Development & Exploitation of Gene Tools For Metabolic Engineering In Saccharolytic Clostridia

MUHAMMAD EHSAAN, MSc

Thesis submitted to the University of Nottingham for the degree of

Doctor of Philosophy

Clostridia Research Group (CRG), Molecular Medical school,

Centre for Biomolecular Sciences

DEC 2013

Declaration

Unless otherwise acknowledged, the work presented in this thesis is my own. No part has been submitted for another qualification in the University of Nottingham or anywhere else.

Muhammad Ehsaan

DEC 2013

Abstract

C. acetobutylicum ATCC 824 is a well characterized microorganism known for its ability to produce solvents using the Acetone-Butanol-Ethanol (ABE) fermentation process. It can utilize a variety of C_5 and C_6 sugars, but cannot directly access the complex lignocellulose plant cell wall material which is the most abundant source of carbon in nature. Sophisticated genetic tools are required to enhance the substrate utilisation ability of the organism by incorporating synthetic operons using a synthetic biology approach. Efficient tools were developed for making precise alterations to the *C. acetobutylicum* genome using either heterologous *pyrE* or *codA* genes as counterselection markers. In the case of the former, the utility of the method was also demonstrated in *Clostridium difficile.* The robustness and reliability of the methods were demonstrated through the creation of in-frame deletions in two genes (*spo0A*, *amylase*) using *pyrE* and also two genes (Cac1502 and Cac2071 (*spo0A*) using *codA*. The *pyrE* system is reliant on the initial creation of a *pyrE* deletion mutant using Allele Coupled Exchange (ACE), that is auxotrophic for uracil and resistant to fluoroorotic acid (FOA). This enables the subsequent modification of target genes by allelic exchange using a heterologous *pyrE* allele from *C. sporogenes* as a counter-/negative-selection marker in the presence of FOA. Following modification of the target gene, the strain created is rapidly returned to uracil prototrophy using ACE, allowing mutant phenotypes to be characterised in a *pyrE* proficient background. Crucially, wild-type copies of the inactivated gene may be introduced into the genome using ACE concomitant with correction of the *pyrE* allele. This allows complementation studies to be undertaken at an appropriate gene dosage, as opposed to the use of multicopy autonomous plasmids. The rapidity of the 'correction' method (5–7 days) makes *pyrE* strains attractive hosts for mutagenesis studies.

Publications related to this project

Heap JT, **Ehsaan M**, Cooksley CM, Ng YK, Cartman ST, Winzer K, Minton NP. (2012), **Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker**. Nucleic Acids Research. 40(8):e59

Ng YK*, **Ehsaan M***, Philip S, Collery MM, Janoir C, Collignon A, Cartman ST, Minton NP. (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. doi:10.1371/journal.pone.0056051

*These authors contributed equally to this work.

Acknowledgements

Praise is due to almighty ALLAH for all He Subḥānahu ūta'āla has given to me, His compassion and mercifulness to allow me to complete my PhD. After that, I would like to thank my supervisors, *Professor* Nigel Minton and Dr. Klaus Winzer for giving me the opportunity to carry out my studies for PhD degree in such a prominent group in the Clostridia world under their kind supervision, goodwill and mentoring over the years. I am particularly grateful to Drs. John Heap and Steve Cartman for all the help and support given to me throughout my studies. John's assistance, suggestions and guidance in the lab has played a major role in enhancing my lab expertise and critical thinking, writing skills and leading to the successful publications came out of this work.

I would also like to express my sincerest gratitude to all friends, colleagues in Clostridia Research Group (CRG) at Nottingham specially Drs. Katalin Kovacs, Katrin Schwarz, Sarah Keuhne, Tom Bailey, David Walker, Dave Burns, Yen Kuan Ng (Rachel Ng), Aleksandra Kubiak (Ola), Sheryl Philip, Mark M. Collery, Ying Zhang for their immense support, continual guidance, motivation, and encouragement throughout this study. I would also like to say my thanks to all the current and past members of CRG.

I would also like to thank TMO Renewables Ltd for funding which finally made my goal possible.

Finally I am extremely grateful to my family, especially my parents for their support, patience, encouragement and prayers for the successful completion of this study.

List of Figures

List of Tables

Contents

Abbreviations

ATCC	American Type Culture Collection
BLAST	Basic Local Alignment Search Tool
CDEPT	Clostridial Directed Enzyme Pro-drug Therapy
CFU	Colony Forming Unit(s)
CRG	Clostridia Research Group
2xYTG	Two-times Yeast-Tryptone-Glucose medium
CGM	Clostridial growth medium
CBM	Clostridial Basal Medium
CBMx	Clostridial Basal Medium with xylose
CBMcmc	Clostridial Basal Medium with Carboxymethyl cellulose
T6	T6 growth medium
EtBr	Ethidium bromide
EtOH	Ethanol
5-FOA	5-fluorooratic acid
CD	Cytosine Deaminase
5-FC	5-fluorocytosine
5-FU	5-fluorouracil
IPTG	Isopropyl β-D-1-thiogalactopyranosid
LB	Luria-Bertani medium
MCS	Multiple Cloning Site
NADH	Nicotinamide Adenine Dinucleotide, reduced
NCIMB	National Collection of Industrial, Marine and Food Bacteria
NTR	Nitroreductase
OD	Optical Density
ORF	Open Reading Frame
PBS	Phosphate-Buffered Saline

PCR	Polymerase Chain Reaction
SOE PCR	Splicing by overlap extension PCR
ACE	Allele-Coupled Exchange
SHA	Short Homology arm
LHA	Long Homology arm
RHA	Right Homology Arm
LHA	Left Homology Arm
mRNA	Messenger RNA
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
RBS	Ribosome Binding Site
DMSO	Dimethylsulfoxide
EDTA	Ethylenediaminetetraacetic acid
EPB	Electroporation Buffer
SDS	Sodium dodecyl sulfate
TAE	Tris Base-Acetic Acid-EDTA Buffer
Ta	Annealing temperature
Tm	Melting temperature
Tm	Thiamphenicol
Tm^S	Thiamphenicol-sensitive
Tm^R	Thiamphenicol-resistance
Erm	Erythromycin
Erm^S	Erythromycin-sensitive
Erm^R	Erythromycin-resistance
Cm^R	Chloramphenicol-resistance
RE	Restriction enzymes
tRNA	Transfer RNA
UV	Ultraviolet
CDI	*Clostridium difficile* Infection

RAM	Retrotransposition-Activated selectable Marker
°C	Degree Celsius
kb	Kilobase pair
bp	Base pair
rpm	Revolutions per minute
µg	Microgram
g	Gram
mg	Milligram(s)
sec	Second(s)
min	Minute(s)
h	Hour(s)
%	Percentage
µl	Microlitre
ml	Millilitre(s)
µM	Micromolar
mM	Millimolar
M	Molar
o/n	Overnight
w/v	Weight per volume
v/v	Volume per volume
V	Volt
kV	Kilovolt(s)
cm	Centimetre
µF	Microfarad(s)
Ω	Resistance
U	Units
g	Gravity
∞	Infinity

Chapter One: General Introduction

1 Introduction

Members of the genus *Clostridium* are well known for their pathogenicity, therapeutic use and industrial applications. In this chapter, the general and industrial importance of these organisms, is reviewed, together with a discussion of the currently available genetic modification systems that may be deployed to both better understand the biology of the genus, and to facilitate the exploitation of their beneficial properties.

1.1 Clostridia

The genus *Clostridium* belongs to the phylum Firmicutes which consists of Gram- positive anaerobic bacilli which survive in the presence of unfavourable conditions such as oxygen, high temperatures and disinfectants due to spore formation (Jones & Woods, 1986; Paredes et al., 2005). This is a very large genus ranking second after *Streptomyces,* consisting of more than 100 species ranging from human pathogens to industrially important species. Toxin producing clostridia include *Clostridium botulinum, Clostridium baratii and Clostridium butyricum* (Hatheway, 1995) while *Clostridium perfringens* and *Clostridium sordellii* cause gas gangrene and necrotizing infections (Bhatnagar et al., 2012). The most common human pathogen *Clostridium difficile,* produces toxin A and B and is responsible for antibiotic associated Pseudomembranous *Coli*tis (PMC) (Bartlett et al., 1978). Spores of some *Clostridium species* can also be used as a delivery system for tumour targeting prodrugs due to their localisation in the necrotic regions found at the centre of tumours, *eg., Clostridium sporogenes, Clostridium oncolyticum* and *Clostridium histolyticum* (Minton, 2003). Due to the absence of the oxygen in

hypoxic/necrotic regions, the spores germinate and the genetically engineered clostridia express enzymes *eg., E. coli* Cytosine deaminase (CD) or nitroreductase (NTR) which convert the harmless prodrugs 5-fluorocytosine (5-FC) and CB1954 into a toxic anticancer drug 5-fluorouracil (5-FU) and a potent DNA-crosslinking cytotoxic agent, respectively, which kill the cancerous cells. This kind of therapy is called Clostridial-Directed Enzyme Prodrug therapy (CDEPT) (Minton, 2003; Theys et al., 2006). Industrially important strains include *Clostridium thermocellum* which use lignocellulosic waste material at high temperature to produce ethanol (Nakayama et al., 2011) while saccharolytic clostridia utilize a variety of substrates from monosaccharides to polysaccharides to produce solvents (acetone, butanol, and ethanol) *eg., Clostridium acetobutylicum* and *Clostridium beijerinckii* (Lopez-Contreras et al., 2000).

1.2 *Clostridium acetobutylicum*

C. acetobutylicum is a Gram-positive spore forming anaerobic *Bacillus* representing a mixture of various strains isolated in the early part of the twentieth century, which produce different fermentation products (Johnson & Chen, 1995). Chaim Weizmann isolated a number of strains during 1912-1914 for the production of solvents by fermentation and one of these was named BY (Gabriel, 1928). The majority of these strains showed promising results in the Acetone Butanol Ethanol (ABE) fermentation process for the production of acetone. At the time, acetone was required by the British Government for munitions manufacture to support their efforts in World War I (Jones & Woods, 1986). The type strain ATCC 824 of *C. acetobutylicum* is the original strain isolated from Connecticut garden soil in 1924 (Weyer & Rettger, 1927).

It was the most extensively used strain in the ABE fermentation process consuming starch and molasses as substrates. Shortage of molasses and its high price triggered the efforts towards the isolation of different strains able to grow on other substrates.

1.3 Interest in bio-butanol production

The increasing rate of increase in the world population has raised concerns over currently available energy resources and increased the demand to search for alternative sources of energy. The currently available resources that have been used so far include 85% of fossil fuels *eg.*, coal, oil, natural gas etc. Furthermore, these energy sources are now decreasing due to excessive usage. It will take millions of years to restore fossil fuels. Moreover high energy prices and the political instability in the oil-rich countries have stimulated the search for alternatives. This would enable individual nations to become self-reliant in the energy sector and to protect their economies from market volatility. Global warming and high carbon dioxide emissions from fossil fuels, which are estimated to rise 55 percent in the next two decades, have also contributed to the need to exploit renewable resources. Their use would be less harmful to environment. Renewable resources include solar energy, wind energy, bio-energy, geothermal energy, hydropower, ocean energy and biomass (Carere et al., 2008; Shunmugam, 2009). Greenhouse gas emissions can be reduced significantly by using biomass from plants (Carere et al., 2008; Crutzen et al., 2008). As this energy source is based on plant material whose use will not result in release of toxic chemicals, such as sulphur or carbon dioxide to the environment, they are environmentally friendly fuels (Shunmugam, 2009).

In terms of the biofuel produced, biobutanol has many advantages over bioethanol as it has (i) a higher energy content; (ii) can make use of existing petrol supply and distribution channels; (iii) can be blended with petrol at higher concentrations, (iv) offers better fuel economy than petrol-ethanol blends; (v) has, unlike ethanol, potential to be used as aviation fuel and (vi) is less evaporative than ethanol (Durre, 2007; Shunmugam, 2009; Zverlov et al., 2006).

1.4 ABE fermentation and *C. acetobutylicum*

Butanol production has a long history as an industrial fermentation process as reviewed by Jones and Woods (Jones & Woods, 1986). It was first reported by Louis Pasteur in 1862 (Durre, 2008). Although many investigators studied butanol production during the latter part of the 19th century, it was only in the early 20th century (1905) that Schardinger reported the production of acetone by fermentation (Jones & Woods, 1986). In the first quarter of the 20th century, a chemist, Chaim Weizmann, from Berlin started working in Perkins's lab at Manchester University on a project initiated by Strange and Graham Ltd concerned with the production of synthetic rubber to overcome its shortage in Britain. Shortly thereafter, Weizmann terminated his contract with the company and continued optimization of the fermentation process at Manchester University (Gabriel & Crawford, 1930; Jones et al., 2000). He isolated and studied many bacterial strains; one was isolated from corn and named *Bacillus granulobacter pectinouorum*, which was later re-named *Clostridium acetobutylicum* (Weisgal & Carmichael, 1963). This strain has the ability to ferment a variety of starchy substrates to produce a mixture of

solvents containing butanol, acetone and ethanol (Durre, 2008; Jones et al., 2000).

