of which generated a DNA fragment consistent with successful insertion of the group II intron at the *pyrE* locus (**figure 23**).

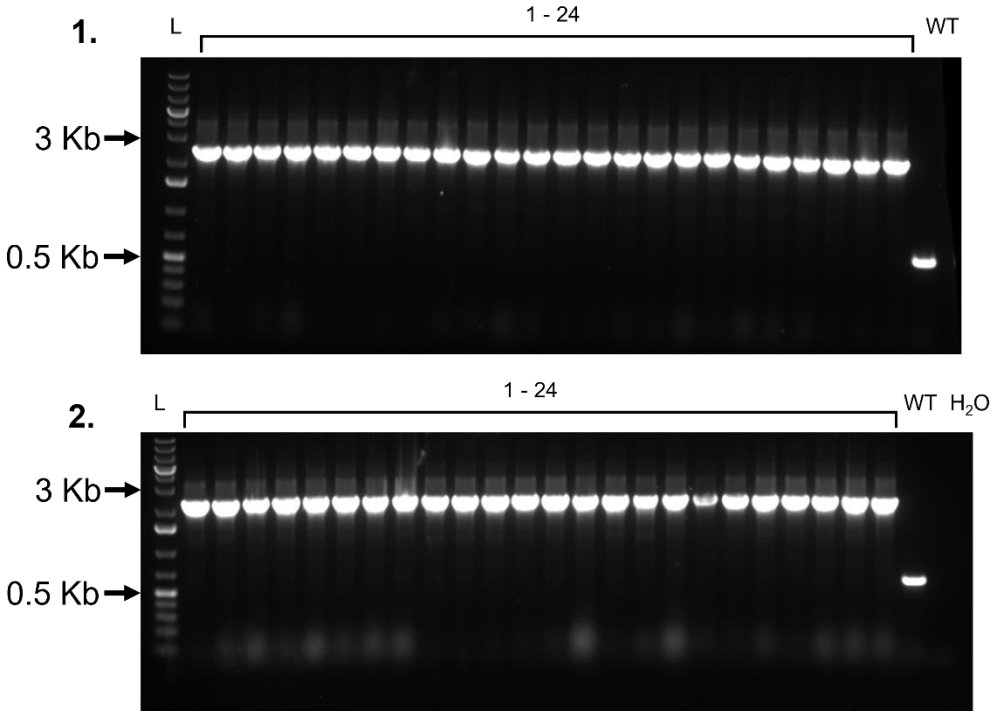

Figure 23 – Agorose gel electrophoresis cPCR of colonies formed on ermB::RAM (1) and catA::RAM (2) selection plates using ClosTron vectors targeting PyrE at base 381 in C. butyricum. Primers Cbut_pyrE_F and Cbut_pyrE_R.

L = Generuler 1kb plus DNA ladder (ThermoFischer), lanes 1 – 24 = screened colonies, and WT = wild type C. butyricum gDNA as a template.

Equivalent retargeted plasmids based on pMTL007E-C were constructed for mutant generation at the *pyrE* locus in *C. sporogenes, C. carboxidivorans* Δ7RM and *C. difficile* R20910 designed to insert the group II intron at bases 470|472s, 123|124a, and 414|415s, respectively.

Putative Tm^R mutants were generated as before and screened using appropriate primers flanking *pyrE* in a diagnostic cPCR. In the case of *C. sporogenes*, of the 13 Tm^R clones screened, only three clones appeared to contain an inton insertion (**Figure 24, panel 1**). All 16 of the *C. carboxidivorans* Δ7RM Tm^R clones screened resulted in a large, amplified DNA fragment indicative of insertion (**Figure 24, panel 2).** In contrast, no intron containing colonies of *C. difficile* R20291 were initially found after screening. Mutagenesis was attempted with a new targeting region in parallel with a second attempt of the initial targeting region. One mutant was found amongst colonies from the initial targeting region during screening of the second attempt and none were found in the new targeting region. In total 24 of each targeting region was screened.

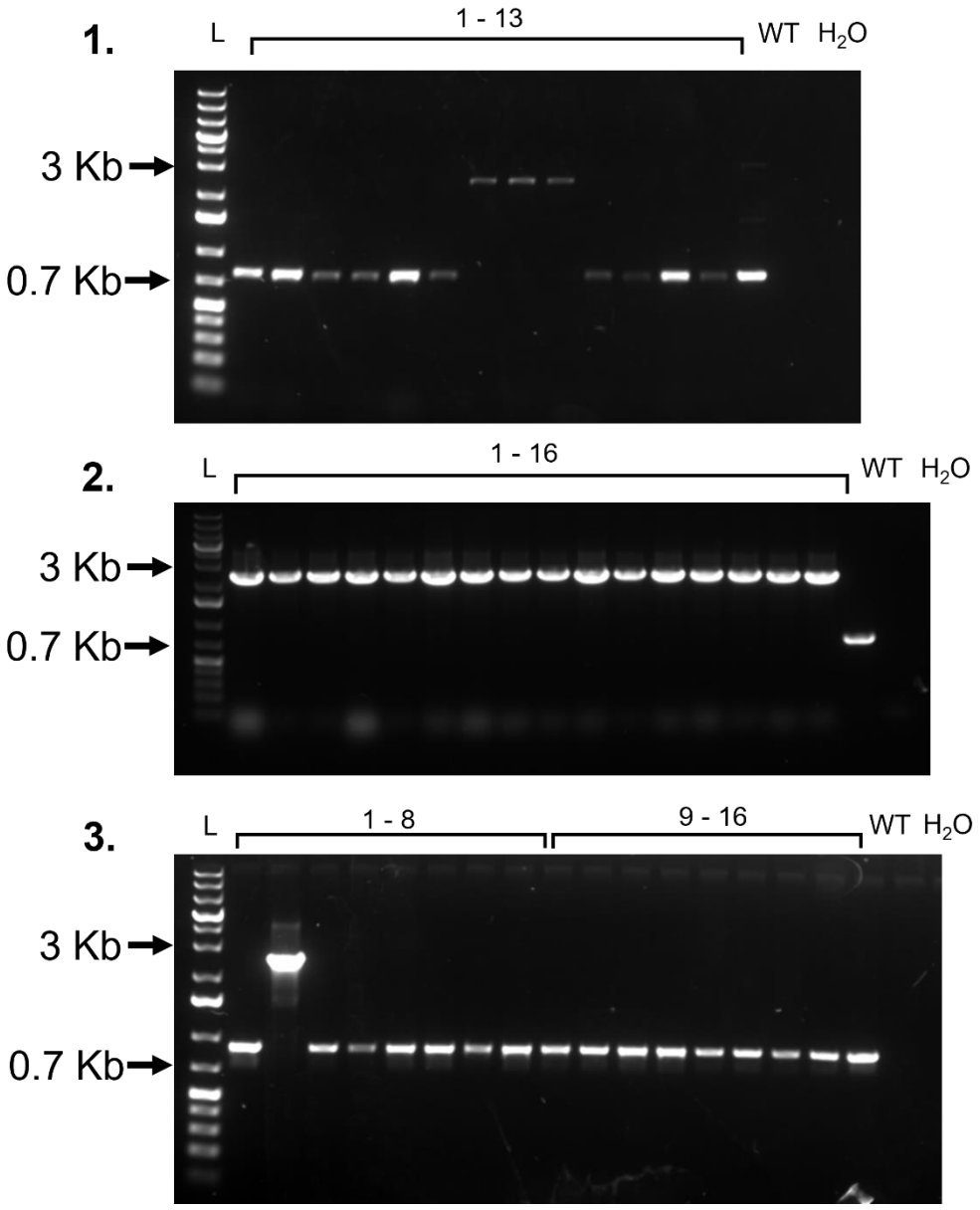

Figure 24 – Agarose gel electrophoresis cPCR of colonies formed on thiamphenicol selection plates after conjugation with catA::RAM ClosTron vector. Primers used in table 6.

Primer pair flanking pyrE. 1 - C. sporogenes. 2 - C. carboxidivorans Δ7RM. 3 - C. difficile R20291, 1-8 = targeting region 414|415s, 9 – 16 = targeting region 213|214s. L = GeneRuler 1kb plus DNA ladder. WT = wild type gDNA template.

5.2.6. Knockout Validation

The *pyrE* gene encodes the enzyme orotate phosphoribosyltransferase, which plays an essential role in pyrimidine biosynthesis. If this gene is knocked out or

otherwise rendered non-functional, then the host cell will display a uracil auxotrophic phenotype (Ng *et al.*, 2013). This provides a simple way of determining if the integration of the group II intron with the *catA*::RAM results in an effective knock out of the target gene by plating the mutants on minimal media agar supplemented with and without uracil (recipe **2.6.1.**).

A *catA*::RAM ClosTron *pyrE* mutant generated in **5.2.2** from *C. butyricum*, *C. difficile* R20291, and *C. carboxidivorans* was plated on both uracil supplemented and uracil deficient minimal agar plates as described in **2.6.1**. After 72 hours *pyrE* mutants from *C. butyricum* and *C. difficile* R20291 exhibited uracil auxotrophy as evidenced by lack a of growth on uracil deficient plates (**figure 25**). *C. carboxidivorans*, however, did not grow on the minimal media.

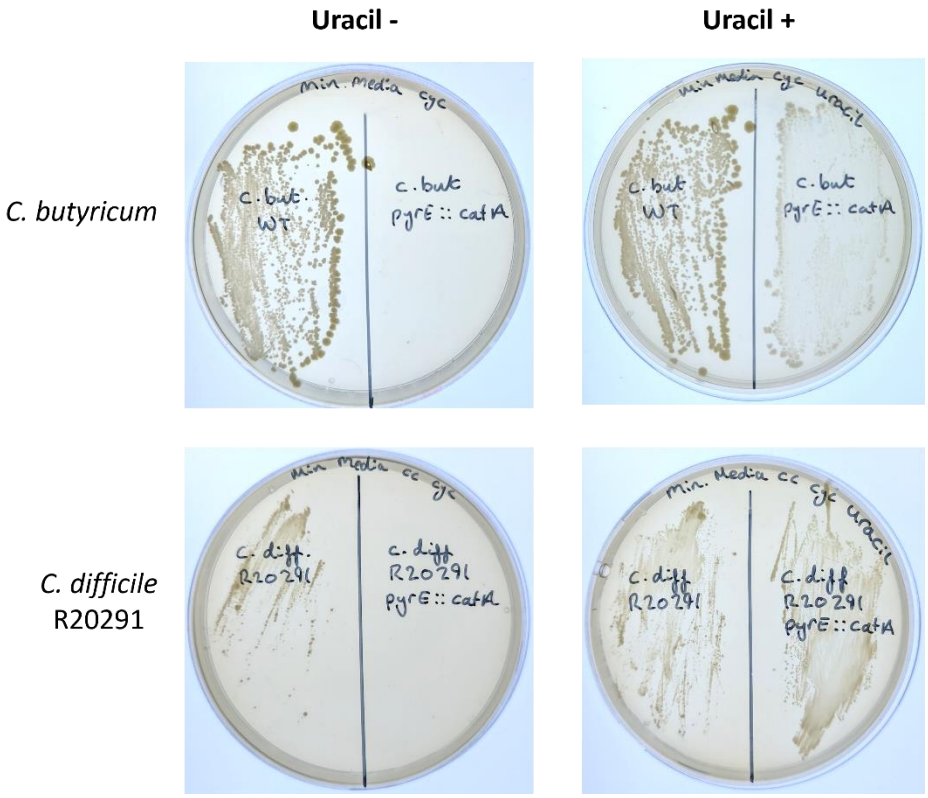

Figure 25 - Wild type (left half of plate) and CatA RAM PyrE mutants (right half of plate) on uracil deficient (left) and uracil supplemented (right) minimal media.

5.2.7. TetM Tn916 Minimum Inhibitory Concentration

The availability of two RAMs based on two different antibiotic resistance genes should allow the generation of double mutants in which the initial mutant is first selected using one antibiotic, and then the second mutant is selected using the other antibiotic. However, such a strategy is reliant on having a third selectable marker by which the two plasmids can be selected on their initial introduction into the cell. *TetM* from the *Enterococcus faecalis* transposon Tn916 (Accession No. P21598) was chosen as a candidate as a new (to *Clostridium*) tetracycline resistance conferring antibiotic marker. This was because the *tetA* marker originally used in the pMTL80000 series of vectors did not confer resistance in *C. acetobutylicum*, *C. beijerinckii*, and *C. botulinum* (Heap *et al.*, 2009) and so an alternative marker would widen the potential applications for double mutant generation using the ClosTron system as well as expanding the molecular toolkit for clostridium in general.

An experiment to determine the efficacy of TetM Tn916 as a selectable marker was performed in a similar manner as when characterising *catA* in *C. butyricum* and *E. coli* (**5.2.3.**) using tetracycline (**figure 26**).

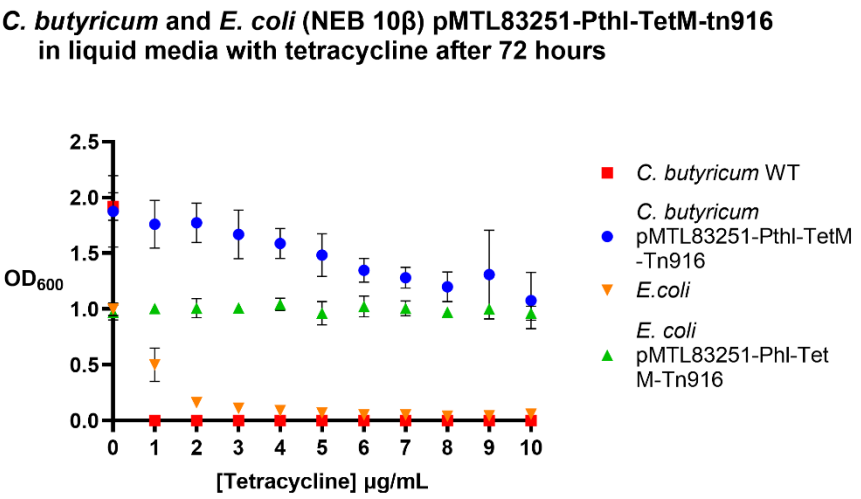

Figure 26 - Growth of C. butyricum and E. coli pMTL83251-Pthl-TetM-Tn916 in liquid media.

OD_{600} values taken 72 hours post inoculation. Red squares: C. butyricum Wild Type. Blue circles: C. butyricum with tetM Tn916 harbouring plasmid. Gold inverted triangles: E. coli NEB 10β. Green triangles: E. coli NEB 10β with tetM harbouring plasmid. N= 3.

Since mutant screening is performed on agar plates the effectiveness of *tetM* on solid agar plates was also investigated. Three liquid cultures of *C. butyricum* pMTL83251-Pthl-TetM supplemented with clarithromycin was prepared, the following day each culture was normalised to an OD_{600} of 1 and serially diluted up to 1:100,000. 100 µL of the final dilution was spread onto BHIS agar plates with increasing levels of tetracycline (**figure 27**).

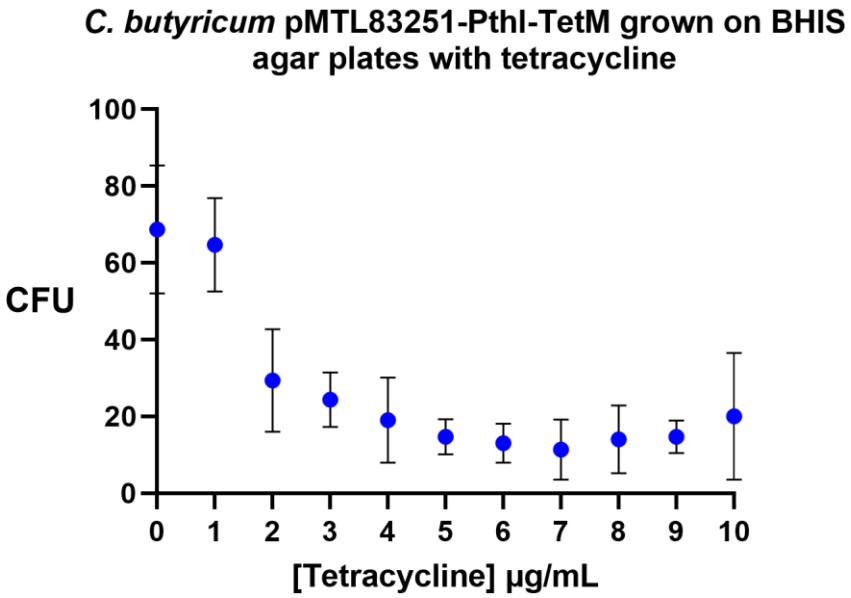

Figure 27 - C. butyricum pMTL83251-Pthl-tetM (blue circles) grown on agar plates supplemented with tetracycline. N= 3.