During the First World War, there was a shortage of acetone in Britain due to blockade of UK supply lines by Germany. At this time, acetone was produced from calcium acetate which was imported from Germany, Austria and the United States. Weizmann was asked by the British government to produce the acetone used as a colloidal solvent for nitrocellulose, which was used to make cordite (smokeless ammunition). The production of acetone by the Weizmann method was so promising that the Admiralty constructed six distilleries in Great Britain. This raised concerns over consumption of food stuff for acetone production and ABE fermentation was halted in England and moved to Canada, France and India during 1916. From 1916-1918 in Canada, 3000 tons of acetone and 6,000 tons of butanol were produced in total. By then the United States was also involved in the war effort. The US and British governments decided to produce acetone in the Midwest corn belt of the US. After the armistice in 1918, there was no need for bulk quantities of acetone, so all these plants were either shutdown or sold to their local governments (Jones et al., 2000; Jones & Woods, 1986).

During World War I, butanol was accumulated and stored as an unwanted by-product of the fermentation process as its own demand did not match that of acetone. In 1920, the US government introduced 'Prohibition' and banned drinkable alcoholic beverages, which created extreme shortages of industrial solvents, such as amyl alcohol, which was also obtained via alcoholic fermentation (Durre, 2008). At that time the automobile industry of Henry Ford was rapidly expanding and required a quick-drying lacquer which would give a

good finish to the car bodies. Butanol and its ester, butyl acetate were ideal solvents for nitrocellulose lacquer (Jones & Woods, 1986; Killeffer, 1927). Until the mid-20th century, approximately 66% of the butanol in the world was produced biotechnologically (Durre, 2008). There was a tremendous increase in the number of the fermenters at the plant in Peoria during 1924-1926 to produce 100 tons of solvent per day, 60 tons per day being used for the production of lacquer. By the end of the 1927, there were 148 fermenters at Peoria in two plants and also new research laboratories were constructed (Jones & Woods, 1986; Killeffer, 1927).

During Second World War, when demand for acetone was at its peak, the fermentation plant at Bromborough was expanded and continuous units were used in the form of multiple columns in an industrial fermentation to meet the demand. Some acetone was imported from the US by her allies. Plants were also built in other parts of the world including Japan, India, South Africa, USSR and China but they did not remain in production for long. When the war was over in 1945, acetone and butanol were still produced in the US by fermentation on a large scale but soon there was a decline in the ABE fermentation process because of the rapidly growing petrochemical industries which could produce acetone and butanol more cost effectively. Moreover, the costs of molasses rose substantially due to its use in cattle feed (Durre, 2008; Jones & Woods, 1986). Generally batch fermentation is carried out on an industrial scale having a capacity of 50,000 to 200,000 gallons due to the biphasic nature (acidogenic and solventogenic) of *C. acetobutylicum* during solvent production. The substrate is first cooked for a certain time and at a specific temperature according to the substrate used, and fermenters filled to

90-95 % of their capacity. Spore cultures are then added by heat activation and mixed up with substrate by bubbling sterile carbon dioxide into the fermenters. Solvent ratios in a batch fermentation varied but a typical ratio of 6:3:1 (butano:acetone:ethanol) was obtained (Jones & Woods, 1986). Due to the toxic nature of butanol, it never exceeded a concentration of more than 2%v/v. To over-come this problem, butanol was continuously removed by gas-stripping below the toxic level to increase the titre by 33% compared to control (Ezeji et al., 2003). In the modern era, the high prices of crude oil and increasing concern over global warming, has led to renewed interest and scientific research into the production of butanol from renewable resources. A number of reviews and research articles have been published describing isolation of new clostridial strains for the ABE fermentation process (Berezina et al., 2009; Durre, 2008; Huang et al., 2004; Lee et al., 2008b; Montoya et al., 2000; Napoli et al., 2010). None of these isolated strains has the potential to grow on raw biomass, crystalline cellulose and to produce significant amounts of butanol. Genetic manipulation methods are being used in various places to make a recombinant strain, which can produce high amounts of butanol from crystalline cellulose.

1.5 Fermentation substrates for solvent production

A variety of raw materials, mainly from food crops, have been used in the ABE fermentation process to produce butanol (Jones & Woods, 1986; Zaldivar et al., 2001). The conventional ABE fermentation process has some limitations due to the high prices of the major substrates used for the production of butanol. Millet, rye, molasses, whey permeate, and corn maize have been used as a feedstock for the production of butanol (Ezeji & Blaschek, 2008; Ezeji et al.,

2007; Qureshi et al., 2006). Other potential substrates could be potatoes, rice, jawari, bajra, tapioca, cassava and Jerusalem artichokes. *C. acetobutylicum* and other saccharolytic strains can metabolize various pentose (C_5) and hexose (C_6) sugars but are unable to degrade pure crystalline cellulose to produce biofuels.

Lignocellulose, a complex natural material composed of lignin, cellulose, and hemicelluloses, is the most abundant renewable resource for the ABE fermentation process. It is produced on large scale from wild plants as well as cultivated crops. The fermentation of this material could use waste materials and produce biofuels and chemicals independent of petrochemical feedstocks (Jones & Woods, 1986; Lee et al., 1985). Xylose is the major component of the hemicellulose in addition to other pentose sugars but the fermentation rate of xylose is lower than of glucose (Fond et al., 1986a; Fond et al., 1986b). The genes for carboxymethyl cellulase and cellobiase are found on the chromosome of *C. acetobutylicum* P270 and it produces cellulolytic enzymes, but they are not as efficient in their ability to degrade amorphous crystalline cellulose compared to cellulolytic bacteria. These enzymes also required molasses for induction (Allcock & Woods, 1981). Various research articles have been published showing the use of pre-treatment procedures, either chemical or physical, to disrupt hemicellulose-lignin-cellulose complexes found in plant cell walls to make cellulose more accessible to the hydrolytic enzymes (Rosgaard et al., 2006; Sheehan & Himmel, 1999).

A novel approach of co-culture has been used to overcome this drawback of *C. acetobutylicum*. A mesophilic *Clostridium* H10 was used for the hydrolysis of cellulose to release sugars followed by the conversion of the sugars into solvents by *C. acetobutylicum* (Fond et al., 1983). Another substrate for ABE

fermentation is xylan, composed of hexose and pentose sugars. It is found in abundance in plant hemicellulose and varies in composition between plant species. *C. acetobutylicum* ATCC 824 has the genes necessary to produce two different kinds of xylanases to convert xylan into xylose for solvent production. However, these xylanases are not very stable at the pH required for solvent production (Bahl et al., 1982; Biely, 1985). Although *C. acetobutylicum* offers great potential for the ABE fermentation process, utilizing various carbohydrates, the activities of many of the enzymes are not yet optimized or fully understood. It was suggested that genetic engineering methods can be used to clone genes of higher activity from other strains into *C. acetobutylicum* to increase the substrate range for ABE fermentation process (Jones & Woods, 1986).

1.6 Isolation of clostridial strains

Clostridium is a very diverse genus of anaerobic, spore-forming, Gram-positive, rod shaped bacteria. Their habitats range from environments with moderate to a very high temperature (Jones & Keis, 1995). The main interest in this genus was to produce solvents in the beginning the 20^{th} century for production of the synthetic rubber from higher alcohols. It was in 1911, when Fernbach isolated a solvent producing strain FB able to ferment potatoes but not maize mash. In the next two years, Weizmann isolated and characterised various strains. He designated a strain called BY which was particularly effective in the ABE fermentation process (Jones & Woods, 1986). This was the most extensively used strain in the ABE fermentation process using starch and molasses as substrates. Shortage of molasses and its high price triggered research towards identifying suitable organisms which could grow on other

substrates, as discussed in Section 1.5. Many solvent producing strains were isolated from soils or mud, roots of leguminous plants, potatoes and cereals (Calam, 1980).

Recently the biodiversity of Colombia was exploited to isolate a number of new solvent producing strains from land cultivated with a variety of crops. A number of these strains were isolated form land where industrial crops cotton and sugarcane were cultivated as well as from grassy land (Montoya et al., 2001; Montoya et al., 2000). Some cellulase producing thermophilic bacteria were isolated from soil and water samples near the hot spring in Egypt by enrichment of the samples with cellulose and they have a high activity against crystalline cellulose avicel, CMC and cellobiose (Ibrahim & El-diwany, 2007). Some mesophilic clostridial strains, which can degrade cellulose, have been isolated from animal and human faeces and municipal waste samples as well as from forest soil (Benoit et al., 1992; Ozkan et al., 2001; Popoff, 1984; Varel et al., 1995; Warnick et al., 2002). Most recently, some saccharolytic strains of clostridia have been isolated from a variety of samples collected from cultivated areas in Germany. These new isolates can ferment a variety of substrates as previously described (Berezina et al., 2009). So new strains and species can be isolated from a variety of environmental samples for example soil, mud, silt, plant decomposed waste, sewage waste collected from different areas of the world to explore their fermentation substrates and products.

1.7 Phylogeny of clostridia

Living organisms were classified on the basis of their phenotypic characteristics into three domains of life archaea, bacteria, and eukaryote

(Woese et al., 1990). These domains were further classified into groups, kingdoms, phyla, classes, orders, families, genera and species. However, although there are similarities between these groups, their genetic makeup can vary significantly between one another. To investigate the relationships among the members of this diverse group of microorganisms, comparative analysis of various housekeeping genes has been undertaken to establish evolutionary relationships. Ribosomal RNA (rRNA) is found to be the most conserved genes during evolution (Smit et al., 2007). Bacteria have three ribosomal genes 16S, 23S, and 5S rRNA which are organised in an operon (Brosius et al., 1981). 16S rRNA is the most extensively studied gene in prokaryotes to understand the emergence of new species and strains and to classify them (Garrity & Holt, 2001). The 16S rDNA sequences from various organisms can be aligned precisely to draw the phylogenetic relationship among them. The invention of PCR, by Kary Mullis in 1983, revolutionized the study of 16S rRNA sequences and a number of bacteria have since been reclassified on the basis of their 16S sequence.

The members of the genus *Clostridium* were classified on the basis of their solvent production profile and the range of substrates they can utilise prior to the use of 16S rRNA for classification (Johnson & Chen, 1995; McCoy et al., 1926; McCoy et al., 1930; Weyer & Rettger, 1927). Simplistically, the majority of these strains were designated *Clostridium acetobutylicum*, until 1995 when molecular approaches were applied and 55 strains isolated during 1915-1931 of solvent producing clostridia were studied and classified into nine groups on the basis of biotyping and DNA fingerprinting. A single prototype strain was selected from each group and subjected to DNA screening by partial 16S rRNA

gene sequence analysis. This study concluded that sequence between nucleotides 830 to 1383 (*E. coli* numbering) was the most variable region of the 16S rRNA and provided a quick method to identify the solvent producing strains. Nine groups were classified into four taxonomic groups; *C. acetobutylicum* ATCC 824^{TM}, *Clostridium beijerinckii* group (with VPI 5481T [= ATCC 25752 T], *Clostridium sp.* strain NRRL B643 (or NCP 262), and "*C. saccharoperbutylacetonicum*" N1-4 (Keis et al., 1995). 16S rRNA sequence was also used to differentiate between *Clostridium thermocellum* and *Thermoanaerobacter* (Ozkan et al., 2001).

A number of phenotypic traits can be used to classify members of the clostridia group but they should be used in connection with established 16s rRNA sequences. The above mentioned four groups can be distinguished from each other on the basis of sensitivity or resistance to rifampicin and production of riboflavin in milk medium. *C. acetobutylicum* ATCC 824^{TM} and *Clostridium sp.* strain NRRL B643 (or NCP 262) were found susceptible to 10 and 100ng of rifampicin per disc while *Clostridium beijerinckii* and *C. saccharoperbutylacetonicum*" N1-4 were resistant (Keis et al., 1995). Strains of *C. acetobutylicum* group form curd in milk medium within 24 hours while other three groups do not form curd even after incubation for 10 days at 4°C (Johnson et al., 1997). It is concluded that classification of new isolates should be carried out on the basis of the 16S rRNA sequence, %DNA relatedness and DNA fingerprinting analysis.