5.3 Discussion

5.3.1 catA::RAM Testing

The *catA* gene from *Paenibacillus silvae* has been shown to be an effective antibiotic marker in *C. butyricum* when expressed on a plasmid as well as when in a single copy integrated into the chromosome. When the *td* group I intron was inserted into *catA* and used as a RAM in the ClosTron system it was shown to have the same integration efficiency as the *ermB*::RAM when targeting *pyrE* in *C. butyricum*. However, only one targeting region was tested. To make a more confident assessment of the *catA*::RAM performance in comparison to the *ermB*::RAM more targeting regions in different species must be compared. Lack of growth on uracil deficient minimal media indicated that the *catA*::RAM harbouring intron has successfully inactivated the *pyrE* gene.

The *catA*::RAM was successfully used to select for *pyrE* mutants in *C. carboxidivorans* Δ7RM and was integrated with 100% efficiency. However, neither *C. carboxidivorans* Δ7RM *pyrE* mutant or its parent strain grew on the minimal media with or without uracil. This is likely because the minimal media used is lacking an essential nutrient for *C. carboxidivorans*.

In *C. sporogenes* only 3 of 13 colonies screened were mutants, this is could be because *C. sporogenes* has two chloramphenicol acetyl transferase genes (locus tags: CLSPO_c20020 and CLSPO_c20330) in its genome and so likely exhibits some degree of resistance to thiamphenicol. It is possible that plating cultures on thiamphenicol plates could be selecting for mutants within the population that have a natural increased resistance to thiamphenicol that has not been provided by integration of the *catA*::RAM. This was anticipated during the design of the experiment, and it was decided that for *C. sporogenes* the concentration of thiamphenicol on selection plates should be increased from 7.5 µg/mL to 15 µg/mL to ensure integrant selection. This is in line with experiments using *C. sporogenes* from the literature, which typically use 15 µg/mL of thiamphenicol when selecting for transconjugants (Heap *et al.*, 2010; Kuwana *et al.*, 2024). However, *C. butyricum* also has native chloramphenicol acetyltransferase genes (locus tags: ATN24_00695, ATN24_10415, ATN24_10680) but the same observations were not found when testing the *catA*::RAM. The minimum inhibitory concentration of thiamphenicol is only 2 µL/mL in *C. butyricum* and so natural variation in thiamphenicol tolerance amongst individuals in a population is unlikely to exceed the 7.5 µg/mL used during *catA*::RAM integrant selection. It could be that the minimum inhibitory

concentration of thiamphenicol in *C. sporogenes* is closer to 15 µg/mL allowing some particularly resistant individuals to escape selection. However, without knowing the minimum inhibitory concentration of thiamphenicol to *C. sporogenes* this is merely speculative.

A more likely explanation is that the targeting region chosen for the type I intron was not particularly efficient at integrating into the *pyrE* locus and could be integrated elsewhere in the chromosome, conferring resistance to thiamphenicol. Considering the Perutka algorithm used to generate intron targeting regions was developed 20 years ago (Perutka *et al.*, 2004) and that Sigma Aldrich use their own (presumably better) propriety paid-for-use algorithm, it would be worthwhile to revise this algorithm to produce more efficient targeting region sequences.

A similar lack of efficiency was observed when trying to integrate a *catA*::RAM harbouring intron into the *C. difficile* R20291 chromosome, attempted with two different targeting regions only yielded 1 of the 52 screened colonies showing integration at *pyrE*. Again, this is likely due to inefficiency of both targeting regions especially when considering *C. difficile* R20291 is sensitive to thiamphenicol. The targeting region *spo0A* 177|178a in *C. difficile* R20291 was shown to have 100% efficiency using the *ermB*::RAM with 6 of 6 colonies screened being integrants (Heap et al., 2010). In a bid to test the CatRAM in *C. difficile* R20291 with a targeting region known to be efficient, this targeting region was used in a *catA*::RAM ClosTron vector but yielded no colonies on thiamphenicol selective plates when conjugated.

An identical CatA (Accession No. PZT55867) was also found to be present in a second strain of *Paenibacillus silvae*, also isolated at this time in China. Since this date three further sequences have emerged that retain the LTQEI motif, but exhibit very small differences elsewhere in the sequence, two from further isolates of *Paenibacillus silvae* (WP_188593043.1 and WP_248061231.1), both of which share 98% identity) while a third (WP_145402878.1) is from *Paenibacillus xylanexedens* and shares 99% identity.

Furthermore, performing a search of the UniProt Knowledgebase (UniProtKB) (Consortium, 2020) for the chloramphenicol O-acetyltransferase Enzyme Class number (EC. 2.3.1.28) produces 4,218 amino acid sequences (accessed 29 September 2024). Downloading and screening these results for the appropriate amino acid motif that allows the insertion of the group I *td* intron results in 7 additional RAM candidates.

Interestingly, one of these can be found in a strain of *Clostridium pasteurianum* isolated in the USA featuring the amino acid motif DPRE (Accession No. KRU12945). Given that this Cat gene can be found in a fellow member of the *Clostridium* genus, it has the potential to be more effective in conferring antibiotic resistance than *catA* from *Paenibacillus sylvae*, due to its close phylogenetic proximity. However, it should be noted that a separate entry with an identical sequence, from an isolate of *C. pasteurianum* from Russia, is annotated as virginiamycin A acetyltransferase (Accession No. AJA51047). Therefore, it would be worth characterising its functionality experimentally in other members of *Clostridium*.

Three of the chloramphenicol acetyltransferase proteins can be found in strains belonging to the *Bacteroides* genus, all of which were isolated in the United Kingdom and share a strong homology to each other (≥ 94 % identity). Two have the amino acid motif GTQEK and are from strains of *Bacteroides faecis* and *Bacteroides thetaiotaomicron* (Accession No. CUP02961 and CUP18257, respectively). The third, also from a *Bacteroides thetaiotaomicron* isolate, features the amino acid motif STQEK (Accession No. CUQ02832).

Another potential RAM candidate was sourced from metagenomic sequencing data from healthy human faeces samples acquired in the United Kingdom (Browne *et al.*, 2016). It reveals that the genome of an uncultured species from the genus *Blautia* encodes a type III chloramphenicol acetyltransferase featuring the amino acid motif TTQET (Accession No. SCH66085).

Two further candidate putative chloramphenicol acetyltransferases with the appropriate amino acid motifs can be found from members of the *Parabacteroides* genus. The first of which can be found from genomic sequencing data of *Parabacteriodes sp.* strain D13, used as a reference genome for the Human Microbiome Project (Turnbaugh *et al.*, 2007), which contains a *catB* gene that when translated contains the motif DPRD (Accession No. EEU52731). The second is a type III chloramphenicol acetyltransferase found in a strain of *Parabacteriodes merdae* isolated in the USA featuring the motif ATQEI (Accession No. MTV00432).

5.3.2 Double ClosTron Knockout

Since erythromycin and thiamphenicol are both being used for selection of a double integrant strain, a third selectable marker was required for maintaining the ClosTron plasmid during the generation of the second knockout. The pMTL80000 series of vectors has selectable markers for tetracycline and spectinomycin (*tetA* and *aad9*, respectively), but common conjugal donor strains CA434 and sExpress are resistant to spectinomycin and the TetA gene has only showed functionality in a limited number of Clostridium. For these reasons, it was decided that an alternative selectable marker conferring resistance to tetracycline should be used.

The *tetM* gene from the *Enterococcus* transposon *Tn916* was chosen and tested in a similar manner to *catA* by constructing a plasmid that contained that gene controlled by an upstream thiolase promoter. In liquid culture *tetM* appeared to function well, maintaining an OD_{600} above 1 in *C. butyricum* up to a tetracycline concentration of 10 μg/mL, albeit with a downward trend as antibiotic concentration increased. However, when a similar experiment was performed on solid media the number of colonies would decrease dramatically at concentrations of tetracycline at 2 μg/mL or higher. In a separate experiment, when re-streaking individual colonies from clarithromycin to tetracycline supplemented agar plates no colonies would grow.

These results suggest that TetM Tn916 may only confer resistance to a small population of transconjugant *C. butyricum* cells. Attempting to use it as a

marker to select for transconjugants could diminish the efficiency of DNA transfer to a point that it is unusable.

Without enough time to test another antibiotic marker the attempts at producing a double knockout was abandoned. In this instance of modifying *pyrE* deficient mutants, a workaround could have been made by providing a recombinant *pyrE* gene on the ClosTron vector along with an antibiotic selection marker specific to *E. coli*, such as ampicillin. The plasmid could be maintained in the *E. coli* donor strain using the antibiotic and the recombinant *pyrE* gene would restore uracil biosynthesis upon successful DNA transfer and would thus grow on uracil deficient media. The transconjugant colonies could then be restreaked onto plates supplemented with antibiotic appropriate for the RAM being selected for.

If this approach were to work, it would only serve as a proof of concept for the delivery of two separate group II introns harbouring different RAMs at different loci. The system itself would be limited for use in *pyrE* mutant strains so is unlikely to see much use in wider research given that constraint. A better alternative would be to develop more RAMs in a group of microorganisms that have a wider range of selectable markers to choose from and avoid this issue altogether.

5.4 Conclusions

The *catA*::RAM has been shown to be capable of acting as a means for selecting for group II integrants and inactivating gene function as part of the ClosTron mutagenesis system. This allows for the use of the ClosTron system in organisms

that are resistant to erythromycin but sensitive to thiamphenicol. In *C. sporogenes* and *C. difficile* R20291 the *catA*::RAM appears to integrate at low frequencies, but a more thorough investigation is required to determine if there is any difference in the integration frequencies when compared to the old *ErmB*::RAM. This could be done by comparing the efficiency of integration between ClosTron vectors targeting the same region harbouring either a CatA or ErmB RAM in more species.

Given more time it would also be prudent to test the *catA*::RAM on important members of clostridium such as *C. botulinum* and *C. acetobutylicum* to encourage its use in wider research.

The results demonstrate that there is potential left in utilising group II introns for gene knockout studies that has perhaps been neglected in favour of more *in vogue* gene editing methods such as CRISPRcas9. When considering the access to large sequence databases of genes and proteins there is scope to mine these for the *td* exon sequence to create many more RAMs and perhaps also exploit genes other than antibiotic markers, which could enable more sophisticated functions in the technology.

Chapter 6 – Conclusion and Further Studies

This body of work describes the full genetic domestication of *C. carboxidivorans* P7 through the application of synthetic biology tools to achieve in-frame deletions of seven nuclease encoding genes associated with *C. carboxidivorans'* restriction modification systems, the most achieved to date in a single strain. The substantial levels of competency observed in the final domesticated strain shows that even seemingly impregnable barriers to DNA transfer can be overcome with the correct approach. This presents a case study where the same workflow can be applied to the domestication of other genetically incalcitrant species. Furthermore, *C. carboxidivorans* Δ7RM represents a platform for biochassis that can underpin a circular economy through the valorisation of waste carbon and the sustainable production of chemicals and fuels from greenhouse gas emissions. Additionally, the strain can be used in academic research to further elucidate the chain elongation mechanism and inspire metabolic engineering strategies for other closely related medium-chain organic acid and alcohol producing members of *Clostridium*, such as *C. drakei*, *C. kluyveri*, and *C. muellerianum*. Unfortunately, full characterisation of the fermentation profile of *C. carboxidivorans* Δ7RM and *C. carboxidivorans* Δ7RMΔ*hytA* could not be completed due to difficulties encountered when trying to adapt strains to autotrophic growth and time limitations. Analysis of the genome sequence of the domesticated strain and the wild-type strain displaying the same issues indicate that this was not a result of the

domestication process but rather an issue with the inoculum preparation protocol.

Chapter 5 describes the first time a new RAM has been available in the ClosTron technology since initially developed (Heap *et al.*, 2007) and the first time that this technology has been used for the mutagenesis of a strain of *C. carboxidivorans*. The development of the new *catA* RAM could rejuvenate ClosTron as a mutagenesis technology attracting renewed interest from scientists looking for an alternative knockout technology. Furthermore, the work demonstrates the importance of the development and maintenance of large bioinformatics databases, which have the potential to breathe new life into older technology. It also serves as a proof of concept for how researchers could develop their own RAM for a specific group of organisms should it be required.

6.1. Further Studies

First and foremost, the priority for further research should be focused on the characterisation of *C. carboxidivorans* Δ7RM autotrophic fermentations to ensure that its metabolic profile is comparable to that of the wild-type strain. This can be followed by metabolic engineering studies to improve growth and alcohol production when using CO_2 or CO as a carbon source.

Knockout targets for metabolic engineering can be inferred by utilising the Transposon Insertion Sequencing technology used to determine the essential gene set of *C. autoethanogenum* (Woods *et al.*, 2022). By inoculating a continuous gas fermentation using a transposon library of *C. carboxidivorans*

Δ7RM, knockout targets that are beneficial for autotrophic growth can be deduced by analysing which transposon disrupted coding regions are overrepresented within the population. A tentative assessment of the viability of this was performed by conjugating the transposase/transposon harbouring plasmid pMTL-CW21 (based on the pMTL-CW20 plasmid from (Woods *et al.*, 2022)) into *C. carboxidivorans* Δ7RM using the sExpress conjugal donor, producing between 5 and 20 colonies per conjugation. Transposon Insertion Sequencing is dependent on high levels of DNA transfer to be effective as to minimise the risk of artificially overrepresenting transposon insertion points within a population. The low levels of conjugation efficiency observed means that alterations would need to be made to the pMTL-CW21 plasmid, such as a different Gram-positive replicon, and the conjugation protocol optimised to improve DNA transfer.

Another metabolic engineering strategy to increase ethanol, butanol, and hexanol production, as well as cell growth on CO_2 and H_2, is to identify bottlenecks in enzyme expression. This can be performed through proteomic and transcriptomic analysis of samples of *C. carboxidivorans* growing on CO_2 and H_2, allowing the expression levels of enzymes involved in solventogenesis and the WLP to be determined. Enzyme-encoding genes that exhibit comparatively low levels of expression can then be identified as potential bottlenecks, and their promoter regions replaced with a strong constitutive promoter, such as P_{thl} from *C. acetobutylicum* ATCC 824, using RiboCas.

C. carboxidivorans' unique carbon chain-elongating capability can be exploited to produce non-native products that are inaccessible in other carbon fixing chassis. For example, ethylene glycol is used as a precursor in the production of polyethylene terephthalate (Djapovic *et al.*, 2021), which is widely used in packaging, plastic bottles, and fabrics (Muringayil Joseph *et al.*, 2024). However, ethylene glycol is synthesised via the hydration of ethylene oxide, which is sourced from petroleum (Yang *et al.*, 2020), creating a demand for sustainable sources of ethylene glycol. One proof-of-concept study by Bourgade *et al.* (2022) demonstrated that ethylene glycol can be produced by *C. autoethanogenum*. A six-step synthetic metabolic pathway was designed *in silico* using cheminformatics tools to convert acetate—a natural product of *C. autoethanogenum*—into ethylene glycol. The appropriate enzymes were cloned onto a pMTL80000 vector and conjugated into *C. autoethanogenum,* with the resulting strains grown in fructose-containing media. After 96 hours the maximum concentration of ethylene glycol produced was 7.61 mM. Whilst this was conducted under heterotrophic growth it demonstrates that non-native product formation is feasible in acetogens, albeit with further strain optimisation required if the same results are to be achieved in gas fermentation. The synthetic metabolic pathway designed in this study consists of a chain of sequential functional group modifications starting from acetate; therefore, a similar approach could be applied to convert hexanoate produced by *C. carboxidivorans* to 1, 6-hexanediol, a petroleum derived precursor to polyurethane (Datta & Kasprzyk, 2018). The production of precursor chemicals from CO and CO_2 which are then used for manufacturing plastics would serve

as a carbon sink for greenhouse gases that would otherwise be released into the atmosphere.

Organisms that implement the WLP pathway do so under strict energy limitations. One mole of ATP is consumed from the conversion of formate to formyl-THF and one mole is produced when producing acetate from acetyl-phosphate, resulting in a net ATP yield of zero. This means that acetogens must generate ATP by coupling membrane-bound ATPases to the proton or Na^+ gradients that are formed by WLP reducing equivalents (Schuchmann & Müller, 2014). When overexpressing energy-consuming native or heterologous pathways, ATP availability dictates the limits of product formation in acetogens. Therefore, devising means to boost ATP production within the cell can expand the possibilities metabolic engineering. Through genome-scale modelling, transcriptomics, and ^{13}C metabolite tracing, it was shown that *C. drakei* couples a glycine synthase-reductase pathway (GSRP) to the WLP during autotrophic growth (Song *et al.*, 2020). This allows methylene-THF to either continue along the WLP pathway to produce acetyl-CoA or be converted into glycine via the GSRP, which is in turn reduced to acetyl-phosphate producing ATP when converted to acetate. In the same study, the GSRP enzyme-encoding genes of *C. drakei* were overexpressed in the acetogen *Eubacterium limosum* ATCC 8486, which lacks this pathway, leading to increased rates of in the rates of growth, acetate production, and CO_2 consumption. Alignment of 16S RNA sequences shows that *C. carboxidivorans* P7 and *C. drakei* are very closely related organisms (Bruant *et al.*, 2010), and BLAST analysis performed by Song *et al.* (2020) revealed that *C. carboxidivorans* P7 possesses the enzyme-encoding

genes required for the GSRP. Since no genetic tools are available for metabolic engineering in *C. drakei, C. carboxidivorans* Δ7RM could be used to exploit this discovery by overexpressing GSRP-related enzymes to increase ATP generation during autotrophic growth.

The domestication process in this study describes the use of methylation donor strains that harbour the methyltransferase encoding genes of *C. carboxidivorans* to pretreat plasmid DNA *in vivo*. By its nature this technique produces donor strains that are only compatible for the delivery of DNA to the one specific organism from which the methyltransferase genes originate. For each genetically incalcitrant organism a new donor strain must be developed, which requires identification of active RMS associated methyltransferases and subsequent cloning into a donor strain. This specialised approach can be very effective for a single organism, but these steps could be bypassed by developing a more "generalised" donor strain that has a broader application. As discussed in **1.3.3.**, bacteriophages have evolved a means to modify and protect their DNA from RMSs in way that differs from mimicking the methylation profile of their target host by incorporating unusual base modifications in their genome. The *mom* operon of the coliphage Mu converts adenine to N6-(1-acetamido)-adenine, some *Bacillus subtilis* phages have thymine completely replaced by hydroxymethyluracil (Hoet *et al.*, 1992), and the T4 coliphage modifies cytosine to hydroxymethyl cytosine or glycosylated hydroxymethyl cytosine, which has even been shown to impair DNA binding between some CRISPR-Cas systems and their target (Vlot *et al.*, 2018). The mechanisms behind these unusual base

modifications could be implemented into a donor strain that could create species non-specific RMS silent DNA.

The findings of this study are not limited to developing non-model organisms for industrial purposes. In recent years there have been efforts to exploit phages to modify bacterial genomic DNA *in situ* (Brödel *et al.*, 2024; Hsu *et al.*, 2020; Lam *et al.*, 2021; Nethery *et al.*, 2022). The research conducted in this area has typically focused on the *in-situ* modification of *E. coli*. Developing this technology to target other species that make up the microbiota of the gut and important pathogens is of paramount importance if it is to find widespread therapeutic use. The results in **3.2.6.** demonstrate that not all RMSs found within bacteria are made equal, having varying contributions to the amenability of DNA uptake in a species. The techniques demonstrated in this work can be used to characterise the RMSs of bacteria of importance in human health, which could lead to the targeted *in situ* genomic modification thereof.

Appendix

CatA from *Paenibacillus silvae*

>tr|A0A2W6NIU3|A0A2W6NIU3_9BACL Chloramphenicol acetyltransferase OS=Paenibacillus silvae OX=1325358 GN=catA PE=3 SV=1

MKFNRIDFDVWNRTEVFNHYMNQNTSFSLTQEIDVSVLYRFIKQHKYRFTPALIFLITTV
VNSHPAFRTSYNCEGDLGYWDRLEPLYTIFDRRSESFSAVWTSVTNDFETFHRAYTADVE
KYNGSGSLFPKTPVPENTFSFSVIPWTSFTSFNLNINNNSRYLLPIITAGRLIHQGNLIY
LPISLQLHHAVCDGYHAGLFMNSVQELAHHPNSCIF

TetM Tn916

>sp|P21598|TET9_ENTFL Tetracycline resistance protein TetM from transposon Tn916 OS=Enterococcus faecalis OX=1351 GN=tetM PE=1 SV=1

MKIINIGVLAHVDAGKTTLTESLLYNSGAITELGSVDKGTTRTDNTLLERQRGITIQTGITSF
QWENTKVNIIDTPGHMDFLAEVYRSLSVLDGAILLISAKDGVQAQTRILFHALRKMGIPTIFF
INKIDQNGIDLSTVYQDIKEKLSAEIVIKQKVELYPNVCVTNFTESEQWDTVIEGNDDLLEKY
MSGKSLEALELEQEESIRFQNCSLFPLYHGSAKSNIGIDNLIEVITNKFYSSTHRGPSELCGN
VFKIEYTKKRQRLAYIRLYSGVLHLRDSVRVSEKEKIKVTEMYTSINGELCKIDRAYSGEIVI
LQNEFLKLNSVLGDTKLLPQRKKIENPHPLLQTTVEPSKPEQREMLLDALLEISDSDPLLRYY
VDSTTHEIILSFLGKVQMEVISALLQEKYHVEIEITEPTVIYMERPLKNAEYTIHIEVPPNPF
WASIGLSVSPLPLGSGMQYESSVSLGYLNQSFQNAVMEGIRYGCEQGLYGWNVTDCKICFKYG
LYYSPVSTPADFRMLAPIVLEQVLKKAGTELLEPYLSFKIYAPQEYLSRAYNDAPKYCANIVD
TQLKNNEVILSGEIPARCIQEYRSDLTFFTNGRSVCLTELKGYHVTTGEPVCQPRRPNSRIDK
VRYMFNKIT

Candidate *tetM*::RAMs

>QIA88364.1 TetM/TetW/TetO/TetS family tetracycline resistance ribosomal protection protein [Lactobacillus johnsonii]

MKKITTGILAHVDAGKTTLSEGMLYKSGTLRKLGAVDKGTAYLDSDDLEKKRGITIFSHIARI
QTGNSELQILDTPGHIDFAQEMEETLSVLDYAILVVSASEGVTGYTRTLWSLLKKHQIPVFIF
VNKMDTLKADKENILKQLNELDDNFIEFGNQDADFYEKVATADETTLDQYLELGKIEDSAVKK
LINQRKIFPVYFGAALKLIGIDEFLAGLDKWTDGKKYTNDFGARIFKVSYDEKGERLTWVKIT
GGSLKAKTEIFPDEKVNEIRCYNGTKYQVIPQAEASEIIAVSGLKSTYPGQGLGFENDQTNFT
VQPVLTYAVKVDSANTNACLQALKQLEDENPQLHVKWNKQTQEISIDVLGKIQLEILQQLLRD
RFNLEVEFTQGKILYQESIQASVEGVGHFEPLRHYAEVHLLLTPGKNGSGLVFKNKCSLEVLP
KKWQDQVMESLSNKEHLGVLTGSPITDIEITLVGGRGSNVHTVGGDFREATYRAVRQGLMELK
AKKQVYLLEPWYQFTLRINQNQVGRAINDIERIGGKFELGESSGNVTTITGQAPVAQMQDYAT

EVRNYTHGSGQLECLFLGYRECKDSAAIIEEMAYDPLSDINNTPNSVFCSHGAGHTVVWDEVP
SHAQYPYLG

>NNN30134.1 TetM/TetW/TetO/TetS family tetracycline resistance ribosomal protection protein [Streptomyces sp. S3(2020)]

MHVLNLGILAHVDAGKTSLTERLLHSVGVIDELGSVDAGSTRTDSLALERQRGITIKSAVVSF
AVDDVTVNLIDTPGHPDFIAEVERVLGVLDGAVLVVSAVEGVQAQTRVLMRTLQRLKIPTLIF
VNKIDRRGARYDGVLRALSERLTPAVVPMGRAVGLGTRQAAFAPDRVPVDVLADHDDELLAAY
VEGTLSQDRVRTALVARTRQALVHPVFFGSAVTGAGVPELVAGIRELLPRADGDPDGPVSGTV
FKVERGPAGEKVAYVRMFSGTLRTRDRVLFGEARDEGRVTAVSVFDHGTDVREDAVVAGRIAR
LWGLTDVMIGDAVGDPREAHGHFFAPPTLETVVVPGPDTDRRALHLALAQLAEQDPLIALRHD
EVRQETSVSLYGEVQKEVIQATLAEEFGLAVGFRETTPLCIERLAGTGAAAEFIKKDANPFLA
TVGLRVDPAPAGSGVAFRLEVELGAMPYAFFKAVEDTVRETLGQGLNGWQVTDCTVTMTHSGY
WPRQSHAHQGFDKSMSSTGADFRGLTPLVLTEALRQAGSQVYEPMHRFRIEAPADTLGALLPV
LAALQAVPRTTETRGGSCVLKGAVPAARVHGLEQRLPGLTRGEGELESGFDHYAPVVRGDVPR
RPRTDHNPLNRKEYLLNVMRRVGS

>WP_025642386.1 MULTISPECIES: tetracycline resistance ribosomal protection protein [Bacillota]

MKIINIGILAHVDAGKTTVTEGLLYKSGAINKIGRVDNGTTITDSMELERDRGITIRASTVSF
NYNDTKVNIIDTPGHMDFIAEVERTLRVLDGAVLVISAKEGIQVQTKIIFNTLAKLNIPTLIF
VNKIDRKGVCLDEIYTQIKRKLTPNLAIMQSVKIKDKGDFELTNVRDDKVIQSQIIEKLLDIN
DYLAEKYINGDVITEKEYDNVFLDEVNSCNLYPVLHGSALKDIGIDELLFAITNYLPVNNDNI
TDNLSAYVYKIDRDEESRKITFLRVFSGNIKTRQEVPINDTEETFKIKSLESIMNGEIVKVDQ
VNSGDIAIISNANSLKIGDFIGEKYDRVLDIKIAQPALRASIKPYDLSKRSKLIGALFELTEE
DPFLDCEINGDTGEIILKLFGNIQMEIIESLLKNRYKIDAKFGELKTIYKERPKRNSKAVIHI
EVPPNPYWASIGLSIEPLPIGSGLLYKTEVSYGYLNNSFQNAVKDAVEKACKEGLYGWEVTDL
KVTFDYGLYYSPVSTPSDFRNLTPYVFWEALRKAGTEILEPYLKYTVQVPNDFCGRVMSDLRK
MRASIEDIIAKGEETTLSGKIPVDTSKSYQSELLSYSNGKGIFITEPYGYDIYNGESITNDIR
NNDNDSSKEGLRYLFQKQSEI

>tr|A0A7W1HN69|A0A7W1HN69_9ACTN TetM/TetW/TetO/TetS family tetracycline resistance ribosomal protection protein OS=Rubrobacteraceae bacterium OX=2740537 GN=H0W52_01925 PE=3 SV=1

MRTLNLGILAHVDAGKTTLTERLLHAVGVIDEIGRVDDGSTQTDTLTLERQRGITIKSAVVSF
VVGDVTVNLIDTPGHPDFIAEVERVLGVLDGAVLVVSAVEGVQSQTRLLMRTLQRLYIPTLIF
VNKIDRSGAQYESLLQSISERLTQAIIPMGSASGLGTRGALYTAHTASDHDFTSGLIDLIADN
DDAFLAAYIDDEATVSYGRLRGELAEQTGKALVHPVFFGSAITGAGVDELISGITELLPAAEG
DADSPLSGTVFKVERSQGGEKIAYVRMFSGTVRTRDRLRFRRDEEEKVTGVSVFERGSSVQRD
SVAAGRIGKLWGLGEVRIGDAIGEQRTTQERHYFAPPTLETVVVPSRPADKGALHVALTQLAE
QDPLINVRQDDSRQEIFVSLYGEVQKEVVGATLANDYDIDVEFRETTTICVERPIGVGTAVEL
LPRARSPTTPFLATIGLRVEPAALDSGVQFGLDVKVGSIPTHVYKTVGAFHEAMERTVLETLR
QGIYGWEVTDCSVTMTDCDYQAPPRGWPGTTASDFRLLTPLVLMGALEQAGTAVCEPIHRFHL
EIPPDTFGATVSAMARLPAAVQTQKMRRSSYVLEGEVPAARVHELQQQLPALTRGEGLVECEF
DSYRAVGGKIPTRPRTDYNPLNRKEYLLHFMRRV

>tr|A0A7W1TLS0|A0A7W1TLS0_9CHLR TetM/TetW/TetO/TetS family tetracycline resistance ribosomal protection protein

OS=Herpetosiphonaceae bacterium OX=2720503 GN=H0X37_13515 PE=3 SV=1

MTSLNLGILAHVDAGKTSLTERLLFMTGVIDKLGSVDAGNTLTDSLALERQRGITIKSAVASF
VIDDVTVNLIDTPGHPDFIAEVERVLNVLDGAVLVISAVEGVQPQTRVLMRALQRLHIPTLLF
INKIDRGGANYERVFQSIAEKLQPAIMLMGSAHEQGSRSAGYSPYGVGDAAFMARLVELLAGQ
DEALLAAYVSDEASVSYCQLRDELVVQTKQARVHPVFLGSAMTGAGVDALIAGIKELLPPAKG
DADAEVSGTVFKIERGASREKIAYVRLFAGTVGVRDRLHWGRDHEGRVTAISVFERGAAVRRG
SAGAGQIAKLWGLTEIQIGDEIGTVRTTGERRFFAPPTLETVIVPTNPADRGRLHVALVQLAE
QDPLINLRQDDIRQELFLSLYGEVQKEVIQETLLTDFHIDVQFRETTMICIERVIGSGSDVEM
LGKAANPFLATVGLRIDPAPLNSGIDFRLDAKVDSMPLFVYKSVEEFRKTMEETVQDTLRQGL
YGWQVTDCRVTLTQSGYVSPSSSARDFRLLTPLVLMEALQQAGASVCEPMHHYHLEIPTNALG
ATVAVLARLQAMPQTQEMRGSSYLLEGDIQAARVHELQQRVPGLTSGEGVLEAAFDHYEPIHG
TVPMRPRTDRNPLNRKEYLLHVLRRV

>tr|A0A5M9ZFD9|A0A5M9ZFD9_9BIFI TetM/TetW/TetO/TetS family tetracycline resistance ribosomal protection protein OS=Bifidobacterium callitrichos OX=762209 GN=EMB92_03840 PE=3 SV=1

MTRIVAGIVAHVDAGKTTLSEALLYRTGEIRKLGRVDHGDAFLDTNALEKARGITIFAHQALV
EHGDLRLTLLDTPGHVDFAAETERVLRVLDYAILVVSGIDGVQGYTETLWRLLRRYDVPVFLF
VNKADAPGFDRDAILAQLHARLSDAIHPLPTVGADPSADGSDGSAAVPFGDEIEDIAALDEHA
MEEYFDAGAITLDRVRAMIAARELFPVFFGSALKLDGVEEFLDGFAAYAREPQWPADFGARVF
KISHDDKHNRLTWLRVTGGTLKAKSLIDGEGAEKGAEKIDQVRVYNGARFDIAAELPAGSVCA
VTGLERTFPGEGLGIEPDAESPEMQPVLTYTVLPAGAAGAGADSMAGTGGESAVGASDESTPA
DRPRFDDLTLHRVLTALRELEDEDPLLHVVWVERVQEIHVQLMGAVQLEIIQQTLHERFGLDV
SFGAGSILYRETITRPIEGVGHFEPLRHYAEAHILLEPGEPGSGVHVASALSVNELDRNWQRL
ILTHLTEREHLGVLTGSPLTDVKMTLVAARAHLKHTEGGDFRQATYRAIRQGLMEARSGVVGH
AVVGTDTARPEYEIDEDEETQESNDAPGASKGTASNGTAADADARAEAVARMAKAVAADPGNC
VVLEPWYRFRLEVPQDMLGRAMADIQRMSGTFDPAVSDGEYALIEGLAPVSEMRDYAMDVNAY
THGRGRFSATFGGYRPCHDQARVIEQAAYDPESDLDNTPDSVFCAHGAGYPVKWYKVPEFMHL
DYATA