1.8 Clostridial strain degeneration

It has been reported in the literature for more than 100 years, that *C. acetobutylicum* loses the ability to produce solvents if serially sub-cultured for long periods of time (Hartmanis et al., 1986; Kutzenok & Aschner, 1952), though the mechanisms involved are currently unknown. A series of morphological and biochemical changes are described leading to loss of ability to produce solvents and spores. Although spore formation is not required for solvent production, it appears to have some indirect regulatory involvement in solventogenesis (Kashket & Cao, 1995). It was only recently shown that *C. acetobutylicum* strains contain a 210-kb plasmid called pWEIZ which carries the pivotal genes required for solvent production. The plasmid was later renamed pSOL1 (Cornillot & Soucaille, 1996). During an investigation of the two degenerated stains M5 and DG1 and *C. acetobutylicum* ATCC 824, it was confirmed that the degenerated strains do not contain the genes for solvent production and had lost pSOL1. These degenerated strains have been complemented for acetone and butanol genes cloned on plasmids (Cornillot et al., 1997). *C. acetobutylicum* ATCC 824 produces an extracellular α-amylase which has been purified but strain DG1 which lost pSOL1, does not produce an α-amylase when grown on glucose suggesting that this gene might be located on pSOL1 and could be used as marker to monitor degeneration of strains and loss of pSOL1 (Sabathe et al., 2002).

Other saccharolytic clostridial strains do not carry such a plasmid, and the solvent genes are located in the chromosome. However, they are not immune to degeneration, which still occurs upon continuous culture. During the fermentation of glucose to acetic and butyric acids at a high rate by *C.*

beijerinckii NCIMB 8052 (formerly *C. acetobutylicum*), large amount of acids accumulate which tends to decrease the pH of the medium. This kills most of the cell population and also causes the cells to degenerate and lose the ability to make solvents (Kashket & Cao, 1993). This can be prevented by the addition of acetate to the medium (Chen & Blaschek, 1999; Gu et al., 2009). There are many other factors involved in the degeneration of a strain and most extensively studied is *spo0A* (Harris et al., 2002; Ravagnani et al., 2000). It was shown that inactivation of this gene in *C. beijerinckii* completely abolishes the formation of acetone and results in only low levels of butanol being produced. In the case of *C. acetobutylicum*, sporulation was severely affected without fully degeneration of the strain which suggests that solvent production is dependent on the presence of a functional *spo0A* gene. Over-expression of *spo0A* accelerated the sporulation process as well as solvent formation compared to the wild type strain (Harris et al., 2002).

1.9 Metabolic pathway of *C. acetobutylicum*

Production of solvents by *C. acetobutylicum* via the ABE fermentation process is divided into two phases; acidogenic and solventogenic as shown in Figure 1.1 (Lee et al., 2008b). During the first phase, *C. acetobutylicum* produces hydrogen, carbon dioxide, acetate and butyrate as the major products and this all occurs during the exponential growth phase. As the cells enter the stationary growth phase, metabolism undergoes a shift to solvent production, the solventogenesis phase (Andersch et al., 1983; Hartmanis & Gatenbeck, 1984). In the solventogenesis phase, the acids are converted into solvents by the action of two sets of dehydrogenases. This change from one phase to another is the result of alterations in gene expression involved in the central metabolic

pathway (Durre et al., 1987). The end product of the glycolysis of hexose sugars is pyruvate produced by the Embden–Meyerhof–Parnas (EMP) pathway. 1mol of hexose sugar produces 2mol of pyruvate as well as 2mol of adenosine triphosphate (ATP) and 2mol of reduced nicotinamide adenine dinucleotide (NADH) (Gheshlaghi et al., 2009; Rogers, 1986). The solvent-producing clostridia also metabolize pentose sugars via the pentose phosphate pathway (Ounine et al., 1983). The pentoses are phosphorylated to pentose 5-phosphate and converted by transketolase and transaldolase to fructose-6-phosphate and glyceraldehyde-3-phosphate, which enter the glycolytic pathway. The fermentation of 3mol pentose to pyruvate produces 5mol ATP and 5 mol NADH (Rogers, 1986). The end product of glycolysis is pyruvate which is then cleaved by pyruvate-ferredoxin oxidoreductase in the presence of coenzyme A (CoA) to form acetyl-CoA and CO2 reduced ferredoxin (Jones & Woods, 1986). Acetyl-CoA is then converted into either oxidized products (*ie.*, acetone, acetate, or CO2) or reduced products (*ie.*, butanol, ethanol, or butyrate). A cascade of four enzymes (thiolase, β-hydroxybutyryl-CoA dehydrogenase, and crotonase and butyryl-CoA dehydrogenase) is required to form butyryl-CoA from acetyl-CoA which serves as an intermediate for the formation of butyrate and butanol from acetyl-CoA (Hartmanis & Gatenbeck, 1984; Seedorf et al., 2008). A change in the flow of carbon triggers the switch from acidogenic phase to solventogenic phase. The important intermediates during this phase for the production of ethanol and butanol are acetyl-CoA and butyryl-CoA respectively, which are then reduced into acetylaldehyde and butyraldehyde by two sets of dehydrogenases. Butyryl-CoA is reduced to form butanol by two dehydrogenases; butyraldehyde dehydrogenase and butanol

dehydrogenase while acetyl-CoA is reduced by acetoaldehyde dehydrogenase and ethanol dehydrogenase (Jones & Woods, 1986). CoA-transferase then converts acetoacetyl-CoA into acetoacetate resulting in the conservation of CoA unit which can be reused in the pathway. Acetoacetate decarboxylase then converts acetoacetate into acetone in an irreversible decarboxylation reaction.

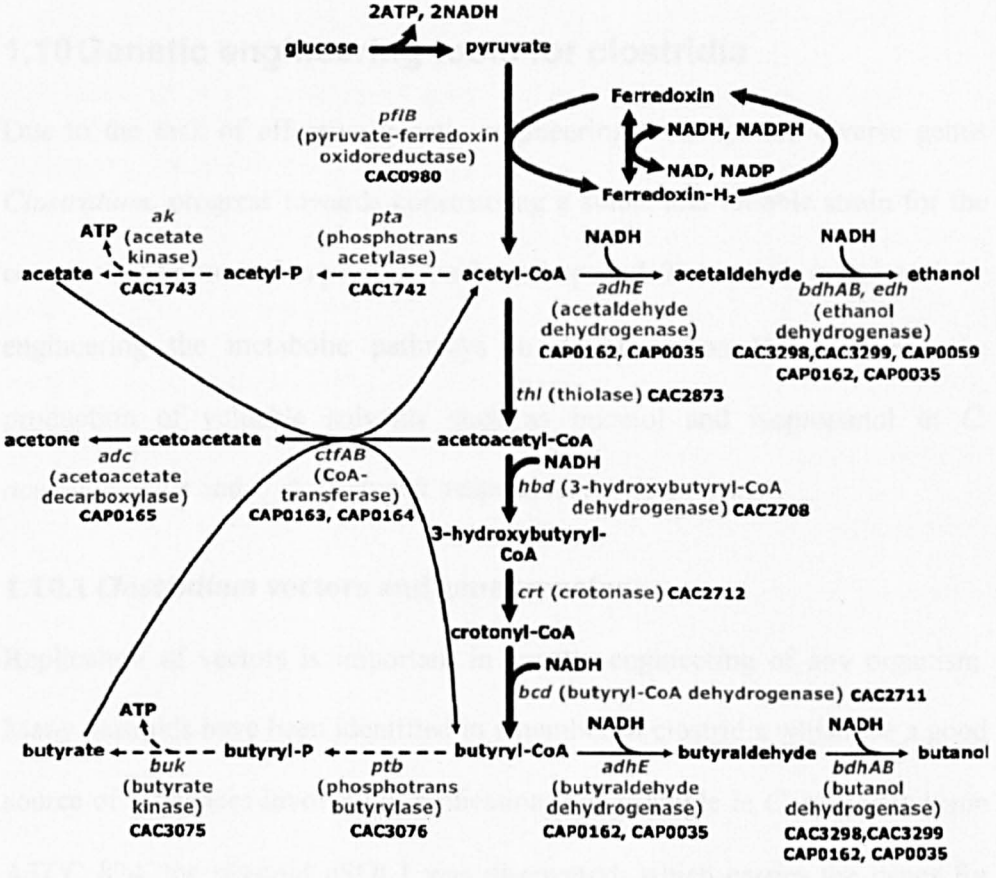


Figure 1.1 Metabolic pathways in *C. acetobutylicum*

Reactions during acidogenesis and solventogenesis are indicated by dotted and solid arrows, respectively. Bold arrows indicate reactions which activate the whole fermentative metabolism. Grey letters indicate genes and enzymes for the reactions. CAC and CAP numbers are the ORF numbers in the chromosome and megaplasmid, respectively. Taken from (Lee et al., 2008b).

It has also been shown that the uptake of acetate and butyrate only occurs when acetoacetyl decarboxylase is induced. The optimum pH for this enzyme is 5.0 and the acidic condition is favourable for the production of acetone which is a

final end product in the metabolic pathway of *C. acetobutylicum*. In contrast some strains of *C. beijerinckii* reduce acetone to form isopropanol by the action of isopropanol dehydrogenase (Andersch et al., 1983). Acetyl-CoA is reduced into acetaldehyde and subsequently into ethanol by acetaldehyde dehydrogenase and ethanol dehydrogenase.

1.10 Genetic engineering tools for clostridia

Due to the lack of effective genetic engineering tools for the diverse genus *Clostridium*, progress towards constructing a stable and reliable strain for the commercial fermentation process has been impeded. These tools are pivotal for engineering the metabolic pathways to direct carbon flow towards the production of valuable solvents such as butanol and isopropanol in *C. acetobutylicum* and *C. beijerinckii*, respectively.

1.10.1 *Clostridium* vectors and gene transfer

Replication of vectors is important in genetic engineering of any organism. Many plasmids have been identified in a number of clostridia which are a good source of sequences involved in replication. For example in *C. acetobutylicum* ATCC 824, the plasmid pSOL1 was discovered, which carries the genes for solvent production in this organism (Cornillot et al., 1997), pCB101 and pCB102 from *C. butyricum* (Minton & Morris, 1981), pCD6 from *C. difficile* which encodes a large replication protein (RepA) and contains a 35 bp repetitive region (Purdy et al., 2002), and pCW4, and pIP404 from *C. perfringens* (Garnier & Cole, 1986; Mihelc et al., 1978).

Due to the problems of transferring, cloning and segregational instability of these vectors in clostridia (Jones & Woods, 1986), a number of *E. coli*-

Clostridium and *B. subtilis-Clostridium* shuttle vectors have been constructed containing origins of replication for both organisms (Lee et al., 1992; Minton et al., 1993). They can be easily cloned and manipulated in *E. coli* or *B. subtilis* and then transferred into clostridia. Origins of replication from *B. subtilis* plasmid pIM13, which replicates via a rolling circle mechanism (Monod et al., 1986), and the *Enterococcus faecalis* plasmid pAMβ1, which replicates via a unidirectional θ mechanism (Clewell et al., 1974), have been used very extensively in constructing new vectors for clostridia. In recent years, a large number of shuttle vectors have been constructed for a number of clostridial species and a modular system was devised to facilitate the rapid construction of vectors with minimal cloning steps (Heap et al., 2009). The shuttle vectors constructed so far vary from species to species due to a requirement for different origins of replication. *Clostridium saccharoperbutylacetonicum* strain N1-4 is one of the highest butanol producing species among all the solvent producing clostridia used in industrial ABE fermentation in the past, and belongs to a different phylogenetic group to *C. acetobutylicum* and *C. beijerinckii* (Johnson et al., 1997; Keis et al., 1995). The shuttle vectors constructed for other solventogenic clostridia were not suitable for *Clostridium saccharoperbutylacetonicum* strain N1-4. Therefore a new shuttle vector pNAK1 was constructed based on pTYD101 from strain N1-4 and an *ori* from *E. coli* vector pHSG397 (Nakayama et al., 2007).

The *C. acetobutylicum* ATCC 824 genome contains a type II restriction system *Cac824I* which recognizes the sequence 5`-GCNGC-3`, found very frequently in *E. coli-Clostridium* shuttle vectors and prevents efficient DNA transfer into the cell. To overcome this problem, a new *B. subtilis/C. acetobutylicum* shuttle

vector pFNK1 was constructed to transform *C. acetobutylicum* via electroporation (Mermelstein et al., 1992). Similarly, DNA was transferred via conjugation into *C. difficile* using origin of replication from vector pCD6 to overcome the restriction barriers in this strain (Purdy et al., 2002). An *in vivo* methylation system with *B. subtilis* Phage ϕ3T I *Methyltransferase* was developed to protect the recombinant shuttle-vectors constructed in *E. coli,* before transferring them into *C. acetobutylicum* ATCC 824 (Mermelstein & Papoutsakis, 1993).