>tr|A0A1C5RM25|A0A1C5RM25_9CLOT Tetracycline resistance protein tetM from transposon Tn916 OS=uncultured Clostridium sp. OX=59620 GN=tetM PE=3 SV=1

MGKRMEKSMAEKTEHLILGILAHVDAGKTTMAESMLYHSGTIKKPGRVDHKNAFLDTFEMERS
RGITIFSKQARMNWKGRQYTLLDTPGHVDFSAEMERTLQILDYAILVISAPDGVQGHDMTLWK
LLRRYQIPVFLFINKMDMPGMDRTKILKELQKYLDSGCIDFSDAVKRKEEIEEELAMCSEELM
LEYLERQEIRQEIVKRAIRKREVFPCYFGSALKLLGVEKFLDGIHNMAELPMYPQQFGARVFK
ISRDAQGNRLTHMKITGGRLRVKQILESSSLEEKEKLDQIRLYSGAACQMTDEVQAGEVCAVT
GLQKSLAGMGYGFETEAEAPVLEPVLSYQILLPEGSDVHGTFLKLCQLEEEEPQLHMVWDERT
QEISAKVMGEVQIEVLKNLIYERFQMEVEFGAGSITYKETIAAPVEGVGHFEPLRHYAEVHLL
LEPLERGSGLQFDTDCSEDLLDKNWQRLIMTHLEERKHPGVLTGSEITDMRITLIAGKAHLKH
TEGGDFRQATYRAIRQGLKMADSLLLEPVYQFRLEIPMENVGRAMTDLNKMNGVFQSPELDGE
MAVLNGSAPVACMRDYHKEVTAYSRGRGHLFCTLKGYEVCHNQEEVINQIGYDAEADLENPTG
SIFCAHGAGFLVPWDQVYDYMHMEGSLCQKKESAEEEELPARATSAIYAASRGWGDDSELEEI
FNRTYGGGSGERIGWRRKKTAENGARTVSASTVTISQKDPEKEYLLVDGYNIIFAIPQLKELA
NLNIDSARDKLMDLLCNYQGYRKNTLILVYDAYKVEGGLGSVEKYHNIYVVYTKEAETADQYI
EKTVHEIGKKYHVTVATSDALEQKIIWGAGADRMSAKGLWEEMQQVREEIKEQYLEKSGKGGQ
KLFHNLDEELAEYLEDIRLGRKVIEGDSGK

>tr|A0A5C4MDZ1|A0A5C4MDZ1_9ACTN TetM/TetW/TetO/TetS family tetracycline resistance ribosomal protection protein OS=Mumia sp. Z527 OX=2585212 GN=FHE65_01520 PE=3 SV=1

MTSSSLVLGVVAHVDAGKTSLTERLLYDAGAVASLGSVDAGTTQTDASDLERRRGITIRASVA
TLALGDVAVTIVDTPGHPDFVAEVERSLAILDAAVLVVSAVEGVQPQTVVLWRALRRLGVPTL
VLVNKVDRSGADLERTVLQVRRRLTADVVVLSHVRGIGRGDVVVEAVPGTDPLLVEAAASLDD
DLLARWVDGATITPEEVAAALRAGVRRGALTPVLAGSAITGAGIDRVRDAITDLLAPAPASDG
PGAGTVFAIDRDERGRRAWVRWWSGELRLRERIAPDGKRPAPVTEIAVSRPGGLERSRVVRAG
EVAAVRGLDVRIGDALGTAAGRGTYRFPPPTLEAVVEPCAAAQRIAMFRGLVELAEEDPLIDL
RVDEEEGEAVVRLHGEVQKEVLAAMLDARYGVPVRFSETSAVCIERVVGTGEALDEIEVDANP
YLATIGLRVEPGARGSGVAFSPGVERGNLPPAFIAATEDGVRAALRHGLAGWEVTDCLVTMTR
SGYWPRQSHAHEAFNKAMSSVATDFRSLAPVVLAAALERAGTQVCRPIDRFEIDLPEDTLGAV
LSLVGQLGGKTTGTLPKDGFTVLTGHLPSAAVPELAQRLPDLTGGEAVLAAELDHYAPVPRGT
AAPRRARTGADPRDRTEWFRSVRR

>tr|A0A7G5LS47|A0A7G5LS47_9ACTN TetM/TetW/TetO/TetS family tetracycline resistance ribosomal protection protein OS=Mumia sp. ZJ1417 OX=2708082 GN=H4N58_11505 PE=3 SV=1
MTSSSLTLGVVAHVDAGKTSLTERLLYDAGAVAALGSVDAGTTRTDASDLERRRGITIRA
SVATLAFGDLVVTIVDTPGHPDFVAEVERSLAVLDAAVLVVSAVEGVQPQTVVLWRALRR
LEVPTLLFVNKVDRSGADLDRAVSQVRRRLTPEVVVLSRVRGTGLREVEVEAVPAGDAQV
VEAAASVDDGVLATWVDGATPALGDVARALREGVRRGALTPVLAGSALTGAGIEPLRHAI
TRLLPSAAAPDGPSSGTVFAIDRDERGRRAWVRWWSGALRLREKVAPDGRRPASVTEIAV
SRPGGLRISDAVEAGEIAAVRGLDVRIGDVLGSAAGRASYAFAPPTLQTVVEPYDPTQRI
AMFQGLAELADEDPLIGLRIDGEEGEAVVRLHGEVQKEVVAALLEERYGVLVRFSETSVV
CLERVVGTGEALDVIGVDANPYLATIGLRVEPGTRGSGVVFSPGVERGNLPPAFIAATEE
GVRAALRQGLAGWEVTDCVVTMTRSGYSPRQSHAHEAFNKAMSSVGADFRSLAPVVLMAA
LERAGTHVCRPIDRYEIDLPDDTLGAVLSLIGRLGGKTTGSTPKDGFTVLTGHLPSAAVP
ALAQRLPDLSGGEAALSAELAHHAVVPAGTAAPTRRRTGPDPRDREGWFRDVRR

>tr|A0A5C4V4E3|A0A5C4V4E3_9ACTN TetM/TetW/TetO/TetS family tetracycline resistance ribosomal protection protein OS=Streptomyces sedi OX=555059 GN=FH715_11900 PE=3 SV=1

MPFSTLNIGVLAHVDAGKTSLTERLLFDNGAVARLGSVDAGSTRTDTGELERERGITIRS
AVASFRVGRHQVNLVDTPGHPDFVAEVERAFSVLDAAVLVVSAVEGVQAHTRVLMRSLRA
AGLPTLLFVNKIDRSGARPEALLADIRARLTPAAFPLTAVTDPGTRGARACARALTDPAV
RDEVAEALAERDDTLLARLVEGRPPSVGELRALLAEAVAAGRAHPVLAGSALTGEGVGAL
TEALTAWPLVPDAADSAPPAGTVFAVERSGEGEKVAYLRLFQGRLHARRRVTFRRREPDG
ARGEFTGRISRLEVVTPDERPGGHGEVPLRAGEIGRLHGLPGVRIGDRLGHPPRADGPAR
FAPPSLETVVEPVSPERKVALHAALSALADEDPLIRTRTALDGTLSVLLYGEVQREVLGE
RLRRDFGVEAVFAPATPVYFERPAGVGTSSTELRKRGPNDYWATVGLRVEPLAPGEGRRF
ERRVEWGALPRAFHQAVEDAVGHTLRQGLHGWEVTDCLVTLVRVGWHSPNSVVADFRRLT
PIVLMRALRAAGTRVYEPCQSMELEIPSDTLPMVVGRLTSVGGRVLDSVERGGNWLLTTE
VPTRLAPEVIAALPGLTRGEGTHWSRPEGDRLVRGTPPSRPRSLDDPRDRPLDLADREEG

>tr|A0A444BLI1|A0A444BLI1_9MICO TetM/TetW/TetO/TetS family tetracycline resistance ribosomal protection protein OS=Phycicoccus flavus OX=2502783 GN=EPD83_11780 PE=3 SV=1

MSSPLVLGIVAHVDAGKTSLTERLLLEAGVLDTPGSVDAGTTRTDSMDLERRRGITIRAS
VTTFAAAGLEVTVVDTPGHPDFVAEVERSLTVLDAAVLVVSAVEGVQPQTVVLWRALRRL
GVPTVLFVNKVDRAGADPDVVLERVRRRLTPHLVPLTRVRDAGSAHAEAVPVPLDDETVV

LAVAEVDDDVLRRWAEDRPVGRDRVRAALRSAVRAGRLTPVLAGSAVTGTGVPDLPAVLA
RVVAPDRRPRAEAAPHRPAATVFAVDRDERGRRVWVRMWDGELAVRDRVVVGGRAPHPVT
EVAVSRPDGLRPEPCVGAGEVAAVRGPAARVGDTVGTPPARARYRFAPGRLESLVTPEDP
ADRPALFAALAELADEDPLIALRRSAADAEAAVTLHGEVQREVVAALLEERFGMRARFSP
PGVVLVERVTGRGAALERLGTGGNPYLATVGLAVAPGPTGSGVVFRPGVQPGRLLPAFVA
ATEEGVRTALRSGRYGWEVLDCMVTMTDSLYYPRQSRPHQGFDKSMSTVAADFRLLSQVV
VHAALARAGTVACEPVDRIEVELPAAALGAVLSAVGRLGGTPEGSRGADGWSLVTGTLPT
RSVAELTRLLPDLTGGEGSLVTRPDHHAPVRGEPPRRRAPDPDPRDREAWFRDAAR

>tr|A0A7X7VZY9|A0A7X7VZY9_9CLOT TetM/TetW/TetO/TetS family
tetracycline resistance ribosomal protection protein (Fragment)
OS=Clostridiaceae bacterium OX=1898204 GN=GX477_05725 PE=4 SV=1

MKKLVAGILAHVDAGKTTLSEGLLYLGGRTRKLGRVDRKDAFLDNYGLERARGITIFSKQAVL
DIGDVQVTLLDTPGHVDFSTEMERTLQVLDYAILVISGADGIQGHTRTLWNLLDIYRVPVFIF
VNKMDQPGTDRDKLMDEIKRDLGHGCVDFGQSRGPDFMEQVAMCDEVLLSEYIETGRIDDEHI
RESIRDRRLFPCYFGAALRLEGVEALMQGLAGYTIIPSWPDKFGARVFKISRDEQGNRLSHLK
ITGGVLRVRDVAGNGTWEEKVTQIRIYSGPKFETVNEAEAGTICAVTGLSQTRPGEGLGADEG
IITPVLEPVLSYRIMLPEDADPREILPKLRQLEDEDPALRITWDEQLREIHAKIMGEVQTEIL
QSVIKDRFGIDVSFDAGRIIYKETITNTVEGVGHFEPLCHYAEVHLLMEPGEPGSGLQFAVDC
SDEVLAPNWKNLVLSHLKEKEHKGVLTGSPITDMKITLVSGRADLRHTAGGDFREATYRAVRQ
G

>tr|A0A847ND65|A0A847ND65_9FIRM TetM/TetW/TetO/TetS family
tetracycline resistance ribosomal protection protein (Fragment)
OS=Gracilibacteraceae bacterium OX=2699748 GN=GX301_06455 PE=4
SV=1

MAKLVIGILAHVDAGKTTLSESILYLSGKIGKLGRVDNKDAYLDNYELERARGITIFSKQ
AIFETGGIQITLLDTPGHVDFSAEMERTLRVLDYAVLVISGADGVQGHTKTLWRLFEIYQ
VPVFVFVNKMDQNRMDKDMLIKNMKEQLDDGCIDFGQAETMGFYEQMAMCDEMMMEAYLE
KGHIETEQIKKAVRERKIFPCFFGSALKLEGVEQLMQGIAKYSVIPCYPDEFGAKIFKIT
RDEQGNRLTYLKLTGGKLKVKDVLTNGIWEEKVNQIRIYSGQKFEAVNEIEAGSICAVTG
LSRTRPGEGLGTEEASTVPVLEPVLFYRIILPEGCDPREMIPKLRQIEEEEPELNIVWNE
QLQEIQVRIMGEVQIEILQSLIESRFGVSVSFDEGGILYKETIANVVEGVGHFEPLRHYA
EVHLLLEPGDPGSGLQFGTECSEDMLAKNWQRLILAYLQEKEHKGVLTGSVLTDVKITLV
SGRAHNKHTESGDFREAACRAVRQGLKEAESILLEPYYAFQLELP

Other candidate *catA*::RAMs

>tr|A0A0H3J125|A0A0H3J125_CLOPA Chloramphenicol O-
acetyltransferase OS=Clostridium pasteurianum DSM 525 = ATCC
6013 OX=1262449 GN=vat PE=4 SV=1

MTIPDLNKIYPRNNDHQIVYLKNVITKDNIEVGDYTIYNDFYDDPREFENNNVLYHYPVN
NDKLIIGKFCSIACRAKFIMNSGNHSMKSLSTYTFPIFGEEWDETLNPKDAWDNKGNIEI
GNDVWIGYEAVIMSGVKIGDGAIIGTRAVVTKDIPPYAIVGGTPAKVIRKRYEDKIISKL
MEIKWWNWSYEKIQRNISYIQAGEIEKLS

>tr|A0A174JWE9|A0A174JWE9_9BACE Chloramphenicol
acetyltransferase OS=Bacteroides faecis OX=674529 GN=cat PE=4
SV=1

MKRIIDIENWERKENFNFFRHFQNPQLSITSEVECGGAKQRAKAAGQSFFLHYLYAVLRA

ANEIPEFRYRIDTEGRVVLYDAIDMLSPIKIKENGKFFTTRFPYHNDFDTFYREAKMIIE
AIPEDGDPYAAENGEVADGDYGLILLSATPDLYFTSITGTQEKKSGNNYPLLNAGKAVVR
EGKLVMPIAMTIHHGFIDGHHLSLFYKKVEEFLK

>tr|A0A174L6Z5|A0A174L6Z5_BACT4 Chloramphenicol
acetyltransferase OS=Bacteroides thetaiotaomicron OX=818 GN=cat
PE=4 SV=1

MKQIIDIENWERKENFNFFRHFQNPQLSITSEVECGGARQRAKAAGQSFFLHYLYAVLRA
ANEIPEFRYRIDPDGRVVLYDTIDMLSPIKIKENGKFFTTRFPYHNDFDTFYQEARLIID
AIPEDGDPYAAENGEVADGDYGLILLSATPDLYFTSITGTQEKRSGNNYPLLNAGKAIIK
EGKLVMPIAMTIHHGFIDGHHLSLFYKKVEEFLK

>tr|A0A174SZ25|A0A174SZ25_BACT4 Chloramphenicol
acetyltransferase OS=Bacteroides thetaiotaomicron OX=818 GN=cat
PE=4 SV=1

MKQIIDIENWERKENFNFFRHFQNPQLSITSEVECGGAKQRAKAAGQSFFLHYLYAVLRA
ANEIPEFRYRIDPDGRVVLYDTIDMLSPIKIKENGKFFTTRFPYHDDFDTFYQEARLIID
AIPEDGDPYAAENGEVADGDYGLILLSATPDLYFTSITSTQEKRSGNNYPLLNAGKAIIR
EGRLVMPIAMTIHHGFIDGHHLSLFYKKVEDFLK

>tr|A0A1C5RX40|A0A1C5RX40_9FIRM Chloramphenicol
acetyltransferase 3 OS=uncultured Blautia sp OX=765821 GN=cat3
PE=4 SV=1

MEKQIDLSSWKRKEIFDFFSHASNPYYMVTFRIDVAPLYAYVKEHHLSFYYSLVYLCTQA
INEVDAFRYTIRGTQVFYLDPRIPSFTDLKKDSEYFHIVTMPTINSLAEFNAEARKRSAA
QQFFLDTTQETDRVIYFSCLPWVDLTALTNEHDFSSPDSKNDSIPRIAWGKYVPNGDRLE
LGISIEVNHRLIDGLHIGQFAQRLEKLIGEL

>tr|A0A6N3DT21|A0A6N3DT21_9BACT Chloramphenicol
acetyltransferase OS=Parabacteroides merdae OX=46503 GN=cat3
PE=4 SV=1

MKQVIDLDNWNRKEHFAFFSAFDDPFFGVTTLVDFTDVYRQSKEQNVSFFLYSLHFLLKC
VNETDAFKLRIEKDSVVRYDTIHISPTIGREDGTFGFGFFEYDPDIDLFIQKATQEIERV
KNGTGLSFSKNTSRQDVIRYSALPWFAFSEMKHATSFKNGDSVPRISTGKLMQENSKYLL
PISVCAHHGLMDGRNVAELIRKLSDNQTAL

>tr|C7X4B1|C7X4B1_9BACT Chloramphenicol O-acetyltransferase
OS=Parabacteroides sp. D13 OX=563193 GN=catB PE=4 SV=1

MNPKKRYPRSGDNQTVYLKSVITRPNIDVGDFTIYNDFENDPRDFEKNNVLYHYPINHDR
LIIGKFCSIACGAKFIFNCANHTLKSLSTYTFPLFFEEWGLQKSEVASAWDNKGDIVIGN
DVWIGYDAVIMAGVTIGDGAIIGTRAVVTKDVESYSIVGGIPAKEIRKRFSPDIIARLQK
LQWWNWDTAKIRNSIKAIQNGDLDSLEHNTI

Bibliography

Abrini, J., Naveau, H., & Nyns, E.-J. (1994). *Clostridium autoethanogenum*, sp. nov., an anaerobic bacterium that produces ethanol from carbon monoxide. *Archives of microbiology, 161*, 345-351.