1.10.2 Group II intron retargeting and the ClosTron

Engineering the metabolic pathways of the industrial solvent producing clostridia has been impeded due to lack of effective gene knockout (KO), knockdown (KD) and knock-in (KI) systems despite the fact that a transformation protocol and *in vivo* methylation system were developed more than 20 years ago (Mermelstein & Papoutsakis, 1993). Some success has been reported in the use of antisense RNA. Such systems are particularly useful for *C. acetobutylicum* to KD the expression of those genes which cannot be deleted or mutated (Tummala et al., 2003). However, they have not been widely adopted. As an alternative, an insertional mutagenesis system, TargeTron, was developed based on the bacterial group II intron from *Lactococcus lactis ltrB* (*Ll. LtrB* intron) which was used in a number of bacteria to make mutants more reliably. Group II introns were explored by Lambowitz and co-workers in a series of studies, where they determined the mechanism of intron movement from one location to another in the genome and its insertion to any target site (Karberg et al., 2001; Mohr et al., 2000; Zhong et al., 2003). Bacterial Group II introns are retro-elements and consist of

a catalytic RNA and a multifunctional open reading frame (ORF) encoding for an Intron-Encoded Protein (IEP) LtrA. LtrA controls the mobility of the intron and is provided transiently on a plasmid during mutagenesis and removed afterwards to ensure the stability of mutants. The group II intron (*Ll. LtrB* intron) could be re-targeted to any site in the genome by making specific changes in intron RNA. IEP as well as base pairing of the intron RNA at the target sites are necessary for recognition of specific target sites in the DNA (Lambowitz & Zirnmerly, 2004). This provided the basis for designing an algorithm to be used to reliably predict the target site for the Ll.LtrB derived intron (Karberg et al., 2001; Perutka et al., 2004). These re-targeted introns were called "TargeTrons" (Zhong et al., 2003) which uses an antibiotic marker in the group II intron interrupted by a self-splicing group I intron from phage *td* (Cousineau et al., 1998). This splices out as the group II intron is inserted into the target site, and the antibiotic resistance gene is restored. Such an antibiotic marker is termed as a Retrotransposition-Activated Marker (RAM). Their deployment overcomes the problem of low integration frequencies as they allow integrants to be positively selected. TargeTron-based plasmids carried kanamycin as the RAM marker and ColEI as origin of replication for Gram-negative bacteria. Neither of these is suitable for use in in the genus *Clostridium.*

To overcome these deficiencies, ClosTron technology was developed based on the group II intron and the clostridial origin of replication from pCB102. The selectable RAM marker developed was based on the *ermB* gene of the *Enterococcus faecalis* plasmid pAMβ1, and encodes for resistance to

erythromycin. The final plasmid developed was designated pMTL007 Figure 1.2 (Heap et al., 2007).

The ClosTron system was later refined and streamlined and made to conform to the pMTL8000 modular system, giving rise to the second-generation vector pMTL007C-E2 (Figure 1.3).

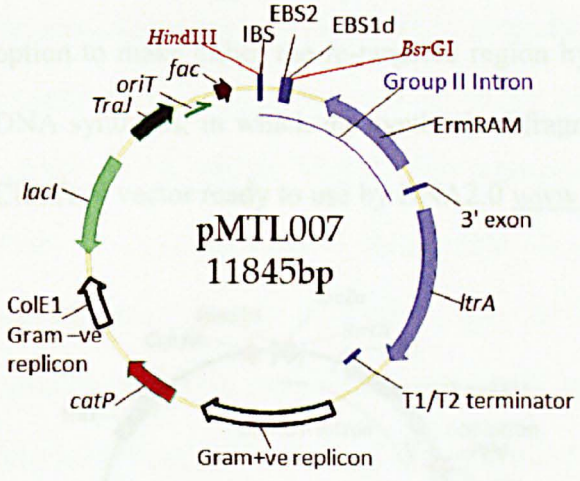


Figure 1.2 First generation ClosTron plasmid, pMTL007

The prototype ClosTron plasmid pMTL007C contains a group II intron from *Ll.LtrB* under the control of the IPTG-inducible *fac* promoter. The ermRAM marker, which is inactivated by phage *td* group I intron, is inserted in the group II intron for positive selection of insertional mutants. It also contains the Gram-positive replicon pCB102 for replication in *Clostridium* species, the Gram-negative replicon ColEI for replication in *E. coli* and antibiotic marker *catP* which confers resistance to chloramphenicol and thiamphenicol in *E. coli* and *Clostridium* species. The re-targeted region is introduced between HindIII and BsrGI restriction endonuclease sites (Heap et al., 2007).

In pMTL007C-E2, the RAM was flanked by FRT sites which could be used to remove the RAM-derived *ermB* gene to allow the subsequent repeated use of the ClosTron to make multiple mutants (Heap et al., 2010b). This has to date only been achieved in *C. acetobutylicum,* where multiple orphan kinase mutants were generated in one strain by three cycles of ClosTron mutagenesis

(Steiner et al., 2011). Attempts to emulate this feat in others species, such as *C. difficile, C. sporogenes,* and *C. botulinum,* have thus far failed.

To make rapid progress towards understanding the molecular biology of solventogenic clostridia, the ClosTron design tool was provided free of charge at http://clostron.com allowing (in a few easy-to-use steps), the research community to design re-targeted ClosTron plasmids. The website provides an option to make either the re-targeted region by conventional SOE PCR or by DNA synthesis, in which the synthesised fragment is cloned into the modular ClosTron vector ready to use by DNA2.0 www.dna20.com.

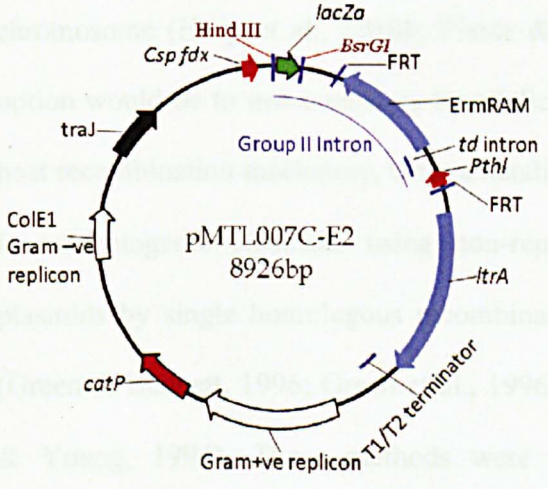


Figure 1.3 Second-generation ClosTron plasmid, pMTL007C-E2

ClosTron plasmid pMTL007C was modified by replacing IPTG-inducible *fac* promoter with the strong *fdx* promoter from *C. sporogenes* to direct expression of the Group II intron and the RAM is flanked by FRT sites to facilitate subsequent FLP-mediated marker removal for multiple intron insertions in a strain. The *E. coli lacI* gene was no longer required and hence removal of it also reduced the plasmid size. Introduction of *lacZalpha* between HindIII and BsrGI facilitated the screening of the re-targeted region between these sites using blue/white screening. It also contains the Gram-positive replicon pCB102 for replication in *Clostridium* species, the Gram-negative replicon ColEI for replication in *E. coli* and antibiotic marker *catP* which confers resistance to chloramphenicol and thiamphenicol in *E. coli* and *Clostridium* species.

1.10.3 Gene knockout by allelic exchange

Although the frequency of making mutants is very high with ClosTron technology, it has a number of drawbacks. A principal weakness of the system is that, as with any insertional mutagen, the insertion of the group II intron can cause polar effects on the genes up-or down-stream of the target site. As a result the observed phenotype may not be a consequence of the inactivation of the target gene, but may be caused by secondary effects on the adjacent gene(s) (Cooksley et al., 2012; Steiner et al., 2011). Additionally, the group II intron is unable to carry foreign "cargo" DNA larger than 1Kb and, therefore, cannot be used for the integration of larger gene (s) or substantial operons into the chromosome (Heap et al., 2010b; Plante & Cousineau, 2006). The preferred option would be to make mutants by allelic exchange, which is reliant on the host recombination machinery. Only a handful of genes have been knocked out in solventogenic clostridia using non-replicating and replication-defective plasmids by single homologous recombination via Campbell-like integration (Green & Bennett, 1996; Green et al., 1996; Liyanage et al., 2001; Wilkinson & Young, 1994). These methods were very tedious requiring laborious screening with a very low reproducibility for identification of the mutants. Moreover, the mutants generated are segregationally unstable due to the presence of two copies of direct repeats in the chromosome (Heap et al., 2007) and requires antibiotic supplementation of growth medium to maintain the mutant genotype. Moreover, generation of multiple mutations in these strains is limited by the number of resistance markers available for use in this genus.

Unmarked and marked gene deletion or replacement methods based on homologous recombination are well established in other organisms using

negative and positive selection markers (Knipfer et al., 1997; Link et al., 1997; Merritt et al., 2007; Wang et al., 2009). The *Bacillus subtilis sacB* gene is the most common negative selection (or counterselection) marker used in Gram-negatives. This gene encodes *levansucrase* (EC2.4.1.10), an enzyme that catalyzes the hydrolysis of sucrose and levan elongation. The majority of the Gram-negative bacteria are sensitive to *levansucrase* which causes the accumulation of levans (high molecular-weight fructose). In this method plasmid containing KO cassettes and the *sacB* gene is transferred to the microorganism. Single crossover integrants are selected in the presence of antibiotic appropriate for the resistance marker present on the plasmid backbone. These are then propagated to allow for the second recombination event to happen and plated onto media supplemented with sucrose to select for double crossover deletion mutants. Cells which have undergone a second recombination and lose the *sacB* containing plasmid represent cells in which the plasmid has excised and can be either double crossover mutants or cells which have reverted to wild type. A number of genes have been deleted using *sacB* in *E. coli* (Donnenberg & Kaper, 1991; Marx, 2008; Ried & Collmer, 1987). A number of negative selection markers have been used in other bacteria for example *pyrF* (*orotidine 5-phosphate decarboxylase*) (Knipfer et al., 1997), *pyrE* (*orotate phosphoribosyl transferase*) (Bitan-Banin et al., 2003), *galK* (*galactokinase*) (Ueki et al., 1996), *codA* (*cytosine deaminase*) (van der Geize et al., 2008), *upp* (*uracil phosphoribosyl transferease*) (Fabret et al., 2002) and *mazF* (Zhang et al., 2006). These mediate the toxicity of their substrates or non-metabolizable analogs, in cells which contain the respective counterselection gene either integrated into chromosome or on autonomous

plasmid. To use these genes as negative selection markers, they need to either be absent or deleted in the organisms of interest while *mazF* relies upon its tight regulation through the use of a strong inducible promoter system.

At the start of this project, there were no reported examples of counter selection systems for use in *Clostridium* spp. During the course of the project, however, a number of systems were described, including the development of *codA* for use in *C. difficile* by members of the Clostridia Research Group at Nottingham. With the exception of *codA*, these systems will be reviewed in the Discussion (Section 4.4) of this thesis.

The *codA* system was developed to make precise changes in *tcdC* in *C. difficile* (Cartman et al., 2012). *CodA* was derived from *E. coli* and codes for cytosine deaminase. It converts non-toxic pyrimidine analogue 5-fluorocytosine (5-FC) to 5-fluorouracil (5-FU) which in the presence of *upp* (*uracil phosphoribosyl transferease*) is toxic to cells. Therefore, the successful use of *codA* as a negative selection maker requires the presence of *upp* and absence of *codA*.

1.11 Strain improvement

A well-known feature of clostridia is the production of solvent by the ABE fermentation process using a wide range of substrates. Due to the low titre of solvents generated, high cost compared to solvent generation from petrochemicals and lack of genetic engineering methods to improve the strains, it's use decreased during the mid-part of the 20th century (Jones & Woods, 1986). Strain improvement has been practised for centuries to improve bacterial strains used for biotechnology applications to maximise product generation and improve profitability. Conventionally, strain improvement was

accomplished by screening, selection of mutations and breeding for specific properties of strains. In addition to traditional mutagenesis strategies, advanced engineering and genetic techniques were also applied to modify the metabolic pathways generating industrially valuable products without significant capital outlay. Metabolic pathways in microbes are controlled by the nucleotide sequences which programme the biological processes which could be improved, reprogrammed to bypass unnecessary pathways and produce the required product using molecular biology techniques (Parekh et al., 2000).

Riboflavin (vitamin B_2) is a very important chemical that is used as a medicine, as a food supplement and as an additive to animal feed. Production of riboflavin is a characteristic trait of *C. acetobutylicum* ATCC 824, but only in very low amounts during fermentation. Although many recombinant strains have been created to increase the solvent concentration, the process is still not commercially viable. To make the process cost effective, the production of additional valuables chemicals is required. In *C. acetobutylicum* ATCC 824, the riboflavin operon consists of a cluster of four genes, namely, *ribG, ribB*, *ribA*, and *ribH*. In an attempt to improve riboflavin production under laboratory conditions, the operon was over-expressed on a plasmid using a promoter from the *C. acetobutylicum ptb* gene. This manoeuvre led to enhanced production of riboflavin, up to 27mg/L in a *Clostridium* culture (Cai & Bennett, 2011).