Ahring, B. K., Licht, D., Schmidt, A. S., Sommer, P., & Thomsen, A. B. (1999, 1999/04/01/). Production of ethanol from wet oxidised wheat straw by *Thermoanaerobacter mathranii*. *Bioresource Technology, 68*(1), 3-9. https://doi.org/https://doi.org/10.1016/S0960-8524(98)00077-7

Ajanovic, A. (2011, 2011/04/01/). Biofuels versus food production: Does biofuels production increase food prices? *Energy, 36*(4), 2070-2076. https://doi.org/https://doi.org/10.1016/j.energy.2010.05.019

Antonicelli, G., Ricci, L., Tarraran, L., Fraterrigo Garofalo, S., Re, A., Vasile, N. S., Verga, F., Pirri, C. F., Menin, B., & Agostino, V. (2023, 2023/11/01/). Expanding the product portfolio of carbon dioxide and hydrogen-based gas fermentation with an evolved strain of *Clostridium carboxidivorans*. *Bioresource Technology, 387*, 129689. https://doi.org/https://doi.org/10.1016/j.biortech.2023.129689

Arber, W., & Dussoix, D. (1962, 1962/07/01/). Host specificity of DNA produced by *Escherichia coli*: I. Host controlled modification of bacteriophage λ. *Journal of Molecular Biology, 5*(1), 18-36. https://doi.org/https://doi.org/10.1016/S0022-2836(62)80058-8

Atanasiu, C., Su, T. J., Sturrock, S. S., & Dryden, D. T. F. (2002). Interaction of the *ocr* gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Research, 30*(18), 3936-3944. https://doi.org/10.1093/nar/gkf518

Auernik, K. S., & Kelly, R. M. (2010). Physiological versatility of the extremely thermoacidophilic archaeon *Metallosphaera sedula* supported by transcriptomic analysis of heterotrophic, autotrophic, and mixotrophic growth. *Applied and Environmental Microbiology, 76*(3), 931-935. https://doi.org/doi:10.1128/AEM.01336-09

Bäumler, M., Schneider, M., Ehrenreich, A., Liebl, W., & Weuster-Botz, D. (2022). Synthetic co-culture of autotrophic *Clostridium carboxidivorans* and chain elongating *Clostridium kluyveri* monitored by flow cytometry.

Microbial Biotechnology, 15(5), 1471-1485. https://doi.org/https://doi.org/10.1111/1751-7915.13941

Benevenuti, C., Botelho, A., Ribeiro, R., Branco, M., Pereira, A., Vieira, A. C., Ferreira, T., & Amaral, P. (2020). Experimental Design to Improve Cell Growth and Ethanol Production in Syngas Fermentation by *Clostridium carboxidivorans*. *Catalysts, 10*(1), 59. https://www.mdpi.com/2073-4344/10/1/59

Bennett, H. J., Pearce, D. M., Glenn, S., Taylor, C. M., Kuhn, M., Sonenshein, A. L., Andrew, P. W., & Roberts, I. S. (2007, Mar). Characterization of *relA* and *codY* mutants of *Listeria monocytogenes*: identification of the CodY regulon and its role in virulence. *Mol Microbiol, 63*(5), 1453-1467. https://doi.org/10.1111/j.1365-2958.2007.05597.x

Bergkessel, M., & Guthrie, C. (2013). Chapter Twenty Five - Colony PCR. In J. Lorsch (Ed.), *Methods in Enzymology* (Vol. 529, pp. 299-309). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-418687-3.00025-2

Bernstein, D. B., Akkas, B., Price, M. N., & Arkin, A. P. (2023, Dec 6). Evaluating *E. coli* genome-scale metabolic model accuracy with high-throughput mutant fitness data. *Mol Syst Biol, 19*(12), e11566. https://doi.org/10.15252/msb.202311566

Besada-Lombana, P. B., McTaggart, T. L., & Da Silva, N. A. (2018, 2018/10/01/). Molecular tools for pathway engineering in *Saccharomyces cerevisiae*. *Current Opinion in Biotechnology, 53*, 39-49. https://doi.org/https://doi.org/10.1016/j.copbio.2017.12.002

Bi, X., Cheng, Y., Xu, X., Lv, X., Liu, Y., Li, J., Du, G., Chen, J., Ledesma-Amaro, R., & Liu, L. (2023). etiBsu1209: A comprehensive multiscale metabolic model for *Bacillus subtilis*. *Biotechnology and Bioengineering, 120*(6), 1623-1639. https://doi.org/https://doi.org/10.1002/bit.28355

Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y.-G., & Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Molecular and cellular biology, 21*(1), 289-297.

Blaisdell, B. E., Campbell, A. M., & Karlin, S. (1996, Jun 11). Similarities and dissimilarities of phage genomes. *Proc Natl Acad Sci U S A, 93*(12), 5854-5859. https://doi.org/10.1073/pnas.93.12.5854

Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., & Shao, Y. (1997, Sep 5). The complete genome sequence of *Escherichia coli* K-12. *Science, 277*(5331), 1453-1462. https://doi.org/10.1126/science.277.5331.1453

Blount, Z. D. (2015). The unexhausted potential of *E. coli*. *eLife, 4*, e05826. https://doi.org/10.7554/eLife.05826

Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology, 151*(8), 2551-2561.

Borriss, R., Danchin, A., Harwood, C. R., Médigue, C., Rocha, E. P. C., Sekowska, A., & Vallenet, D. (2018, Jan). *Bacillus subtilis*, the model Gram-positive bacterium: 20 years of annotation refinement. *Microb Biotechnol, 11*(1), 3-17. https://doi.org/10.1111/1751-7915.13043

Bourgade, B., Humphreys, C. M., Millard, J., Minton, N. P., & Islam, M. A. (2022, 2022/05/20). Design, Analysis, and Implementation of a Novel Biochemical Pathway for Ethylene Glycol Production in Clostridium autoethanogenum. *ACS Synthetic Biology, 11*(5), 1790-1800. https://doi.org/10.1021/acssynbio.1c00624

Brödel, A. K., Charpenay, L. H., Galtier, M., Fuche, F. J., Terrasse, R., Poquet, C., Havránek, J., Pignotti, S., Krawczyk, A., Arraou, M., Prevot, G., Spadoni, D., Yarnall, M. T. N., Hessel, E. M., Fernandez-Rodriguez, J., Duportet, X., & Bikard, D. (2024, 2024/08/01). In situ targeted base editing of bacteria in the mouse gut. *Nature, 632*(8026), 877-884. https://doi.org/10.1038/s41586-024-07681-w

Browne, H. P., Forster, S. C., Anonye, B. O., Kumar, N., Neville, B. A., Stares, M. D., Goulding, D., & Lawley, T. D. (2016, May 26). Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature, 533*(7604), 543-546. https://doi.org/10.1038/nature17645

Bruant, G., Lévesque, M.-J., Peter, C., Guiot, S. R., & Masson, L. (2010). Genomic Analysis of Carbon Monoxide Utilization and Butanol Production by Clostridium carboxidivorans Strain P7T. *PLOS ONE, 5*(9), e13033. https://doi.org/10.1371/journal.pone.0013033

Bruder, M. R., Pyne, M. E., Moo-Young, M., Chung, D. A., & Chou, C. P. (2016). Extending CRISPR-Cas9 technology from genome editing to transcriptional engineering in the genus *Clostridium*. *Applied and Environmental Microbiology, 82*(20), 6109-6119.

Buchholz, K., & Collins, J. (2013, 2013/05/01). The roots—a short history of industrial microbiology and biotechnology. *Applied Microbiology and Biotechnology, 97*(9), 3747-3762. https://doi.org/10.1007/s00253-013-4768-2

Cañadas, I. C., Groothuis, D., Zygouropoulou, M., Rodrigues, R., & Minton, N. P. (2019, 2019/06/21). RiboCas: A Universal CRISPR-Based Editing Tool for *Clostridium*. *ACS Synthetic Biology, 8*(6), 1379-1390. https://doi.org/10.1021/acssynbio.9b00075

Cerveny, L., Straskova, A., Dankova, V., Hartlova, A., Ceckova, M., Staud, F., & Stulik, J. (2013, Mar). Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. *Infect Immun, 81*(3), 629-635. https://doi.org/10.1128/iai.01035-12

Chen, G.-Q., & Wu, Q. (2005, 2005/11/01/). The application of polyhydroxyalkanoates as tissue engineering materials. *Biomaterials, 26*(33), 6565-6578. https://doi.org/https://doi.org/10.1016/j.biomaterials.2005.04.036

Chen, Y., Caruso, L., McClane, B., Fisher, D., & Gupta, P. (2007, 2007/09/01/). Disruption of a toxin gene by introduction of a foreign gene into the chromosome of *Clostridium perfringens* using targetron-induced mutagenesis. *Plasmid, 58*(2), 182-189. https://doi.org/https://doi.org/10.1016/j.plasmid.2007.04.002

Cheng, C., Li, W., Lin, M., & Yang, S. T. (2019, Jul). Metabolic engineering of Clostridium carboxidivorans for enhanced ethanol and butanol production from syngas and glucose. *Bioresour Technol, 284*, 415-423. https://doi.org/10.1016/j.biortech.2019.03.145

Clark, T. A., Murray, I. A., Morgan, R. D., Kislyuk, A. O., Spittle, K. E., Boitano, M., Fomenkov, A., Roberts, R. J., & Korlach, J. (2012, Feb). Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res, 40*(4), e29. https://doi.org/10.1093/nar/gkr1146

Cohen, S. N., Chang, A. C., Boyer, H. W., & Helling, R. B. (1973, Nov). Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci U S A, 70*(11), 3240-3244. https://doi.org/10.1073/pnas.70.11.3240

Cohen, S. N., Chang, A. C. Y., & Hsu, L. (1972). Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of *Escherichia coli* by R-Factor DNA. *Proceedings of the National Academy of Sciences, 69*(8), 2110-2114. https://doi.org/doi:10.1073/pnas.69.8.2110

Consortium, T. U. (2020). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research, 49*(D1), D480-D489. https://doi.org/10.1093/nar/gkaa1100

Cooksley, C. M., Zhang, Y., Wang, H., Redl, S., Winzer, K., & Minton, N. P. (2012, 2012/11/01/). Targeted mutagenesis of the *Clostridium acetobutylicum* acetone–butanol–ethanol fermentation pathway. *Metabolic Engineering, 14*(6), 630-641. https://doi.org/https://doi.org/10.1016/j.ymben.2012.09.001

Cui, G.-z., Hong, W., Zhang, J., Li, W.-l., Feng, Y., Liu, Y.-j., & Cui, Q. (2012, 2012/06/01/). Targeted gene engineering in *Clostridium cellulolyticum* H10 without methylation. *Journal of Microbiological Methods, 89*(3), 201-208. https://doi.org/https://doi.org/10.1016/j.mimet.2012.02.015

Cui, G.-Z., Zhang, J., Hong, W., Xu, C., Feng, Y., Cui, Q., & Liu, Y.-J. (2014, 2014/01/01). Improvement of ClosTron for successive gene disruption in *Clostridium cellulolyticum* using a *pyrF*-based screening system. *Applied Microbiology and Biotechnology, 98*(1), 313-323. https://doi.org/10.1007/s00253-013-5330-y

Datta, J., & Kasprzyk, P. (2018). Thermoplastic polyurethanes derived from petrochemical or renewable resources: A comprehensive review. *Polymer Engineering & Science, 58*(S1), E14-E35. https://doi.org/https://doi.org/10.1002/pen.24633

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011, 2011/03/01). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature, 471*(7340), 602-607. https://doi.org/10.1038/nature09886

Di Leonardo, P. F., Antonicelli, G., Agostino, V., & Re, A. (2022). Genome-scale mining of acetogens of the genus *Clostridium* unveils distinctive traits in

[FeFe]-and [NiFe]-hydrogenase content and maturation. *Microbiology Spectrum, 10*(4), e01019-01022.

Dineen, S. S., McBride, S. M., & Sonenshein, A. L. (2010). Integration of Metabolism and Virulence by *Clostridium difficile* CodY. *Journal of bacteriology, 192*(20), 5350-5362. https://doi.org/doi:10.1128/jb.00341-10

Dineen, S. S., Villapakkam, A. C., Nordman, J. T., & Sonenshein, A. L. (2007, Oct). Repression of *Clostridium difficile* toxin gene expression by CodY. *Mol Microbiol, 66*(1), 206-219. https://doi.org/10.1111/j.1365-2958.2007.05906.x

Djapovic, M., Milivojevic, D., Ilic-Tomic, T., Lješević, M., Nikolaivits, E., Topakas, E., Maslak, V., & Nikodinovic-Runic, J. (2021, 2021/07/01/). Synthesis and characterization of polyethylene terephthalate (PET) precursors and potential degradation products: Toxicity study and application in discovery of novel PETases. *Chemosphere, 275*, 130005. https://doi.org/https://doi.org/10.1016/j.chemosphere.2021.130005

Doll, K., Rückel, A., Kämpf, P., Wende, M., & Weuster-Botz, D. (2018, 2018/10/01). Two stirred-tank bioreactors in series enable continuous production of alcohols from carbon monoxide with *Clostridium carboxidivorans*. *Bioprocess and Biosystems Engineering, 41*(10), 1403-1416. https://doi.org/10.1007/s00449-018-1969-1

Dragosits, M., & Mattanovich, D. (2013, 2013/07/01). Adaptive laboratory evolution – principles and applications for biotechnology. *Microbial Cell Factories, 12*(1), 64. https://doi.org/10.1186/1475-2859-12-64

Dryden, D. T., Cooper, L. P., & Murray, N. E. (1993, Jun 25). Purification and characterization of the methyltransferase from the type 1 restriction and modification system of *Escherichia coli* K12. *J Biol Chem, 268*(18), 13228-13236.

Dryden, D. T., Cooper, L. P., Thorpe, P. H., & Byron, O. (1997, Feb 4). The in vitro assembly of the EcoKI type I DNA restriction/modification enzyme and its in vivo implications. *Biochemistry, 36*(5), 1065-1076. https://doi.org/10.1021/bi9619435

Farrow, K. A., Lyras, D., & Rood, J. I. (2001). Genomic analysis of the erythromycin resistance element Tn5398 from *Clostridium difficile*.