Another hurdle to be overcome in the commercialization of the ABE fermentation process using solventogenic clostridia is that they are sensitive to total solvent concentrations of > 20g/L and sensitive to butanol concentrations > 13g/L. This could be improved either by removing the excess butanol from

the growing culture using gas stripping technique (Ezeji et al., 2003) or by screening for butanol resistant strains such as SA-1 and SA-2 by gradually increasing the concentration of butanol in the medium and screening for strains with improved resistance to butanol. It is generally believed that such strains have adjusted their lipid membrane contents in order to remain viable (Baer et al., 1987; Li et al., 2010; Lin & Blaschek, 1983).

An alternative strategy for increasing butanol tolerance would be to make rational changes to the genome using recombinant DNA approaches, to isolate mutants producing high yield of solvent but with enhanced butanol tolerance (Ezeji et al., 2010). Others investigated the effects of over-expressing the class I stress response operon *groESL*. This approach in *C. acetobutylicum* ATCC 824 resulted in increased production of both butanol, up to 17.1g/L, and acetone, 8.6g/L (Tomas et al., 2003). Transcriptional analysis has revealed that *spo0A* is also involved in butanol tolerance and over-expression resulted in increased butanol production (Alsaker et al., 2004). Borden and Papoutsakis identified a number of transcriptional regulators such as Cac0977, Cac1463, Cac1869, and Cac2495 which were cloned and expressed on a plasmid in *C. acetobutylicum* ATCC 824. Wild type strains transformed with plasmid expressing Cac1869 showed 81% increase in butanol tolerance, while Cac0003 exhibited 24% increase compared to strains with empty plasmids (Borden & Papoutsakis, 2007).

Furthermore, advanced fermentation techniques for example use of liquid-liquid extraction, perstraction, gas stripping and pervaporation could also be used in conjunction with the recombinant approaches to achieve higher butanol concentrations (Ezeji et al., 2010).

1.12 Synthetic biology and metabolic engineering

The advent of recombinant DNA technology in the 1970s through the pioneering work of Cohen and Boyer, *de novo* synthesis of DNA and the application of engineering principles to biology have enabled the scientific community to make changes in the host at the DNA level to build up new artificial biological systems based on well-characterized components. This new field is called synthetic biology (Lanza et al., 2012). A number of international companies have been established to synthesize the desired genes rather than building them by cloning in the laboratory using traditional cloning methods *eg.*, www.blueheronbio.com, www.dna20.com, www.geneart.com,. These developments have had a great impact on the genetic engineering of microbes for the production of biofuels. It has provided the means to construct synthetic pathways of highly important chemicals (such as glucaric acid, lycopene, 1, 3-propandol, butanediol and hydrogen) in a non-native *E. coli* host (Lanza et al., 2012) by assembling the genes from a variety of organisms in a single heterologous host (Weeks & Chang, 2011). A large library of DNA elements, or BioBricks, has been created during the International Genetically Engineered Machine (iGEM) competition, an annual event that started in 2004 housed at Massachusetts Institute of Technology (MIT). However, little information is available on functionality of these elements (Kwok, 2010). Using standard biological procedures, these parts can be assembled together and tested for any desired outcome (Kelly et al., 2009; Knight, 2003).

Artificial DNA synthesis has reduced the cost and time needed to construct complicated genetic operons encoding the pivotal enzymes of metabolic pathways. Other methods, such as Gibson assembly (Gibson et al., 2009;

Merryman & Gibson, 2012), in combination with G-Blocks (Sabel J et al. (2011) IDT Mutagenesis Application Guide Available at www.idtdna.com), have further reduced the time and effort needed for pathway construction. Due to limited and inadequate number of genes tools for *Clostridium* species, a number of synthetic pathways have been expressed in commonly used laboratory organisms such as *E. coli* and *S. cerevisiae* for the production of valuable chemicals and biofuels (Fu et al., 2011; Lee et al., 2008a; Nakamura & Whited, 2003; Shetty et al., 2008; Vroom & Wang, 2008).

Over the years, increasing advances in gene tools have enabled researchers around the world to begin to engineer the metabolic pathways of solvent producing clostridia. Thus, a number of genes from solvent producing bacteria have been cloned and over expressed (Al-Hinai et al., 2012; Lee et al., 2008b; Lutke-Eversloh & Bahl, 2011; Papoutsakis, 2008). The majority of research activity to date has been concerned with *C. acetobutylicum.* However, as yet no recombinant strain has been found to be suitable for the development of a large scale industrial fermentation process. One major reason is that feedstock costs dictate that any process developed must be able to use lignocellulosic material as the source of sugar. Although the butanol producing saccharolytic clostridia such as *C. acetobutylicum* and *C. beijerinckii,* are unable to grow on cellulose, other *Clostridium* species, for example *C. thermocellum* and *C. cellulolyticum,* can degrade cellulose very efficiently using a multienzyme complex, or nanomachine, the cellulosome (Lamed et al., 1983; Madarro et al., 1991). However, cellulosic clostridial species that produce cellulosomes do not produce butanol. Rather they ferment sugar into the less valuable alcohol, ethanol. One attractive notion is, therefore, to use genetic engineering

approaches to create a butanol producing *Clostridium* that produces a cellulosome, making butanol production cellulosic. Two different approaches that could achieve this goal are evident: (i) either introduce the solvent production pathway into a cellulolytic *Clostridium,* or *vice versa*, or; (ii) use a consortium of two strains; one degrading cellulose and the other one producing solvents. The biggest challenge is to assemble and express the cellulosome in solvent producing clostridia.

1.13 Aim of this project

The overall objective of the current project was to develop reproducible methods for making in-frame deletions in saccharolytic clostridia (using *C. acetobutylicum* as the model organism) and if established, to use these methods to make rational changes to the organism that may improve solvent production. As part of this long-term aim, an attempt was made to isolate a cellulosic, butanol-producing *Clostridium* species. At the outset, our primary aim was to develop procedures that relied on the use of *pyrE* alleles, both as a negative selection marker and to generate the required *pyrE* uracil-requiring auxotrophic mutant host strain. The latter was accomplished using a special form of allelic exchange (ACE), with which it was subsequently possible to correct the host back to uracil prototrophy by rapid restoration of the wild type *pyrE* allele. As a system was developed during the course of the project which allowed the creation of in-frame deletions in *C. difficile* through the use of *codA* as a counterselection marker, its utility was also investigated. The specific aims were:-

1. To exemplify the utility of Allele-Coupled Exchange (ACE) for the insertion of large fragments of DNA into the chromosome, and to create the *pyrE* mutants required for the use of this allele as a negative selection marker.
2. To demonstrate the use of heterologous PyrE alleles as negative selection markers for the isolation of in-frame deletions.
3. To undertake the comparative assessment of *codA* as a negative selection marker.
4. To demonstrate ACE can be used to complement in-frame deletions in the chromosome, as opposed to on an autonomous plasmid.
5. To isolate a new solventogenic species most likely a *Clostridium saccharobutylicum* strain and apply the technology developed in this thesis to enhance its commercial properties in terms of high butanol production, tolerance to high butanol concentration, and substrate utilization for fermentation.

Chapter Two: Materials and Methods

2 Materials and Methods

Bacterial strains used are shown in Table 2.1and were stored at -80 °C

Table 2.1 List of bacterial strains

CRG CC#	Strain	Genotype	Reference
210	*E. coli* TOP10	F⁻ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80lacZΔM15 Δ*lacX74* *deoR* *recA1* *araD139* Δ(*ara-leu*)7697 *galU galK rpsL* (Str^R) *endA1* *nupG*	Invitrogen Ltd[1]
685	*E. coli* pAN2	*E. coli* Top10 strain carrying pAN2 plasmid	(Heap et al., 2007)
	C. acetobutylicum ATCC 824		Rostock, Germany SysMO collection
ME1	*C. acetobutylicum* ATCC 824*pyrF*-lambda2.0 mutant	*pyrF* mutant made by ACE method using pMTL-JH2-lambda2.0 vector	This study
ME2	*C. acetobutylicum* ATCC 824*pyrF*-lambda2.3 mutant	*pyrF* mutant made by ACE method using pMTL-JH2-lambda2.3 vector	This study
ME3	*C. acetobutylicum* ATCC 824*pyrF*-lambda4.3 mutant	*pyrF* mutant made by ACE method using pMTL-JH2-lambda4.3 vector	This study
ME4	*C. acetobutylicum* ATCC 824*pyrF*-lambda6.5 mutant	*pyrF* mutant made by ACE method using pMTL-JH2-lambda6.5 vector	This study
1545	*C. acetobutylicum* ATCC 824 *pyrE* mutant	*pyrE* mutant made by ACE method using pMTL-JH12 vector	This study
ME5	*C. acetobutylicum* ATCC 824 *pyrE*+ve	*pyrE* mutant strain repaired by using ACE vector pMTL-JH14	This study
ME6	*C. acetobutylicum*	Activation of promoter less *ermB* marker at *thiolase* locus	This study

	ATCC 824 *ermB*+ve	using ACE pMTL-JH16 vector	
ME7	*C. acetobutylicum* ATCC 824 *pyrE*+ve	Activation of promoter less *pyrE* marker at *thiolase* locus using ACE pMTL-JH31 vector	This study
ME8	*C. acetobutylicum* ATCC 824-(L28)	Integration of L28 in *pyrE* mutant strain by activation of promoter less *ermB* marker at *thiolase* locus using ACE pMTL-JH16::l28 vector	This study
ME9	*C. acetobutylicum* ATCC 824-(L28+L12)	Integration of L12 in *pyrE* mutant strain L28 by activation of promoter less *pyrE* marker at *thiolase* locus using ACE pMTL-JH30::L12 vector	This study
ME10	*C. acetobutylicum* ATCC 824-(L28+L12+L6.5)	Integration of L6.5 in strain L28+L12 by activation of promoter less *ermB* marker at *thiolase* locus using ACE pMTL-JH15::L6.5 vector	This study
ME11	Cac824-CaP0098-485a::CT	*C. acetobutylicum* ATCC 824 ClosTron mutant of CaP0098 at 485a	This study
ME12	Cac824-CaP0098-565s::CT	*C. acetobutylicum* ATCC 824 ClosTron mutant of CaP0098 at 565s	This study
ME13	Cac824-CaP0098-1040a::CT	*C. acetobutylicum* ATCC 824 ClosTron mutant of CaP0098 at 1040a	This study
ME14	Cac824-CaP0168-456s::CT	*C. acetobutylicum* ATCC 824 ClosTron mutant of CaP0168 at 456s	This study
ME15	Cac824-CaP0168-657s::CT	*C. acetobutylicum* ATCC 824 ClosTron mutant of CaP0168 at 657s	This study
ME16	Cac824-CaP0168-918s::CT	*C. acetobutylicum* ATCC 824 ClosTron mutant of CaP0168 at 918s	This study
ME17	*C. acetobutylicum*	*C. acetobutylicum* ATCC 824 *pyrE* mutant strain repaired to	This study

	ATCC 824 *pyrE*+ve	wild type sequence using ACE vector pMTL-ME6	
ME18	*C. acetobutylicum* ATCC 824 *pyrE*+ve	*C. acetobutylicum* ATCC 824 *pyrE* mutant strain repaired using ACE vector pMTL-ME7C	This study
3520	*C. acetobutylicum* ATCC 824 (*pyrE*-ve, *spo0A*-ve)	In-frame deletion mutant of *spo0A* made using pMTL-ME3-Cacspo0AKO (*pyrE* vector with KO cassette)	This study
ME19	*C. acetobutylicum* ATCC 824 (*pyrE*+ve, *spo0A*-ve)	In-frame deletion mutant of *spo0A* repaired to wild type at *pyrE* locus using ACE vector pMTL-ME7C	This study
ME20	*C. acetobutylicum* ATCC 824 (*pyrE*+ve, *spo0A*-ve, 85141)	In-frame deletion mutant of *spo0A* repaired to wild type at *pyrE* locus using ACE vector pMTL-ME7C and transformed with pMTL-85141 vector control strain.	This study
ME21	*C. acetobutylicum* ATCC 824 (*pyrE*+ve, *spo0A*-ve, 85141P*spo0A*)	In-frame deletion mutant of *spo0A* mutant strain repaired to wild type at *pyrE* locus using ACE vector pMTL-ME7C and transformed with pMTL-85141-PcacspoOA complementation by vector.	This study
ME22	*C. acetobutylicum* ATCC 824 (*pyrE*+ P*spo0A*+ve)	In-frame deletion mutant of *spo0A* repaired using ACE vector pMTL-ME7C-PCacspo0A (complementation vector for *spo0A*)	This study
ME23	*C. acetobutylicum* ATCC 824 (*pyrE*+ve, *spo0A*-ve)	In-frame deletion mutant of *spo0A* repaired to wild type at *pyrE* locus using ACE vector pMTL-ME6C	This study
ME24	*C. acetobutylicum* ATCC 824 (*pyrE*+ P*spo0A*+ve)	In-frame deletion mutant of *spo0A* repaired using ACE vector pMTL-ME6C-PCacspo0A (complementation vector for *spo0A*)	This study

ME25	*C. acetobutylicum* ATCC 824 (*pyrE*+ve, X*spo0A*+ve)	In-frame deletion mutant of *spo0A* repaired using ACE vector pMTL-ME6X-Cacspo0A (Over-expression vector for *spo0A*)	This study
ME26	*C. acetobutylicum* ATCC 824(*pyrE*-ve, *amyP*-ve)	In-frame deletion mutant of CaP0168 (*amyP*) mutant made using pMTL-ME3-Cap0168KO (*pyrE* vector with KO cassette)	This study
ME27	*C. acetobutylicum* ATCC 824 (*pyrE*+ve, *AmyP*-ve)	In-frame deletion mutant of CaP0168 mutant strain repaired to wild type at *pyrE* locus using ACE vector pMTL-ME7C	This study
ME28	*C. acetobutylicum* ATCC 824 (*pyrE*+ve, *AmyP*-ve, 85141)	In-frame deletion mutant of CaP0168 mutant strain repaired to wild type at *pyrE* locus using ACE vector pMTL-ME7C and transformed with pMTL-85141 vector control strain.	This study
ME29	*C. acetobutylicum* ATCC 824 (*pyrE*+ve, *amyP*-ve, 85141CaP0168)	In-frame deletion mutant of CaP0168 mutant strain repaired to wild type at *pyrE* locus using ACE vector pMTL-ME7C and transformed with pMTL-85141-CaP0168 complementation by vector.	This study
ME30	*C. acetobutylicum* ATCC 824 (*pyrE*+ve, *amyP*)	In-frame deletion mutant of CaP0168 repaired by using ACE vector pMTL-ME7X-CaP0168 (Over-expression vector for CaP0168)	This study
2403	*C. acetobutylicum* ATCC 824 Cac1502-ve	In-frame deletion mutant of type II Restriction system Cac1502 using pMTL-SC7515-Cac1502KO (*codA* vector with KO cassette)	This study
2404	*C. acetobutylicum* ATCC 824 *spo0A*-ve	In-frame deletion mutant of *spo0A* gene made by using pMTL-SC7515-	This study

		Cacspo0AKO (*codA* vector with KO cassette)	

Plasmids/constructs made and used in this study were stored in *E. coli* top10 at -80 °C.

Table 2.2 List of plasmids

Plasmid	Properties	Reference
pAN2	Derived by inserting the ϕ3T I *methyltransferase* gene of pAN-1 into the *catP* gene of pACYC184	(Heap et al., 2007)
pCR2.1-TOPO	T/A cloning vector	Invitrogen Ltd
pMTL80110	*E. coli* cloning vector (p15A, *catP*)	(Heap et al., 2009)
pMTL83251	Shuttle vector (pCB102, ColEI, *ermB*)	(Heap et al., 2009)
pMTL83151	Shuttle vector (pCB102, ColEI, *catP*)	(Heap et al., 2009)
pMTL85141	Shuttle vector (pIM13, ColEI, *catP*)	(Heap et al., 2009)
pMTL83211	Shuttle vector (pCB102, p15A, *ermB*)	This study
pMTL-ME2	Shuttle vector (ColEI, *catP*, *repH*, *C. sporogenes* ATCC 15579 *pyrE*)	This study
pMTL-ME3	Shuttle vector (pIM13, ColEI, *catP*, *sporogenes* ATCC 15579 *pyrE*)	This study
pMTL-JH2	ACE vector *pyrF* internal fragment (300bp) and 1200bp downstream of *pyrF*, ColEI, *catP*, pIM13)	Heap J.
pMTL-JH2::lambda2.0	ACE vector *pyrF* internal fragment (300bp) and 1200bp downstream of *pyrF*, and 2.0Kb lambda DNA ColEI, *catP*, pIM13)	This study
pMTL-JH2::lambda2.3	ACE vector *pyrF* internal fragment (300bp) and 1200bp downstream of *pyrF*, and 2.3 Kb lambda DNA, ColEI,	This study

	catP, pIM13)	
pMTL-JH2::lambda4.3	ACE vector *pyrF* internal fragment (300bp) and 1200bp downstream of *pyrF*, and 4.3 Kb lambda DNA, ColEI, *catP*, pIM13)	This study
pMTL-JH2::lambda6.5	ACE vector *pyrF* internal fragment (300bp) and 1200bp downstream of *pyrF*, and 6.5 Kb lambda DNA, ColEI, *catP*, pIM13)	This study
pMTL-JH12	ACE vector (*C. acetobutylicum* ATCC 824 *pyrE* internal fragment (300bp), ColEI, *catP*, pIM13)	This study
pMTL-JH12::lambda6.5	ACE vector *pyrE* internal fragment (300bp) and 1200bp downstream of *pyrE*, and 6.5 Kb lambda DNA, ColEI, *catP*, pIM13)	This study
pMTL-JH14	ACE vector (for repairing of Δ*pyrE*, ColEI, pIM13, *CatP*)	This study
pMTL-JH14-Pcspfdx	ACE vector (for repairing of Δ*pyrE* including *C. sporogenes fdx* promoter, ColEI, pIM13, *CatP*)	This study
pMTL-JH15	ACE vector containing last 300bp of *thiolase*, *ermB* lacking promoter, *lacZa*, MCS	This study
pMTL-JH16	ACE vector containing last 300bp of *thiolase*, *ermB* lacking promoter, *lacZa*, MCS and 1200b downstream of *thiolase*	This study
pMTL-JH30	ACE vector containing last 300bp of *thiolase*, *C. sporogenes* ATCC 15579 *pyrE* lacking promoter, *lacZa*, MCS	This study
pMTL-JH31	ACE vector containing last 300bp of *thiolase*, *C. sporogenes* ATCC 15579 *pyrE* lacking promoter, *lacZa*, MCS and 1200b downstream of *thiolase*	This study
pMTL-JH16::L28	ACE vector containing last 300bp of *thiolase*, *ermB* lacking promoter, *lacZa*, MCS, lambda L28Kb and 1200b downstream of *thiolase*	Heap J.
pMTL-JH30::L12	ACE vector containing last 300bp of *thiolase*, ATCC 15579 *pyrE* lacking promoter, *lacZa*, MCS and lambda	Heap J.

	L12Kb	
pMTL-JH15::L6.5	ACE vector containing last 300bp of *thiolase*, *ermB* lacking promoter, *lacZa*, MCS and lambda L6.5Kb	Heap J.
pMTL-ME14	ACE vector containing 300bp of SHA and 1200bp of LHA for *amyB* locus, promoter-less *ermB* marker, *catP*, pIM13 replicon, and inverse ACE fragment for *pyrE/hydA* locus	This study
pMTL007C-E2::CaP0098-485a	ClosTron plasmid retargeted to CaP0098-485a	This study
pMTL007C-E2::CaP0098-565s	ClosTron plasmid retargeted to CaP0098-565s	This study
pMTL007C-E2::CaP0098-1040a	ClosTron plasmid retargeted to CaP0098-1040a	This study
pMTL007C-E2::CaP0168-456s	ClosTron plasmid retargeted to CaP1068-456s	This study
pMTL007C-E2::CaP0168-657s	ClosTron plasmid retargeted to CaP1068-657s	This study
pMTL007C-E2::CaP0168-918s	ClosTron plasmid retargeted to CaP1068-918s	This study
pMTL-ME6	ACE vector (for repairing of Δ*pyrE*, ColEI, pIM13, *CatP*)	This study
pMTL-ME6C	ACE vector (for repairing of Δ*pyrE* and complementation of knockout gene, ColEI, pIM13, *CatP*)	This study
pMTL-ME7C	ACE vector (for repairing of Δ*pyrE* and complementation of knockout gene, ColEI, pIM13, *CatP* and transcription terminator from *L. lactis* IL1403 between *pyrE* and *hydA*	This study
pMTL-ME6X	ACE vector (for repairing of Δ*pyrE* and over-expression of knockout gene using *C. sporogenes fdx* promoter, ColEI, pIM13, *CatP*)	This study
pMTL-ME7X	ACE vector (for repairing of Δ*pyrE* and over-expression of knockout gene using *C. sporogenes fdx* promoter, ColEI, pIM13, *catP* and transcription	This study

	terminator from *L. lactis* IL1403 between *pyrE* and *hydA*	
pMTL-ME3::Cac-spo0A-KO-1.5	Shuttle vector (ColEI, *CatP*, pIM13, *C. sporogenes* ATCC 15579 *pyrE*) with *spo0A* KO cassette 1500bp	This study
pMTL-ME3::CaP0168-KO-1.5	Shuttle vector (ColEI, *CatP*, pIM13, *C. sporogenes* ATCC 15579 *pyrE*) with CaP0168-KO cassette 1500bp	This study
pMTL-ME6C-PCacspo0A	ACE vector (for repairing of Δ*pyrE* and complementation of knockout *spo0A* with its promoter, ColEI, pIM13, *catP*)	This study
pMTL-ME6X-Cacspo0A	ACE vector (for repairing of Δ*pyrE* and over-expression of knockout *spo0A* using *C. sporogenes fdx* promoter, ColEI, pIM13, *CatP*)	This study
pMTL-ME6C-PCaP0168	ACE vector (for repairing of Δ*pyrE* and complementation of knockout gene CaP0168 with its promoter, ColEI, pIM13, *CatP*)	This study
pMTL-ME6X-CaP0168	ACE vector (for repairing of Δ*pyrE* and over-expression of knockout gene CaP0168 using *C. sporogenes fdx* promoter, ColEI, pIM13, *CatP*)	This study
pMTL-SC7515	In-frame deletion vector using *E. coli codA* as negative selection marker, *CatP*, pIM13and *traJ* elements	Cartman S
pMTL-SC7515::Cac-spo0A-KO-1.5	In-frame deletion vector using *E. coli codA* as negative selection marker, *CatP*, pIM13, *traJ* elements and KO cassette for *spo0A*	This study
pMTL-SC7515::Cac1502-KO-1.5	In-frame deletion vector using *E. coli codA* as negative selection marker, *CatP*, pIM13, *traJ* elements and KO cassette for type II restriction system Cac1502	This study

Table 2.3 List of oligonucleotides

Oligonucleotide	Sequence 5`→3`	Reference
SbfI-Cac-*pyrF*-iF1	TAATACCTGCAGGGTGTTTAGGACTTGATACT GATATTACTTATGTACCAGAAGAG	Heap J.
Cac-*pyrF*-NotI-R1	ATAAGCGGCCGCTCATTAACCTTCAAAGTGAG CTTTAGCATACATTTCAGCTG	Heap J.
NheI-Cac-*pyrZ*-F1	ATAAAGCTAGCATTTTGGGGGAATTTTGATGA AGGAAAAGTATACAG	Heap J.
AscI-Cac-*pyrD*-iR1	ATTGGCGCGCCTTGTGATACAACATTATAAGC AACTTCTGATTTTATTCCAAAAGCC	Heap J.
Cac-*pyrI*-sF1	TGTGATGAAATATATAAGGGAGCAAAGGCGC	Heap J.
pyrE-LHAv1.0-F1	CCTGCAGGAGAGTAATGTACTTACCTTTGGGG ATTTCATAAC	This study
pyrE-LHAv1.0short-R1	GCGGCCGCCAAGAAGTATTCCCTTATCACCGT GATCTTTAAC	This study
pyrE-LHAv1.0long-R1	TTACTATTTTACTCCATACTCTTTATAGTACTC ATTAATTC	This study
hydA-RHAv1.0-F2	GCTAGCTAAAATAAATGTGCCTCAACTTAGAT GTTAAGGCACATTTATTTTATATATTATTCATG	This study
hydA-RHAv1.0-R2	GGCGCGCCTGTTGCTGCTTTAAAAGAAAAATC CCATATAGAAAAAGTTCAAGAAGC	This study
Cac-thl-LHAv1.0-F1	TAAATTGATATCTATGCAACAAAAGCAGCTAT TGAAAAAGCAGG	This study
Cac-thl-LHAv1.0-R1	CTCCTTCTTAATCGATCTAGCACTTTTCTAGCA ATATTGCTGTTCC	This study
ClaI-*ermB*-F1	GTGCTAGATCGATTAAGAAGGAGTGATTACAT GAACAAAAATATAAAATATTCTC	This study
ermB-AfeI-R1	ATTCTTTAGCGCTTTATTTCCTCCCGTTAAATA ATAGATAACTATTAAAAATAGAC	This study
Cac-DSthl-RHAv1.0-F1	TGCTAGCAAAGTATTGTTAAAAATAACTCTGT AGAATTATAAATTAGTTCTACAG	This study
Cac-DSthl-RHAv1.0-R1	TGGCGCGCCTTCTTTTTATTGCAGTTGCATTTA TTAAAAATGCACTTACTAAAGC	This study
lacZa-sF2	ACTGGCCGTCGTTTTACAACGTCGTG	This study
Cac0026-sF2	TAGCACAATTGTATTTGGACTTCTTTAAATAAA AACATGG	This study
Cac-*hydA*-sR2	TTGATGATGTTTGTCTTGATGACTCAACATGC	This study
ClaI-*pyrE*-F1	TCTTATCGATTAAGGAGAAGATATAAATGAGT AATATAAATGTTATAGATATATTAAAAGAATC	This study