Microbiology, 147(10), 2717-2728. https://doi.org/https://doi.org/10.1099/00221287-147-10-2717

Fernández-Naveira, Á., Abubackar, H. N., Veiga, M. C., & Kennes, C. (2016, 2016/04/01). Efficient butanol-ethanol (B-E) production from carbon monoxide fermentation by *Clostridium carboxidivorans*. *Applied Microbiology and Biotechnology, 100*(7), 3361-3370. https://doi.org/10.1007/s00253-015-7238-1

Fernández-Naveira, Á., Veiga, M. C., & Kennes, C. (2017). H-B-E (hexanol-butanol-ethanol) fermentation for the production of higher alcohols from syngas/waste gas. *Journal of Chemical Technology & Biotechnology, 92*(4), 712-731. https://doi.org/https://doi.org/10.1002/jctb.5194

Fernández-Naveira, Á., Veiga, M. C., & Kennes, C. (2019, 2019/07/17). Effect of salinity on C1-gas fermentation by *Clostridium carboxidivorans* producing acids and alcohols. *AMB Express, 9*(1), 110. https://doi.org/10.1186/s13568-019-0837-y

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., Korlach, J., & Turner, S. W. (2010, Jun). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods, 7*(6), 461-465. https://doi.org/10.1038/nmeth.1459

Foo, J. L., Jensen, H. M., Dahl, R. H., George, K., Keasling, J. D., Lee, T. S., Leong, S., & Mukhopadhyay, A. (2014). Improving Microbial Biogasoline Production in *Escherichia coli* Using Tolerance Engineering. *mBio, 5*(6), 10.1128/mbio.01932-01914. https://doi.org/doi:10.1128/mbio.01932-14

Förster, J., Famili, I., Fu, P., Palsson, B. Ø., & Nielsen, J. (2003). Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome research, 13*(2), 244-253.

Frost, L. S., Ippen-Ihler, K., & Skurray, R. A. (1994, Jun). Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol Rev, 58*(2), 162-210. https://doi.org/10.1128/mr.58.2.162-210.1994

Gago-Córdoba, C., Val-Calvo, J., Miguel-Arribas, A., Serrano, E., Singh, P. K., Abia, D., Wu, L. J., & Meijer, W. J. J. (2019). Surface Exclusion Revisited: Function Related to Differential Expression of the Surface Exclusion

System of *Bacillus subtilis* Plasmid pLS20. *Front Microbiol, 10*, 1502. https://doi.org/10.3389/fmicb.2019.01502

Gao, H., Qiu, Z., Wang, X., Zhang, X., Zhang, Y., Dai, J., & Liang, Z. (2024, 2024/03/01/). Recent advances in genome-scale engineering in *Escherichia coli* and their applications. *Engineering Microbiology, 4*(1), 100115. https://doi.org/https://doi.org/10.1016/j.engmic.2023.100115

Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H., & Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature, 468*(7320), 67-71.

Garritano, A. N., Song, W., & Thomas, T. (2022). Carbon fixation pathways across the bacterial and archaeal tree of life. *PNAS Nexus, 1*(5). https://doi.org/10.1093/pnasnexus/pgac226

Gassler, T., Sauer, M., Gasser, B., Egermeier, M., Troyer, C., Causon, T., Hann, S., Mattanovich, D., & Steiger, M. G. (2020, Feb). The industrial yeast *Pichia pastoris* is converted from a heterotroph into an autotroph capable of growth on CO(2). *Nat Biotechnol, 38*(2), 210-216. https://doi.org/10.1038/s41587-019-0363-0

Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020, 2020/01/01/). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal, 18*, 9-19. https://doi.org/https://doi.org/10.1016/j.csbj.2019.11.002

Gleizer, S., Ben-Nissan, R., Bar-On, Y. M., Antonovsky, N., Noor, E., Zohar, Y., Jona, G., Krieger, E., Shamshoum, M., Bar-Even, A., & Milo, R. (2019, Nov 27). Conversion of *Escherichia coli* to Generate All Biomass Carbon from CO(2). *Cell, 179*(6), 1255-1263.e1212. https://doi.org/10.1016/j.cell.2019.11.009

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., & Oliver, S. G. (1996). Life with 6000 Genes. *Science, 274*(5287), 546-567. https://doi.org/doi:10.1126/science.274.5287.546

Graupner, S., Weger, N., Sohni, M., & Wackernagel, W. (2001). Requirement of novel competence genes *pilT* and *pilU* of *Pseudomonas stutzeri* for natural transformation and suppression of *pilT* deficiency by a hexahistidine tag on the type IV pilus protein PilAI. *Journal of bacteriology, 183*(16), 4694-4701.

Grosse-Honebrink, A., Schwarz, K. M., Wang, H., Minton, N. P., & Zhang, Y. (2017, 2017/12/01/). Improving gene transfer in *Clostridium pasteurianum* through the isolation of rare hypertransformable variants. *Anaerobe, 48*, 203-205. https://doi.org/https://doi.org/10.1016/j.anaerobe.2017.09.001

Hadi, S. M., Bächi, B., Iida, S., & Bickle, T. A. (1983, Mar 25). DNA restriction--modification enzymes of phage P1 and plasmid p15B. Subunit functions and structural homologies. *J Mol Biol, 165*(1), 19-34. https://doi.org/10.1016/s0022-2836(83)80240-x

Han, Y.-F., Xie, B.-T., Wu, G.-x., Guo, Y.-Q., Li, D.-M., & Huang, Z.-Y. (2020, 2020-September-25). Combination of Trace Metal to Improve Solventogenesis of *Clostridium carboxidivorans* P7 in Syngas Fermentation [Original Research]. *Frontiers in Microbiology, 11*. https://doi.org/10.3389/fmicb.2020.577266

Hattman, S. (1980, Apr). Specificity of the bacteriophage Mu *mom*+ -controlled DNA modification. *J Virol, 34*(1), 277-279. https://doi.org/10.1128/jvi.34.1.277-279.1980

Heap, J. T., Kuehne, S. A., Ehsaan, M., Cartman, S. T., Cooksley, C. M., Scott, J. C., & Minton, N. P. (2010, 2010/01/01/). The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. *Journal of Microbiological Methods, 80*(1), 49-55. https://doi.org/https://doi.org/10.1016/j.mimet.2009.10.018

Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P., & Minton, N. P. (2007, 2007/09/01/). The ClosTron: A universal gene knock-out system for the genus *Clostridium*. *Journal of Microbiological Methods, 70*(3), 452-464. https://doi.org/https://doi.org/10.1016/j.mimet.2007.05.021

Heap, J. T., Pennington, O. J., Cartman, S. T., & Minton, N. P. (2009, 2009/07/01/). A modular system for *Clostridium* shuttle plasmids. *Journal of Microbiological Methods, 78*(1), 79-85. https://doi.org/https://doi.org/10.1016/j.mimet.2009.05.004

Hendriksen, W. T., Bootsma, H. J., Estevão, S., Hoogenboezem, T., de Jong, A., de Groot, R., Kuipers, O. P., & Hermans, P. W. (2008, Jan). CodY of *Streptococcus pneumoniae*: link between nutritional gene regulation and colonization. *J Bacteriol, 190*(2), 590-601. https://doi.org/10.1128/jb.00917-07

Henstra, A. M., Sipma, J., Rinzema, A., & Stams, A. J. M. (2007, 2007/06/01/). Microbiology of synthesis gas fermentation for biofuel production. *Current Opinion in Biotechnology, 18*(3), 200-206. https://doi.org/https://doi.org/10.1016/j.copbio.2007.03.008

Hoet, P. P., Coene, M. M., & Cocito, C. G. (1992). Replication cycle of *Bacillus subtilis* hydroxymethyluracil-containing phages. *Annu Rev Microbiol, 46*, 95-116. https://doi.org/10.1146/annurev.mi.46.100192.000523

Hoseini, S. S., & Sauer, M. G. (2015, 2015/01/19). Molecular cloning using polymerase chain reaction, an educational guide for cellular engineering. *Journal of Biological Engineering, 9*(1), 2. https://doi.org/10.1186/1754-1611-9-2

Hsu, B. B., Plant, I. N., Lyon, L., Anastassacos, F. M., Way, J. C., & Silver, P. A. (2020, 2020/10/06). In situ reprogramming of gut bacteria by oral delivery. *Nature Communications, 11*(1), 5030. https://doi.org/10.1038/s41467-020-18614-2

Hsu, Patrick D., Lander, Eric S., & Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell, 157*(6), 1262-1278. https://doi.org/10.1016/j.cell.2014.05.010

Hsueh, Y. H., Somers, E. B., & Wong, A. C. (2008, Jun). Characterization of the *codY* gene and its influence on biofilm formation in *Bacillus cereus*. *Arch Microbiol, 189*(6), 557-568. https://doi.org/10.1007/s00203-008-0348-8

Hu, J., Yu, F., & Lu, Y. (2012). Application of Fischer–Tropsch Synthesis in Biomass to Liquid Conversion. *Catalysts, 2*(2), 303-326. https://www.mdpi.com/2073-4344/2/2/303

Huang, H., Chai, C., Li, N., Rowe, P., Minton, N. P., Yang, S., Jiang, W., & Gu, Y. (2016). CRISPR/Cas9-based efficient genome editing in *Clostridium ljungdahlii*, an autotrophic gas-fermenting bacterium. *ACS Synthetic Biology, 5*(12), 1355-1361.

Ichiyanagi, K., Beauregard, A., Lawrence, S., Smith, D., Cousineau, B., & Belfort, M. (2002). Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Molecular Microbiology, 46*(5), 1259-1272. https://doi.org/https://doi.org/10.1046/j.1365-2958.2002.03226.x

Iida, S., Streiff, M. B., Bickle, T. A., & Arber, W. (1987, Mar). Two DNA antirestriction systems of bacteriophage P1, *darA*, and *darB*: characterization of darA- phages. *Virology, 157*(1), 156-166. https://doi.org/10.1016/0042-6822(87)90324-2

Ingle, A. P., Saxena, S., Moharil, M. P., Rivaldi, J. D., Ramos, L., & Chandel, A. K. (2025, 2025/03/10). Production of biomaterials and biochemicals from lignocellulosic biomass through sustainable approaches: current scenario and future perspectives. *Biotechnology for Sustainable Materials, 2*(1), 3. https://doi.org/10.1186/s44316-025-00025-2

Janscak, P., Dryden, D. T., & Firman, K. (1998, Oct 1). Analysis of the subunit assembly of the typeIC restriction-modification enzyme EcoR124I. *Nucleic Acids Res, 26*(19), 4439-4445. https://doi.org/10.1093/nar/26.19.4439

Jeyachandran, S., Vibhute, P., Kumar, D., & Ragavendran, C. (2023, 2023/12/15). Random mutagenesis as a tool for industrial strain improvement for enhanced production of antibiotics: a review. *Molecular Biology Reports, 51*(1), 19. https://doi.org/10.1007/s11033-023-08975-4

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology, 31*(3), 233-239.

Jin, H., Chen, L., Wang, J., & Zhang, W. (2014, 2014/03/01/). Engineering biofuel tolerance in non-native producing microorganisms. *Biotechnology Advances, 32*(2), 541-548. https://doi.org/https://doi.org/10.1016/j.biotechadv.2014.02.001

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science, 337*(6096), 816-821.

Johnston, C. D., Cotton, S. L., Rittling, S. R., Starr, J. R., Borisy, G. G., Dewhirst, F. E., & Lemon, K. P. (2019, Jun 4). Systematic evasion of the restriction-

modification barrier in bacteria. *Proc Natl Acad Sci U S A, 116*(23), 11454-11459. https://doi.org/10.1073/pnas.1820256116

Jones, D. T., & Keis, S. (1995). Origins and relationships of industrial solvent-producing *clostridial* strains. *FEMS Microbiology Reviews, 17*(3), 223-232. https://doi.org/10.1111/j.1574-6976.1995.tb00206.x

Jones, D. T., Schulz, F., Roux, S., & Brown, S. D. (2023, Sep 7). Solvent-Producing *Clostridia* Revisited. *Microorganisms, 11*(9). https://doi.org/10.3390/microorganisms11092253

Karyagina, A., Shilov, I., Tashlitskii, V., Khodoun, M., Vasil'ev, S., Lau, P. C., & Nikolskaya, I. (1997). Specific binding of sso II DNA methyltransferase to its promoter region provides the regulation of sso II restriction-modification gene expression. *Nucleic Acids Research, 25*(11), 2114-2120.

Kaunitz, J. D. (2015, Aug). The Discovery of PCR: ProCuRement of Divine Power. *Dig Dis Sci, 60*(8), 2230-2231. https://doi.org/10.1007/s10620-015-3747-0

King, G., & Murray, N. E. (1995). Restriction alleviation and modification enhancement by the Rac prophage of *Escherichia coli* K-12. *Molecular Microbiology, 16*(4), 769-777. https://doi.org/https://doi.org/10.1111/j.1365-2958.1995.tb02438.x

Koller, M., Hesse, P., Bona, R., Kutschera, C., Atlić, A., & Braunegg, G. (2007). Biosynthesis of high quality polyhydroxyalkanoate co- and terpolyesters for potential medical application by the archaeon *Haloferax mediterranei*. *Macromolecular Symposia, 253*(1), 33-39. https://doi.org/https://doi.org/10.1002/masy.200750704

Köpke, M., Mihalcea, C., Liew, F., Tizard, J. H., Ali, M. S., Conolly, J. J., Al-Sinawi, B., & Simpson, S. D. (2011, Aug). 2,3-butanediol production by acetogenic bacteria, an alternative route to chemical synthesis, using industrial waste gas. *Appl Environ Microbiol, 77*(15), 5467-5475. https://doi.org/10.1128/aem.00355-11

Köpke, M., & Simpson, S. D. (2020, 2020/10/01/). Pollution to products: recycling of 'above ground' carbon by gas fermentation. *Current Opinion in Biotechnology, 65*, 180-189. https://doi.org/https://doi.org/10.1016/j.copbio.2020.02.017

Kottenhahn, P., Philipps, G., Bunk, B., Spröer, C., & Jennewein, S. (2023, Dec 12). The Restriction-Modification Systems of *Clostridium carboxidivorans* P7. *Microorganisms, 11*(12). https://doi.org/10.3390/microorganisms11122962

Kottenhahn, P., Philipps, G., & Jennewein, S. (2021). Hexanol biosynthesis from syngas by *Clostridium carboxidivorans* P7–product toxicity, temperature dependence and in situ extraction. *Heliyon, 7*(8).

Krüger, D., & Bickle, T. A. (1983). Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiological reviews, 47*(3), 345-360.

Kuehne, S. A., & Minton, N. P. (2012, 2012/07/18). ClosTron-mediated engineering of *Clostridium*. *Bioengineered, 3*(4), 247-254. https://doi.org/10.4161/bioe.21004

Kumar, B., Bhardwaj, N., Agrawal, K., Chaturvedi, V., & Verma, P. (2020, 2020/03/01/). Current perspective on pretreatment technologies using lignocellulosic biomass: An emerging biorefinery concept. *Fuel Processing Technology, 199*, 106244. https://doi.org/https://doi.org/10.1016/j.fuproc.2019.106244

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessières, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., Cummings, N. J., Daniel, R. A., Denizot, F., Devine, K. M., Düsterhöft, A., Ehrlich, S. D., Emmerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S. Y., Glaser, P., Goffeau, A., Golightly, E. J., Grandi, G., Guiseppi, G., Guy, B. J., Haga, K., Haiech, J., Harwood, C. R., Hénaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M. F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S. M., Levine, A., Liu, H., Masuda, S., Mauël, C., Médigue, C., Medina, N., Mellado, R. P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S. H., Parro, V., Pohl, T. M., Portetelle, D., Porwollik, S., Prescott, A. M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,

Sekowska, A., Seror, S. J., Serror, P., Shin, B. S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H. F., Zumstein, E., Yoshikawa, H., & Danchin, A. (1997, 1997/11/01). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature, 390*(6657), 249-256. https://doi.org/10.1038/36786

Kuwana, R., Dupuy, B., Martin-Verstraete, I., & Takamatsu, H. (2024, 2024-April-24). SpoIVA is an essential morphogenetic protein for the formation of heat- and lysozyme-resistant spores in *Clostridium sporogenes* NBRC 14293 [Original Research]. *Frontiers in Microbiology, 15*. https://doi.org/10.3389/fmicb.2024.1338751

Lakhssassi, N., Baharlouei, A., Meksem, J., Hamilton-Brehm, S. D., Lightfoot, D. A., Meksem, K., & Liang, Y. (2020). EMS-Induced Mutagenesis of *Clostridium carboxidivorans* for Increased Atmospheric CO2 Reduction Efficiency and Solvent Production. *Microorganisms, 8*(8), 1239. https://www.mdpi.com/2076-2607/8/8/1239

Lam, K. N., Spanogiannopoulos, P., Soto-Perez, P., Alexander, M., Nalley, M. J., Bisanz, J. E., Nayak, R. R., Weakley, A. M., Feiqiao, B. Y., & Turnbaugh, P. J. (2021). Phage-delivered CRISPR-Cas9 for strain-specific depletion and genomic deletions in the gut microbiome. *Cell Reports, 37*(5).

Lemos, J. A., Nascimento, M. M., Lin, V. K., Abranches, J., & Burne, R. A. (2008, Aug). Global regulation by (p)ppGpp and CodY in *Streptococcus mutans*. *J Bacteriol, 190*(15), 5291-5299. https://doi.org/10.1128/jb.00288-08

Lennen, R. M., & Pfleger, B. F. (2013). Modulating Membrane Composition Alters Free Fatty Acid Tolerance in *Escherichia coli*. *PLOS ONE, 8*(1), e54031. https://doi.org/10.1371/journal.pone.0054031

Liew, F., Henstra, A. M., Köpke, M., Winzer, K., Simpson, S. D., & Minton, N. P. (2017, Mar). Metabolic engineering of *Clostridium autoethanogenum* for selective alcohol production. *Metab Eng, 40*, 104-114. https://doi.org/10.1016/j.ymben.2017.01.007

Liew, F., Henstra, A. M., Winzer, K., Köpke, M., Simpson, S. D., & Minton, N. P. (2016). Insights into CO2 Fixation Pathway of *Clostridium*

autoethanogenum by Targeted Mutagenesis. *mBio, 7*(3), 10.1128/mbio.00427-00416. https://doi.org/doi:10.1128/mbio.00427-16

Liew, F., Martin, M. E., Tappel, R. C., Heijstra, B. D., Mihalcea, C., & Köpke, M. (2016, 2016-May-11). Gas Fermentation—A Flexible Platform for Commercial Scale Production of Low-Carbon-Fuels and Chemicals from Waste and Renewable Feedstocks [Review]. *Frontiers in Microbiology, 7*. https://doi.org/10.3389/fmicb.2016.00694

Liew, F. E., Nogle, R., Abdalla, T., Rasor, B. J., Canter, C., Jensen, R. O., Wang, L., Strutz, J., Chirania, P., De Tissera, S., Mueller, A. P., Ruan, Z., Gao, A., Tran, L., Engle, N. L., Bromley, J. C., Daniell, J., Conrado, R., Tschaplinski, T. J., Giannone, R. J., Hettich, R. L., Karim, A. S., Simpson, S. D., Brown, S. D., Leang, C., Jewett, M. C., & Köpke, M. (2022, 2022/03/01). Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. *Nature Biotechnology, 40*(3), 335-344. https://doi.org/10.1038/s41587-021-01195-w

Liou, J. S.-C., Balkwill, D. L., Drake, G. R., & Tanner, R. S. (2005). *Clostridium carboxidivorans* sp. nov., a solvent-producing *Clostridium* isolated from an agricultural settling lagoon, and reclassification of the acetogen *Clostridium scatologenes* strain SL1 as *Clostridium drakei* sp. nov. *International Journal of Systematic and Evolutionary Microbiology, 55*(5), 2085-2091. https://doi.org/https://doi.org/10.1099/ijs.0.63482-0

Loenen, W. A. M., & Murray, N. E. (1986, 1986/07/05/). Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage λ. *Journal of Molecular Biology, 190*(1), 11-22. https://doi.org/https://doi.org/10.1016/0022-2836(86)90071-9

Louca, S., Mazel, F., Doebeli, M., & Parfrey, L. W. (2019). A census-based estimate of Earth's bacterial and archaeal diversity. *PLOS Biology, 17*(2), e3000106. https://doi.org/10.1371/journal.pbio.3000106

Majerczyk, C. D., Sadykov, M. R., Luong, T. T., Lee, C., Somerville, G. A., & Sonenshein, A. L. (2008, Apr). *Staphylococcus aureus* CodY negatively regulates virulence gene expression. *J Bacteriol, 190*(7), 2257-2265. https://doi.org/10.1128/jb.01545-07

Malke, H., & Ferretti, J. J. (2007, Jun). CodY-affected transcriptional gene expression of *Streptococcus pyogenes* during growth in human blood. *J*

Med Microbiol, 56(Pt 6), 707-714. https://doi.org/10.1099/jmm.0.46984-0

McCarthy, S., Ai, C., & Blum, P. (2018). Chapter Four - Enhancement of *Metallosphaera sedula* Bioleaching by Targeted Recombination and Adaptive Laboratory Evolution. In G. M. Gadd & S. Sariaslani (Eds.), *Advances in Applied Microbiology* (Vol. 104, pp. 135-165). Academic Press. https://doi.org/https://doi.org/10.1016/bs.aambs.2018.03.002

Mermelstein, L. D., & Papoutsakis, E. T. (1993, Apr). In vivo methylation in *Escherichia coli* by the *Bacillus subtilis* phage phi 3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. *Appl Environ Microbiol, 59*(4), 1077-1081. https://doi.org/10.1128/aem.59.4.1077-1081.1993

Merz, A. J., So, M., & Sheetz, M. P. (2000, 2000/09/01). Pilus retraction powers bacterial twitching motility. *Nature, 407*(6800), 98-102. https://doi.org/10.1038/35024105

Mohr, G., Smith, D., Belfort, M., & Lambowitz, A. M. (2000). Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes & development, 14*(5), 559-573.

Mojica, F. J., Díez-Villaseñor, C. s., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution, 60*, 174-182.

Molino, A., Chianese, S., & Musmarra, D. (2016, 2016/01/01/). Biomass gasification technology: The state of the art overview. *Journal of Energy Chemistry, 25*(1), 10-25. https://doi.org/https://doi.org/10.1016/j.jechem.2015.11.005

Muh, U., Pannullo, A., Weiss, D., & Ellermeier, C. (2019). A xylose-inducible expression system and a CRISPRi-plasmid for targeted knock-down of gene expression in *Clostridioides difficile*. *J Bacteriol, 201*(10.1128), 00711-00718.

Muringayil Joseph, T., Azat, S., Ahmadi, Z., Moini Jazani, O., Esmaeili, A., Kianfar, E., Haponiuk, J., & Thomas, S. (2024, 2024/06/01/). Polyethylene terephthalate (PET) recycling: A review. *Case Studies in Chemical and*

Environmental Engineering, 9, 100673. https://doi.org/https://doi.org/10.1016/j.cscee.2024.100673

Murray, N. E. (2000, Jun). Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev, 64*(2), 412-434. https://doi.org/10.1128/mmbr.64.2.412-434.2000

Nagaraju, S., Davies, N. K., Walker, D. J. F., Köpke, M., & Simpson, S. D. (2016). Genome editing of *Clostridium autoethanogenum* using CRISPR/Cas9. *Biotechnology for Biofuels, 9*, 1-8.

Nagaraju, S., & Koepke, M. (2019). *Microorganism with modified hydrogenase activity* (USA Patent No. 10,358,661). https://patents.google.com/patent/US10358661B2/en

Ndaba, B., Chiyanzu, I., & Marx, S. (2015, Dec). n-Butanol derived from biochemical and chemical routes: A review. *Biotechnol Rep (Amst), 8*, 1-9. https://doi.org/10.1016/j.btre.2015.08.001

Nethery, M. A., Hidalgo-Cantabrana, C., Roberts, A., & Barrangou, R. (2022). CRISPR-based engineering of phages for in situ bacterial base editing. *Proceedings of the National Academy of Sciences, 119*(46), e2206744119.

Ng, Y. K., Ehsaan, M., Philip, S., Collery, M. M., Janoir, C., Collignon, A., Cartman, S. T., & 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), e56051. https://doi.org/10.1371/journal.pone.0056051

Nicolaou, S. A., Gaida, S. M., & Papoutsakis, E. T. (2010, 2010/07/01/). A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation. *Metabolic Engineering, 12*(4), 307-331. https://doi.org/https://doi.org/10.1016/j.ymben.2010.03.004

Noyer-Weidner, M., Jentsch, S., Pawlek, B., Günthert, U., & Trautner, T. A. (1983, May). Restriction and modification in *Bacillus subtilis*: DNA methylation potential of the related bacteriophages Z, SPR, SP beta, phi 3T, and rho 11. *J Virol, 46*(2), 446-453. https://doi.org/10.1128/jvi.46.2.446-453.1983

Oliveira, P. H., Ribis, J. W., Garrett, E. M., Trzilova, D., Kim, A., Sekulovic, O., Mead, E. A., Pak, T., Zhu, S., Deikus, G., Touchon, M., Lewis-Sandari, M., Beckford, C., Zeitouni, N. E., Altman, D. R., Webster, E., Oussenko, I., Bunyavanich, S., Aggarwal, A. K., Bashir, A., Patel, G., Wallach, F., Hamula, C., Huprikar, S., Schadt, E. E., Sebra, R., van Bakel, H., Kasarskis, A., Tamayo, R., Shen, A., & Fang, G. (2020, 2020/01/01). Epigenomic characterization of *Clostridioides difficile* finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. *Nature Microbiology, 5*(1), 166-180. https://doi.org/10.1038/s41564-019-0613-4

Orsi, E., Nikel, P. I., Nielsen, L. K., & Donati, S. (2023, 2023/10/21). Synergistic investigation of natural and synthetic C1-trophic microorganisms to foster a circular carbon economy. *Nature Communications, 14*(1), 6673. https://doi.org/10.1038/s41467-023-42166-w

Oultram, J. D., Loughlin, M., Swinfield, T.-J., Brehm, J. K., Thompson, D. E., & Minton, N. P. (1988). Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. *FEMS Microbiology Letters, 56*(1), 83-88. https://doi.org/10.1111/j.1574-6968.1988.tb03154.x

Parapouli, M., Vasileiadis, A., Afendra, A. S., & Hatziloukas, E. (2020). *Saccharomyces cerevisiae* and its industrial applications. *AIMS Microbiol, 6*(1), 1-31. https://doi.org/10.3934/microbiol.2020001

Park, H., He, H., Yan, X., Liu, X., Scrutton, N. S., & Chen, G.-Q. (2024, 2024/03/01/). PHA is not just a bioplastic! *Biotechnology Advances, 71*, 108320. https://doi.org/https://doi.org/10.1016/j.biotechadv.2024.108320

Parte, A. C. (2018). LPSN – List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *International Journal of Systematic and Evolutionary Microbiology, 68*(6), 1825-1829. https://doi.org/https://doi.org/10.1099/ijsem.0.002786

Peplow, M. (2015, 2015/11/01). Industrial biotechs turn greenhouse gas into feedstock opportunity. *Nature Biotechnology, 33*(11), 1124-1125. https://doi.org/10.1038/nbt1115-1123

Perutka, J., Wang, W., Goerlitz, D., & Lambowitz, A. M. (2004, 2004/02/13/). Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *Journal of Molecular*

Biology, *336*(2), 421-439. https://doi.org/https://doi.org/10.1016/j.jmb.2003.12.009

Phillips, J. R., Atiyeh, H. K., Tanner, R. S., Torres, J. R., Saxena, J., Wilkins, M. R., & Huhnke, R. L. (2015, 2015/08/01/). Butanol and hexanol production in *Clostridium carboxidivorans* syngas fermentation: Medium development and culture techniques. *Bioresource Technology, 190*, 114-121. https://doi.org/https://doi.org/10.1016/j.biortech.2015.04.043

Poudel, S., Dunham, E. C., Lindsay, M. R., Amenabar, M. J., Fones, E. M., Colman, D. R., & Boyd, E. S. (2018, 2018-August-03). Origin and Evolution of Flavin-Based Electron Bifurcating Enzymes [Original Research]. *Frontiers in Microbiology, 9*. https://doi.org/10.3389/fmicb.2018.01762

Pyne, M. E., Bruder, M., Moo-Young, M., Chung, D. A., & Chou, C. P. (2014, 2014/05/01/). Technical guide for genetic advancement of underdeveloped and intractable *Clostridium*. *Biotechnology Advances, 32*(3), 623-641. https://doi.org/https://doi.org/10.1016/j.biotechadv.2014.04.003

Pyne, M. E., Moo-Young, M., Chung, D. A., & Chou, C. P. (2013, 2013/04/09). Development of an electrotransformation protocol for genetic manipulation of *Clostridium pasteurianum*. *Biotechnology for Biofuels, 6*(1), 50. https://doi.org/10.1186/1754-6834-6-50

Quinn, R. (2013, Mar). Rethinking antibiotic research and development: World War II and the penicillin collaborative. *Am J Public Health, 103*(3), 426-434. https://doi.org/10.2105/ajph.2012.300693

Ragsdale, S. W., & Pierce, E. (2008, 2008/12/01/). Acetogenesis and the Wood–Ljungdahl pathway of CO2 fixation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1784*(12), 1873-1898. https://doi.org/https://doi.org/10.1016/j.bbapap.2008.08.012

Ramió-Pujol, S., Ganigué, R., Bañeras, L., & Colprim, J. (2015, Sep). Incubation at 25 °C prevents acid crash and enhances alcohol production in *Clostridium carboxidivorans* P7. *Bioresour Technol, 192*, 296-303. https://doi.org/10.1016/j.biortech.2015.05.077

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013, 2013/11/01). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols, 8*(11), 2281-2308. https://doi.org/10.1038/nprot.2013.143

Rao, D. N., Dryden, D. T. F., & Bheemanaik, S. (2013). Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Research, 42*(1), 45-55. https://doi.org/10.1093/nar/gkt616

Rawlings, D. E. (2002). Heavy metal mining using microbes. *Annual Review of Microbiology, 56*(1), 65-91. https://doi.org/10.1146/annurev.micro.56.012302.161052

Redfern, E. (2021). *Developing Synthetic Biology Tools for Acetogens* [Ph. D. Thesis, University of Nottingham]. Nottingham. https://eprints.nottingham.ac.uk/id/eprint/69764

Rizzo, M. G., De Plano, L. M., & Franco, D. (2020, 2020/06/01). Regulation of filamentation by bacteria and its impact on the productivity of compounds in biotechnological processes. *Applied Microbiology and Biotechnology, 104*(11), 4631-4642. https://doi.org/10.1007/s00253-020-10590-3

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S., Dryden, D. T., Dybvig, K., Firman, K., Gromova, E. S., Gumport, R. I., Halford, S. E., Hattman, S., Heitman, J., Hornby, D. P., Janulaitis, A., Jeltsch, A., Josephsen, J., Kiss, A., Klaenhammer, T. R., Kobayashi, I., Kong, H., Krüger, D. H., Lacks, S., Marinus, M. G., Miyahara, M., Morgan, R. D., Murray, N. E., Nagaraja, V., Piekarowicz, A., Pingoud, A., Raleigh, E., Rao, D. N., Reich, N., Repin, V. E., Selker, E. U., Shaw, P. C., Stein, D. C., Stoddard, B. L., Szybalski, W., Trautner, T. A., Van Etten, J. L., Vitor, J. M., Wilson, G. G., & Xu, S. Y. (2003, Apr 1). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res, 31*(7), 1805-1812. https://doi.org/10.1093/nar/gkg274

Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2003, Jan 1). REBASE: restriction enzymes and methyltransferases. *Nucleic Acids Res, 31*(1), 418-420. https://doi.org/10.1093/nar/gkg069

Rückel, A., Hannemann, J., Maierhofer, C., Fuchs, A., & Weuster-Botz, D. (2021, 2021-April-15). Studies on Syngas Fermentation With *Clostridium carboxidivorans* in Stirred-Tank Reactors With Defined Gas Impurities [Original Research]. *Frontiers in Microbiology, 12*. https://doi.org/10.3389/fmicb.2021.655390

Saha, S., & Rao, D. N. (1995, Apr 7). ATP hydrolysis is required for DNA cleavage by EcoPI restriction enzyme. *J Mol Biol, 247*(4), 559-567. https://doi.org/10.1016/s0022-2836(05)80137-8

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988). Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science, 239*(4839), 487-491. https://doi.org/doi:10.1126/science.2448875

Saldanha, R. J., Pemberton, A., Shiflett, P., Perutka, J., Whitt, J. T., Ellington, A., Lambowitz, A. M., Kramer, R., Taylor, D., & Lamkin, T. J. (2013, Sep 18). Rapid targeted gene disruption in *Bacillus anthracis*. *BMC Biotechnol, 13*, 72. https://doi.org/10.1186/1472-6750-13-72