	AAATGC	
pyrE-NotI-R1	AAGATAAGCGGCCGCTTATTTTTGTTCTCTACT ACCTGGTTTTACAAAAGG	This study
M13F	GTAAAACGACGGCCAG	Invitrogen Ltd
M13R	CAGGAAACAGCTATGACC	Invitrogen Ltd
Cac-*atpB*-sR1	ATGATACTGGTATTGTAACCTTTTCTAAAAGGT TCATAGG	This study
Cac-*thl*-sF1	ACTTGCTAAGATAGTTTCTTATGGTTCAGCAGG	This study
L28_F1	CAGAGCCTGTGCACGATTTA	This study
L28_R1	AGTGAAACCGGGTACATTGC	This study
L28_R3	TAAACCGGCATACAGCAACA	This study
L12_F2	CCACCTCTTCCACCATCAGT	This study
L12_R2	CAGGCTTAACCATGCATTCC	This study
L6.5_F2	TGGACGCCAGAAAATTAAGG	This study
L12_R4	TATCAACCTGGTGGTGAGCA	This study
L6.5_R2	GCCGCACAGATGGTTAACTT	This study
SC7-R	AGATCCTTTGATCTTTTCTACGGGGTCTGACGC TCAGTGG	Cartman et al 2012
SC7-F	GACGGATTTCACATTTGCCGTTTTGTAAACGAA TTGCAGG	Cartman et al 2012
Csp-pyrE-HpaI-sF1	AATATTGTTAACTAAGGAGAAGATATAAATGA GTAATATAAATGTTATAGATATATTAAAAGAA TCAAAT	This study
Csp-pyrE-HpaI-sR1	AATATTGTTAACTTATTTTTGTTCTCTACTACCT GGTTTTACAAAAGGT	This study
PmeI-Cac-*spo0A*-L-F2	ATCTTGTTTAAACGAATCTTGCATTTGTAAATG CGCATTACCTAAC	This study
Cac-*spo0A*-L-R2	CCAAGAATAATAGGACATGCTATTGAAGTTGC ATGGTCACGTG	This study
Cac-*spo0A*-R-F1	CAACTTCAATAGCATGTCCTATTATTCTTGGAA GGTTTTCTGC	This study
PmeI-Cac-*spo0A*-R-R1	AAGATGTTTAAACAGGAATGAGTGGAAGTCC	This study
Cac-1501-sF2	GGATTTTTCTCATTTACCCCAGA	This study
Cac-1504-sR1	GCTGTAGTGTAAACTTGTTCTTTG	This study
Cac-*spo0A*-sF2	CTCACCCTTTCTTTCCATCAC	This study

Cac-*spo0A*-sR2	CCTAAAAGTATGATGATTAAAGTTACAGATGC	This study
NotI-Cac*spo0A*-F1	TAAGCGGCCGCAAATTGAGTTTATTAAGTATA ACCCTATATATAGG	This study
EcoRI-Cac*spo0A*-R1	TAATACGAATTCTAACTTAGCTAACTTTATTTT TAAGTC	This study
NdeI-Cac*spo0A*-F1	ATATCATATGGAAAGTAGAAAAATAAGTGTT	This study
Pcac_P-0168-F1	AAAAGCGGCCGCAGATTAAAAATAATTGCAAC AAGTGTGTTGAC	This study
Cac_P-0168-R1	TTTTGGATCCTTAATTATTTGTTGAAGAAAAAT TATAGTTATTACTGT	This study
Cac_P-0168-F2	AAAAATTAATGAGTAAACGTTCTAAATTGTTA AAAAGAAG	This study
PmeI-Cac1502-LF1	ACTGTTTAAACATGCTGAAAAGTTCACTATGTT TG	This study
Cac1502-LR1	GATAAACAGTCTCTACATAGTAACCACCTC	This study
Cac1502-RF1	GAGGTGGTTACTATGTAGAGACTGTTTATCAA G	This study
PmeI-Cac1502-RR1	ACTGTTTAAACTTCACCCCTTAATAAATTTTAG TATTC	This study
27F	AGAGTTTGATCMTGGCTCAG	(Brosius et al., 1978)
1492R	TACGGYTACCTTGTTACGACTT	(Brosius et al., 1978)
Cpa-TT-Oligo-1	CTAGTATAAAAATAAGAAGCCTGCATTTGCAG GCTTCTTATTTTTATG	This study
Cpa-TT-Oligo-2	CTAGCATAAAAATAAGAAGCCTGCAAATGCAG GCTTCTTATTTTTATA	This study
Cac0355-NdeI-sF1	ATATCATATGAATTTTAACTTAGTCACAGCAAC TTCAACATC	This study
Cac0355-EcoRI-sR1	ATATGAATTCCTAGCCAATATTAAGTTTTGTAA TTTCTTC	This study
AM1_2-Fw	TGCTCAACTTGGCACTTACG	This study
AM3_Rev	GTCCCCGCCAACATTTACTA	This study
Pcac_P-0168-F1	AAAAGCGGCCGCAGATTAAAAATAATTGCAAC AAGTGTGTTGAC	This study
Cac_P-0168-R1	TTTTGGATCCTTAATTATTTGTTGAAGAAAAAT TATAGTTATTACTGT	This study
Cac_P-0168-F1	AAAACCCGGGTTGAGTAAACGTTCTAAATTGT TAAAAAGAAG	This study

Cac_P-0168-F2	AAAAATTAATGAGTAAACGTTCTAAATTGTTA AAAAGAAG	This study
CaP0168-Int-sF1	GTTCCCGCATCATAAGTTATC	This study
cac-*amyP*-sF2	GCTCTGGCACAACTTTACGA	This study
cac-*amyP*-sR2	GCACCATTTTGTGATGTGAAA	This study
CaP0168-Int-sF2	TCATATGCGCTTGCTTCAAC	This study
CaP0168-Int-sR2	AACCGGTTTTCGTTTTGATG	This study
RH2_Cap0168_F1	ATATGCTAGCTAAGTGAAAAATCAATGGCTTG TACAATC	This study
RH2_Cap0168_R1	ATATGGCGCGCCATTGCCATAATTTCAAACCT GACTAAG	This study
LH_Cap0168_F2	ATATCCTGCAGGTTAATTGGATATGGACCTGT G	This study
LH_Cap0168_R2	ATATGCGGCCGCATCGTAACTCCCAAGTTGAG A	This study
FRT_ErmB_F1	ATATGCGGCCGCGAAGTTCCTATACTTTCTAGA GAATAGGAACTTCTAAGAAGGAGTGATTACAT GAACAAAAATATAAAATATTCTCAAAAC	This study
FRT_ErmB_R1	TATAGACGCGTGAAGTTCCTATTCTCTAGAAA GTATAGGAACTTCTTATTTCCTCCCGTTAAATA ATAGATAAC	This study
SOE_pyrE_F1	ATATACGCGTCTATTTTACTCCATACTCTTTAT AGTAC	This study
SOE_pyrE_R1	GGAATTCACAATTTTCTCTTCTAGAAAGTAATG TACTTACCTTTGGGGA	This study
SOE_hydA_F1	TCCCCAAAGGTAAGTACATTACTTTCTAGAAG AGAAAATTGTGAATTCC	This study
SOE_hydA_R1	ATATACGCGTTAAAATAAATGTGCCTCAACTT AGATGT	This study
catP-F	TCGGCGCGCCGACTGGGCCGGCCAGTGGGCAA GTTGAAAAATTCAC	(Heap et al., 2009)
catP-R	AGGTTTAAACTTAGGGTAACAAAAAACACCGT ATTTCTAC	(Heap et al., 2009)
ME3-sR2	GAGCGAAGCGAATAAGCGTC	This study
ME3-sF2	ACCGTATTACCGCCTTTGAGTGAG	This study
cac-amyP-sF3	GCTTTGGATACCCGAGCTTT	This study

*Restriction sites are underline in the primers/oligonucleotides

2.1 Microbiological materials and methods

2.1.1 Aerobic bacterial strains and culture conditions

E. coli TOP10 (Invitrogen) or any derivative was grown aerobically in Luria-Bertani (LB) medium or on LB agar (Sambrook et al., 1989) at 37°C with shaking at 200rpm where required. Where appropriate, medium was supplemented with chloramphenicol, 12.5μg/ml in broth and 25μg/ml in agar.

2.1.2 Anaerobic bacterial strains and culture conditions

C. acetobutylicum ATCC 824 was grown in CGM (Hartmanis & Gatenbeck, 1984) or 2xYTG (Harris et al., 2000) or CBM (Obrien & Morris, 1971) broth or on agar at 37 °C in an anaerobic workstation (Don Whitley, Yorkshire, UK) under an atmosphere of N2:H2:CO2 (80:10:10, vol:vol:vol). The media for *Clostridium* strains were pre-reduced overnight. The same anaerobic conditions were used for the new clostridial strains isolated during the course of this project. The media were supplemented with Thiamphenicol 7.5μg/ml in broth and 15μg/ml and erythromycin 10-40μg/ml in solid agar where needed.

2.1.3 Preparation of electro-competent *E. coli* cells

Electrocompetent *E. coli* cells were prepared as follows: 200ml of LB broth was inoculated with 2ml of overnight LB culture *E. coli* (grown with selection if necessary). The cells were incubated at 37 °C with shaking at 200rpm to an OD_{600nm} of 0.5-0.7 (an optical density indicative of exponential growth). The cells were then harvested by chilling on ice for 15-30min followed by centrifugation at 4000xg for 15min at 4°C. The pellet was re-suspended in 200ml of ice-cold sterile deionised water twice and centrifuged as before. The pellet was re-suspended in 4ml of 10% v/v glycerol and centrifuged as before.

The pellet was then re-suspended to a final volume of 0.6-0.8ml in 10% v/v glycerol in water. 50µl of aliquots cells were stored at -80°C for long time use.

2.1.4 Preparation of chemical-competent *E. coli* cells

Chemical competent *E. coli* cells were prepared as follows: 200ml of LB broth was inoculated with 2ml of overnight LB culture *E. coli* (grown with selection if necessary). The cells were incubated at 37°C with shaking at 200rpm to an OD_{600nm} of 0.5-0.7 (an optical density indicative of exponential growth). The cells were then harvested by chilling on ice for 30min followed by centrifugation at 4000x*g* for 6min at 4°C. The pellet was re-suspended in 200ml of ice-cold 0.1M $MgCl_2$ and centrifuged as before. The pellet was re-suspended gently in 200ml ice cold 0.1M $CaCl_2$ and incubated on ice for 20min followed by centrifugation as before. The pellet was re-suspended in 8.6ml of 0.1M $CaCl_2$ and 1.4ml of 100% v/v glycerol was added to it. 200µl aliquots of cells, enough for four transformations, were store at -80°C for future use.

2.1.5 Transformation by electroporation of *E. coli* cells

50µl of the electrocompetent *E. coli* cells prepared in the above section were thawed on ice for 10min and plasmid DNA or dialysed ligation reactions were added. Dialysis was performed using 0.025µm membrane filters (Millipore) by pipetting 20µl of ligation reaction onto the surface of a filter floating on sterile water, and incubation at room temperature for 10min. The *E. coli*/DNA mixture was then transferred to a pre-chilled 2mm gap Electroporation cuvette (BioRad). A pulse of (2.5kV, 25µF, 200Ω) was delivered to the cuvette using a BioRad Gene Pulser according to the manufacturer's instructions. 300µl of S.O.C. medium (Invitrogen) was added immediately and the cell suspension was incubated at 37°C for 30-60min with shaking at 200rpm. Serial dilutions

were made and spread on LB agar plates supplemented with appropriate antibiotics.

2.1.6 Transformation by heat shock of *E. coli* cells

50µl of the chemically competent cells were thawed on ice for 30min and plasmid DNA or ligation reactions were added to them. The mixture was then incubated on ice for 30min followed by heat-shock at 42°C in a water bath for 2min. The mixture was then incubated on ice for 1 min followed by addition of 1ml of LB broth. The reaction was then incubated at 37°C with shaking at 200rpm for 30-60min. The suspension was centrifuged at high speed 20000xg for 1min and supernatant was discarded. The pellet was re-suspended in 200µl of LB and spread on LB plates supplemented with appropriate antibiotics.