Salzer, R., Joos, F., & Averhoff, B. (2014, 2014/01/15). Type IV Pilus Biogenesis, Twitching Motility, and DNA Uptake in *Thermus thermophilus*: Discrete Roles of Antagonistic ATPases PilF, PilT1, and PilT2. *Applied and Environmental Microbiology, 80*(2), 644-652. https://doi.org/10.1128/AEM.03218-13

Santos Correa, S., Schultz, J., Lauersen, K. J., & Soares Rosado, A. (2023, May). Natural carbon fixation and advances in synthetic engineering for redesigning and creating new fixation pathways. *J Adv Res, 47*, 75-92. https://doi.org/10.1016/j.jare.2022.07.011

Schuchmann, K., & Müller, V. (2014, 2014/12/01). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nature Reviews Microbiology, 12*(12), 809-821. https://doi.org/10.1038/nrmicro3365

Scott, P. T., & Rood, J. I. (1989, 1989/10/30/). Electroporation-mediated transformation of lysostaphin-treated *Clostridium perfringens*. *Gene, 82*(2), 327-333. https://doi.org/https://doi.org/10.1016/0378-1119(89)90059-0

Seys, F. M., Rowe, P., Bolt, E. L., Humphreys, C. M., & Minton, N. P. (2020, Apr 23). A Gold Standard, CRISPR/Cas9-Based Complementation Strategy Reliant on 24 Nucleotide Bookmark Sequences. *Genes (Basel), 11*(4). https://doi.org/10.3390/genes11040458

Sharp, P. M. (1986, Jan). Molecular evolution of bacteriophages: evidence of selection against the recognition sites of host restriction enzymes. *Mol*

Biol Evol, 3(1), 75-83. https://doi.org/10.1093/oxfordjournals.molbev.a040377

Shen, S., Gu, Y., Chai, C., Jiang, W., Zhuang, Y., & Wang, Y. (2017, 2017/09/01/). Enhanced alcohol titre and ratio in carbon monoxide-rich off-gas fermentation of *Clostridium carboxidivorans* through combination of trace metals optimization with variable-temperature cultivation. *Bioresource Technology, 239*, 236-243. https://doi.org/https://doi.org/10.1016/j.biortech.2017.04.099

Shen, S., Wang, G., Zhang, M., Tang, Y., Gu, Y., Jiang, W., Wang, Y., & Zhuang, Y. (2020, 2020/10/22). Effect of temperature and surfactant on biomass growth and higher-alcohol production during syngas fermentation by *Clostridium carboxidivorans* P7. *Bioresources and Bioprocessing, 7*(1), 56. https://doi.org/10.1186/s40643-020-00344-4

Shen, Y., Brown, R., & Wen, Z. (2014, 2014/12/31/). Enhancing mass transfer and ethanol production in syngas fermentation of *Clostridium carboxidivorans* P7 through a monolithic biofilm reactor. *Applied Energy, 136*, 68-76. https://doi.org/https://doi.org/10.1016/j.apenergy.2014.08.117

Shen, Y., Brown, R. C., & Wen, Z. (2017, 2017/02/01/). Syngas fermentation by *Clostridium carboxidivorans* P7 in a horizontal rotating packed bed biofilm reactor with enhanced ethanol production. *Applied Energy, 187*, 585-594. https://doi.org/https://doi.org/10.1016/j.apenergy.2016.11.084

Som, S., & Friedman, S. (1994). Regulation of EcoRII methyltransferase: effect of mutations on gene expression and in vitro binding to the promoter region. *Nucleic Acids Research, 22*(24), 5347-5353.

Song, Y., Lee, J. S., Shin, J., Lee, G. M., Jin, S., Kang, S., Lee, J.-K., Kim, D. R., Lee, E. Y., Kim, S. C., Cho, S., Kim, D., & Cho, B.-K. (2020). Functional cooperation of the glycine synthase-reductase and Wood–Ljungdahl pathways for autotrophic growth of Clostridium drakei. *Proceedings of the National Academy of Sciences, 117*(13), 7516-7523. https://doi.org/doi:10.1073/pnas.1912289117

Steiner, E., Dago, A. E., Young, D. I., Heap, J. T., Minton, N. P., Hoch, J. A., & Young, M. (2011). Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in

Clostridium acetobutylicum. *Molecular Microbiology, 80*(3), 641-654. https://doi.org/https://doi.org/10.1111/j.1365-2958.2011.07608.x

Stewart, F. J., Panne, D., Bickle, T. A., & Raleigh, E. A. (2000, May 12). Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J Mol Biol, 298*(4), 611-622. https://doi.org/10.1006/jmbi.2000.3697

Swinton, D., Hattman, S., Crain, P. F., Cheng, C. S., Smith, D. L., & McCloskey, J. A. (1983). Purification and characterization of the unusual deoxynucleoside, alpha-N-(9-beta-D-2'-deoxyribofuranosylpurin-6-yl)glycinamide, specified by the phage Mu modification function. *Proceedings of the National Academy of Sciences, 80*(24), 7400-7404. https://doi.org/doi:10.1073/pnas.80.24.7400

Taylor, I., Patel, J., Firman, K., & Kneale, G. (1992). Purification and biochemical characterisation of the Eco R124 type I modification methylase. *Nucleic Acids Research, 20*(2), 179-186. https://doi.org/10.1093/nar/20.2.179

Thunuguntla, R., Atiyeh, H. K., Huhnke, R. L., & Tanner, R. S. (2024, 2024/02/01/). CO2-based production of C2-C6 acids and alcohols: The potential of novel *Clostridia*. *Bioresource Technology Reports, 25*, 101713. https://doi.org/https://doi.org/10.1016/j.biteb.2023.101713

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007, 2007/10/01). The Human Microbiome Project. *Nature, 449*(7164), 804-810. https://doi.org/10.1038/nature06244

van Schaik, W., Château, A., Dillies, M. A., Coppée, J. Y., Sonenshein, A. L., & Fouet, A. (2009, Oct). The global regulator CodY regulates toxin gene expression in *Bacillus anthracis* and is required for full virulence. *Infect Immun, 77*(10), 4437-4445. https://doi.org/10.1128/iai.00716-09

Vasileiadou, A. (2024). From Organic Wastes to Bioenergy, Biofuels, and Value-Added Products for Urban Sustainability and Circular Economy: A Review. *Urban Science, 8*(3), 121. https://www.mdpi.com/2413-8851/8/3/121

Vasu, K., & Nagaraja, V. (2013, Mar). Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev, 77*(1), 53-72. https://doi.org/10.1128/mmbr.00044-12

Vees, C. A., Herwig, C., & Pflügl, S. (2022, 2022/06/01/). Mixotrophic co-utilization of glucose and carbon monoxide boosts ethanol and butanol productivity of continuous *Clostridium carboxidivorans* cultures. *Bioresource Technology, 353*, 127138. https://doi.org/https://doi.org/10.1016/j.biortech.2022.127138

Vlot, M., Houkes, J., Lochs, S. J. A., Swarts, D. C., Zheng, P., Kunne, T., Mohanraju, P., Anders, C., Jinek, M., van der Oost, J., Dickman, M. J., & Brouns, S. J. J. (2018, Jan 25). Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR-Cas effector complexes. *Nucleic Acids Res, 46*(2), 873-885. https://doi.org/10.1093/nar/gkx1264

Walkinshaw, M. D., Taylor, P., Sturrock, S. S., Atanasiu, C., Berge, T., Henderson, R. M., Edwardson, J. M., & Dryden, D. T. F. (2002, 2002/01/01/). Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Molecular Cell, 9*(1), 187-194. https://doi.org/https://doi.org/10.1016/S1097-2765(02)00435-5

Wang, S., Dong, S., Wang, P., Tao, Y., & Wang, Y. (2017). Genome editing in *Clostridium saccharoperbutylacetonicum* N1-4 with the CRISPR-Cas9 system. *Applied and Environmental Microbiology, 83*(10), e00233-00217.

Wang, S., Hong, W., Dong, S., Zhang, Z. T., Zhang, J., Wang, L., & Wang, Y. (2018, 2018/10/01/). Genome engineering of *Clostridium difficile* using the CRISPR-Cas9 system. *Clinical Microbiology and Infection, 24*(10), 1095-1099. https://doi.org/https://doi.org/10.1016/j.cmi.2018.03.026

Wang, S., Huang, H., Kahnt, J., Mueller, A. P., Köpke, M., & Thauer, R. K. (2013). NADP-specific electron-bifurcating [FeFe]-hydrogenase in a functional complex with formate dehydrogenase in *Clostridium autoethanogenum* grown on CO. *Journal of bacteriology, 195*(19), 4373-4386.

Wang, Y., Zhang, Z.-T., Seo, S.-O., Choi, K., Lu, T., Jin, Y.-S., & Blaschek, H. P. (2015, 2015/04/20/). Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *Journal of Biotechnology, 200*, 1-5. https://doi.org/https://doi.org/10.1016/j.jbiotec.2015.02.005

Warren, R. A. J. (1980). Modified Bases in Bacteriophage DNAs. *Annual Review of Microbiology, 34*(Volume 34, 1980), 137-158. https://doi.org/https://doi.org/10.1146/annurev.mi.34.100180.001033

Wasels, F., Jean-Marie, J., Collas, F., López-Contreras, A. M., & Ferreira, N. L. (2017). A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. *Journal of Microbiological Methods, 140*, 5-11.

Weigele, P., & Raleigh, E. A. (2016, 2016/10/26). Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. *Chemical Reviews, 116*(20), 12655-12687. https://doi.org/10.1021/acs.chemrev.6b00114

Wen, Z., Minton, N. P., Zhang, Y., Li, Q., Liu, J., Jiang, Y., & Yang, S. (2017). Enhanced solvent production by metabolic engineering of a twin-*clostridial* consortium. *Metabolic Engineering, 39*, 38-48.

Wilson, G. G. (1991). Organization of restriction-modification systems. *Nucleic Acids Research, 19*(10), 2539-2566. https://doi.org/10.1093/nar/19.10.2539

Wirth, S., & Dürre, P. (2021, 2021/07/01/). Investigation of putative genes for the production of medium-chained acids and alcohols in autotrophic acetogenic bacteria. *Metabolic Engineering, 66*, 296-307. https://doi.org/https://doi.org/10.1016/j.ymben.2021.04.010

Wolfgang, M., Lauer, P., Park, H. S., Brossay, L., Hébert, J., & Koomey, M. (1998). PilT mutations lead to simultaneous defects in competence for natural transformation and twitching motility in piliated *Neisseria gonorrhoeae*. *Molecular Microbiology, 29*(1), 321-330.

Woo, T. D. H., Oka, K., Takahashi, M., Hojo, F., Osaki, T., Hanawa, T., Kurata, S., Yonezawa, H., & Kamiya, S. (2011). Inhibition of the cytotoxic effect of *Clostridium difficile in vitro* by *Clostridium butyricum* MIYAIRI 588 strain. *Journal of Medical Microbiology, 60*(11), 1617-1625. https://doi.org/https://doi.org/10.1099/jmm.0.033423-0

Woods, C., Humphreys, C. M., Rodrigues, R. M., Ingle, P., Rowe, P., Henstra, A. M., Köpke, M., Simpson, S. D., Winzer, K., & Minton, N. P. (2019, Oct). 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

Woods, C., Humphreys, C. M., Tomi-Andrino, C., Henstra, A. M., Köpke, M., Simpson, S. D., Winzer, K., & Minton, N. P. (2022, Apr 12). Required gene set for autotrophic growth of *Clostridium autoethanogenum*. *Appl*

Environ Microbiol, 88(7), e0247921. https://doi.org/10.1128/aem.02479-21

Worden, R. M., Grethlein, A. J., Jain, M. K., & Datta, R. (1991, 1991/05/01/). Production of butanol and ethanol from synthesis gas via fermentation. *Fuel, 70*(5), 615-619. https://doi.org/https://doi.org/10.1016/0016-2361(91)90175-A

Wozniak, K. J., & Simmons, L. A. (2022). Genome Editing Methods for *Bacillus subtilis*. *Methods Mol Biol, 2479*, 159-174. https://doi.org/10.1007/978-1-0716-2233-9_11

Xie, M., & Fussenegger, M. (2018, 2018/08/01). Designing cell function: assembly of synthetic gene circuits for cell biology applications. *Nature Reviews Molecular Cell Biology, 19*(8), 507-525. https://doi.org/10.1038/s41580-018-0024-z

Xu, T., Li, Y., Shi, Z., Hemme, C. L., Li, Y., Zhu, Y., Van Nostrand, J. D., He, Z., & Zhou, J. (2015). Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. *Applied and Environmental Microbiology, 81*(13), 4423-4431.

Yang, Q., Yang, Q., Xu, S., Zhu, S., & Zhang, D. (2020, 2020/11/10/). Technoeconomic and environmental analysis of ethylene glycol production from coal and natural gas compared with oil-based production. *Journal of Cleaner Production, 273*, 123120. https://doi.org/https://doi.org/10.1016/j.jclepro.2020.123120

Yang, S.-J., Kataeva, I., Hamilton-Brehm, S. D., Engle, N. L., Tschaplinski, T. J., Doeppke, C., Davis, M., Westpheling, J., & Adams, M. W. W. (2009). Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe *Anaerocellum thermophilum* DSM 6725. *Applied and Environmental Microbiology, 75*(14), 4762-4769. https://doi.org/doi:10.1128/AEM.00236-09

Yasui, K., Kano, Y., Tanaka, K., Watanabe, K., Shimizu-Kadota, M., Yoshikawa, H., & Suzuki, T. (2009, Jan). Improvement of bacterial transformation efficiency using plasmid artificial modification. *Nucleic Acids Res, 37*(1), e3. https://doi.org/10.1093/nar/gkn884

Zeldes, B. M., Keller, M. W., Loder, A. J., Straub, C. T., Adams, M. W. W., & Kelly, R. M. (2015, 2015-November-05). Extremely thermophilic microorganisms as metabolic engineering platforms for production of

fuels and industrial chemicals [Review]. *Frontiers in Microbiology, 6*. https://doi.org/10.3389/fmicb.2015.01209

Zhang, J., Hong, W., Zong, W., Wang, P., & Wang, Y. (2018). Markerless genome editing in *Clostridium beijerinckii* using the CRISPR-Cpf1 system. *Journal of Biotechnology, 284*, 27-30.

Zhang, J., Taylor, S., & Wang, Y. (2016, 2016/10/01/). Effects of end products on fermentation profiles in *Clostridium carboxidivorans* P7 for syngas fermentation. *Bioresource Technology, 218*, 1055-1063. https://doi.org/https://doi.org/10.1016/j.biortech.2016.07.071

Zhang, J., Yan, X., Park, H., Scrutton, N. S., Chen, T., & Chen, G.-Q. (2024, 2024/02/01/). Nonsterile microbial production of chemicals based on Halomonas spp. *Current Opinion in Biotechnology, 85*, 103064. https://doi.org/https://doi.org/10.1016/j.copbio.2023.103064

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.

Zhang, L., Zhao, R., Jia, D., Jiang, W., & Gu, Y. (2020, 2020/12/01). Engineering *Clostridium ljungdahlii* as the gas-fermenting cell factory for the production of biofuels and biochemicals. *Current Opinion in Chemical Biology, 59*, 54-61. https://doi.org/https://doi.org/10.1016/j.cbpa.2020.04.010

Zhang, X., Lin, Y., & Chen, G.-Q. (2018). Halophiles as Chassis for Bioproduction. *Advanced Biosystems, 2*(11), 1800088. https://doi.org/https://doi.org/10.1002/adbi.201800088

Zhang, Y., Grosse-Honebrink, A., & Minton, N. P. (2015). A universal mariner transposon system for forward genetic studies in the genus *Clostridium*. *PLOS ONE, 10*(4), e0122411. https://doi.org/10.1371/journal.pone.0122411

Zhong, J., Karberg, M., & Lambowitz, A. M. (2003). Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Research, 31*(6), 1656-1664. https://doi.org/10.1093/nar/gkg248

Zhou, J., Horton, J. R., Blumenthal, R. M., Zhang, X., & Cheng, X. (2021, 2021/06/08). *Clostridioides difficile* specific DNA adenine methyltransferase CamA squeezes and flips adenine out of DNA helix. *Nature Communications, 12*(1), 3436. https://doi.org/10.1038/s41467-021-23693-w

Zhu, Y., Liu, X., & Yang, S. T. (2005). Construction and characterization of pta gene-deleted mutant of *Clostridium tyrobutyricum* for enhanced butyric acid fermentation. *Biotechnology and Bioengineering, 90*(2), 154-166.