2.1.7 Transformation of *C. acetobutylicum* ATCC 824

This method is described by (Mermelstein & Papoutsakis, 1993). *C. acetobutylicum* ATCC 824 was streaked onto a CGM agar plate and incubated for 24h. Inocula were taken directly from -80 frozen stocks or alternatively, a spore stock was heat shocked at 80°C for 10min in a water bath before plate inoculation. 10ml of 2xYTG pH5.2 or CGM broth was then inoculated with a loop of fresh *C. acetobutylicum* ATCC 824 cells, mixed by vortexing and serially diluted, 10^{-1}, 10^{-2}, and 10^{-3} into fresh medium before incubation overnight at 37°C in an anaerobic workstation. The culture inoculated with the lowest inoculum which showed growth, was then subcultured into 60ml fresh 2xYTG pH5.2 (pre-reduced overnight) followed by incubation at 37°C anaerobically until an OD_{600nm} of 1.1 was reached. The culture was then harvested by centrifugation at 4°C for 10min at 4000x*g*. Supernatant was discarded and the pellet was re-suspended in 10ml of ice cold Electroporation

buffer (EPB; 284mM sucrose, 100mM sodium phosphate, pH 7.4) (pre-reduced overnight) by vigorous vortexing followed by centrifugation as before. The pellet was re-suspended in 2.3ml of ice cold EPB. 4-6μg of methylated (as appropriate) plasmid DNA was added to the pre-chilled 0.4cm gap Electroporation cuvette and 570μl of the competent cells were also added and incubated for 2min on ice. The cuvette was dried and Electroporation was done using 2.0kV, 25μF and ∞Ω. 1ml of warm 2xYTG was added immediately and incubated for 1-3hr at 37°C anaerobically. Cells were harvested by centrifugation and cell pellet resuspended in 500μl of 2xYTG and plated onto appropriate plates.

2.2 Molecular biological materials and methods

2.2.1 Plasmid DNA extraction and purification

Plasmid DNA from *E. coli* was isolated by alkaline lysis method using QIAprep Spin Miniprep (Qiagen) or Zyppy™ Plasmid Miniprep Kits according to manufacturer's instructions.

2.2.2 Extraction and purification of chromosomal DNA

Chromosomal DNA was extracted using the Qiagen DNeasy® Blood & Tissue Kit according to the manufacturer's instructions. Gram-positive cells were first treated with lysozyme (20mg/ml in phosphate-buffered saline [PBS]) at 37°C for 30min.

2.2.3 Spectrophotometric quantification of DNA

DNA present within a sample was quantified using a Nanodrop ND-1000 spectrophotometer. The Nanodrop measures the OD_{260} of the sample given that

with a 1 cm path, a 50μg/ml solution of double-stranded DNA has an OD_{260} of 1.0.

2.2.4 Polymerase Chain Reaction (PCR)

Amplification of DNA for screening was carried out using NEB BioLabs, (UK) Taq DNA polymerase with Thermopole buffer in accordance with the manufacturer's instructions. For cloning of fragments below 3Kb, a high fidelity Failsafe polymerase was used from Epicentre in conjunction with Failsafe Premix E buffer, and for fragments over 3Kb, Novagen KOD Hot Start DNA polymerase system was used in accordance with the manufacturer's instructions.

A typical PCR reaction was carried out using initial denaturation at 95-98°C for 2 min for 1 cycle, followed by 25-30cycles of denaturation at 95-98°C for 30sec, annealing (Tm-5) for 30sec and extension at 72°C for 30sec to 1min for each Kb of DNA and a final extension at 72°C for 5-10min.

2.2.5 Splicing Overlap Extension (SOE) PCR

Splicing by Overlap Extension PCR was carried out to splice two DNA fragments together using appropriate primers in a two-step reaction mixture, without the use of restriction endonucleases (Horton et al., 1990; Horton et al., 1989).

2.2.6 Restriction endonuclease digestion of DNA

Restriction digests were carried out using enzymes from New England Biolabs or Promega according to the manufacturer's instructions. A typical reaction of 50μl contained 1X buffer, bovine serum albumin (1μg/ml final concentration)

was added if required, 5-10U of enzyme, DNA 0.5-1.0µg and was incubated at 37°C for 1hr.

2.2.7 Blunting of DNA ends by T-4 DNA polymerase

Many restriction enzymes produce DNA ends with either 3` or 5` overhanging ends. T4 DNA polymerase from NEB Biolabs was used which catalyzes the synthesis of the DNA in the 5` to 3` direction and has exonuclease activity in the 3` to 5` direction. It fills in the 5` overhang and removes the 3`overhang to make blunt ends. A typical reaction consisted of 1x any NEB buffer, 5U of T4 DNA polymerase and a mixture of 100µM each dNTP (Promega) followed by incubation at 37°C for 30min.

2.2.8 De-phosphorylation of DNA using antarctic phosphatase

Linearised DNA was dephosphorylated to avoid self-ligation using Antarctic Phosphatase (AP) from NEB Biolabs according to the manufacturer's instructions. Dephosphorylation was carried out in a volume of 20µl with 1X AP buffer and 5U of AP enzyme followed by incubation at 37°C for 30-60min.

2.2.9 Ligation of DNA using T-4 DNA ligase

Ligation of DNA fragments was performed using T4 DNA ligase from NEB Biolabs or Promega according to the manufacturer's instructions. Reaction was performed in a volume of 20µl; vector and insert were added in ratios of 1:1 or 1:3, 1X T4 DNA ligase buffer and 10U of T4 DNA ligase followed by incubation at 4°C overnight.

2.2.10 Agarose gel electrophoresis

PCR products, restriction fragments and plasmid DNA were separated by electrophoresis at 90V for 40-60min (depending on the size of expected

fragment) through 0.8% w/v agarose gels in TAE containing 0.5μg/ml ethidium bromide. DNA was subsequently visualised under ultra-violet light.

2.2.11 Extraction of DNA from agarose gels

DNA fragments were visualised under UV light using ethidium bromide and the desired band was excised from the agarose gel using a clean scalpel. The DNA was then extracted and purified using Qiaquick Gel Extraction Kit (Qiagen Ltd UK) according to the manufacturer's instructions.

2.2.12 Purification of DNA from PCR reaction mixtures

DNA was purified from reaction mixtures using the Qiagen QIAquick PCR Purification Kit in accordance with the manufacturer's instructions.

2.2.13 Annealing pairs of oligonucleotides

Synthetic DNA inserts with specified cohesive ends were generated by annealing primers which were complementary to each other. These cohesive ends are designed into the primers so the desired over/under-hang is present to use directly for ligating into a plasmid backbone. Equal amount of oligonucleotide pair (100μM) were added to a PCR tube and diluted with Elution Buffer (EB) (Qiagen) to a ratio of 1:100 to yield 1μM of each oligonucleotide. This mixture was then heated in a PCR thermo cycler to 95°C for 30sec and allowed to cool to room temperature at a rate of 1°C per second. This reaction anneals the two oligonucleotides together, allowing the resultant insert to be used directly in a ligation reaction.

2.2.14 Southern blotting

Southern blot analysis was performed using lambda DNA as probe to verify the lambda phage DNA insertions using five micrograms of genomic DNA

digested with EcoRI (NEB) overnight. The blot was carried out using a DIG high prime labelling and detection kit (Roche) according to the manufacturer's instructions.

2.2.15 Sporulation assay

Spore stocks of test strains were heat treated at 80 °C for 10 min before plating on pre-reduced CBM agar and incubation for 24 hrs.

Triplicate pre-cultures were prepared by resuspending a loopful of colonies from each strain into three universal tubes containing 10ml CBMS medium (CBM + 0.5% w/v $CaCO_3$, 5% w/v glucose) and incubated overnight.

The following day, triplicate sporulation assay cultures (30ml CBMS medium in 50ml-Falcon tubes) were inoculated to an initial OD_{600nm} 0.048-0.0065 from the pre-cultures. 200µl samples were removed at 0hr, and after incubation for 24, 48, 72, 96, and 120hrs. Samples were heated at 80 °C for 10 min in a water bath and dilutions (10^{1}-$10^{-6)}$ were prepared in 96 well microtitre. Triplicate plates 20µl volumes were spotted onto CBM agar medium and plates incubated for 48hrs. The number of colonies was counted and Heat resistant Colony Forming Units per millilitre (CFU/ml) was calculated (Ann-Kathrin Kotte PhD student CRG 2013, Personal communication).

2.2.16 Gram staining

Gram staining was performed as previously described (Sambrook et al., 1989).

2.2.17 Iodine assay for halo detection

Amylolytic activity (halos formation) of *C. acetobutylicum* ATCC 824 on CBM medium containing 2% w/v starch and 0.5% w/v glucose after incubation for 48hrs, was detected by staining with iodine (Sabathe et al., 2002).

2.3 Bioinformatics methods

2.3.1 DNA sequencing

DNA sequencing was performed by Source Bioscience sequencing facility in Nottingham.

2.3.2 Sequence editor and plasmid map designer

In silico vector maps were designed using a combination of Vector NTI (http://www.invitrogen.com) and GENtle (http://gentle.magnusmanske.de).

2.3.3 Analysis of the sequencing results

Construction and analysis of contigs from sequencing results was performed using sequencing assistant features in GENtle.

2.3.4 Designing of oligonucleotides or primers

PCR primers or synthetic DNA inserts were designed manually or using Primer3 available at http://frodo.wi.mit.edu/primer3/ or Oligo Analysis Tool available at http://www.operon.com/tools/oligo-analysis-tool.aspx.

2.4 Isolation of new clostridial species

2.4.1 Screening of new clostridial species for biofuel production

Samples from soil, degraded plant material, and composts were collected in small sterile universal tubes with a sterile scoop from the University of Nottingham campuses and canal sides.

2.4.2 Screening for solvent producing clostridial strain

1g of soil degraded plant material or compost was re-suspended in 10ml of pre-reduced RCM and heated at 70°C for 10min to kill all the vegetative cells. The media containing soil samples were then incubated for 2-3 days at 37°C in an anaerobic workstation as described above. The cultures were checked each day for growth and gas production. 50-100µl of the liquid culture was spread on T6 agar (per liter 6 g of tryptone, 2g of yeast extract, 0.5g of KH_2PO_4, 0.3g of $MgSO_4.7H_2O$, 10mg of $FeSO_4.7H_2O$, 3g of ammonium acetate, and 0.5g of cysteine hydrochloride. The pH was adjusted to 6.5 with NaOH (Kashket & Cao, 1993). Single colonies were inoculated in 2ml T6 broth to screen for acetone production by Rothera's test using one small drop of 5% (w/v) sodium nitroprusiate and 25% (w/v) ammonium hydroxide. In the presence of acetone, a permanganate colour ring was formed (Rothera, 1908). Acetone producing colonies were then checked for spore formation by phase contrast microscopy.

2.4.3 Substrate utilization such as xylose, avicel and CMC

The soil, degraded plant material or composts were treated as described above in CBM broth with 2% w/v xylose, 1% w/v avicel and 1% w/v carboxymethyl cellulose (CMC) instead of glucose followed by incubation for 3 days at 37°C in an anaerobic workstation. The cultures were checked daily for growth and

gas production. 50-100µl of the cultures was spread on CBMx and CBM_{CMC} agar containing xylose (x) and carboxymethyl cellulose (CMC) as a carbon source. Pure cultures were obtained by re-streaking single colonies twice on the same media. Single colonies were then inoculated in CBM broth with xylose or avicel to check for growth. Strains isolated from this stage were then screened for 16S rRNA sequences.

2.4.4 Genetic identification of screened microorganisms

Strains isolated in the above step were then inoculated in 1ml of CBM with 2% w/v xylose or 1% w/v avicel and incubated for 24hrs at 37°C in an anaerobic workstation as described above. Bacteria were pelleted by centrifugation of the cultures (21000x*g* for 5 min) and total genomic DNA was prepared using Qiagen DNeasy Blood & Tissue Kit according to the protocol described by the manufacturer. The 16S rRNA gene ca. 1500bp, was amplified by PCR with two universal primers 27F and 1492R (Brosius et al., 1978). The amplified bands were purified from an agarose gel using a Qiaquick gel extraction kit (Qiagen) according to the manufacturer's instructions and sequenced using the same primers. A variety of species were identified using the nucleotide sequence BLAST search engine available at http://blast.ncbi.nlm.nih.gov.

2.4.5 Microscopic observation of new isolates

1-2µl of the growing culture was spotted on a microscope slide and covered with a cover slip. Bacteria were visualised using 10X, 40X and 100X lenses.

2.5 Plasmid construction

2.5.1 Allele Coupled Exchange (ACE) vectors

C. acetobutylicum ATCC 824 shuttle plasmid pMTL85141 (Heap et al., 2009) was used to construct all the Allele Coupled Exchange (ACE) vectors. These differ only in target locus-specific Short Homology Arm (SHA) and Long homology Arm (LHA) sequences inserted between the SbfI/NotI and NheI/AscI restriction sites of the plasmids. All the ACE vectors shared commonly used features such as the chloramphenicol/thiamphenicol-resistance marker *catP* and the origin of replication from *Bacillus subtilis* plasmid pIM13 (Monod et al., 1986). The pIM13 replicon exhibits segregational instability in *C. acetobutylicum* (Heap et al., 2009) and single-crossover clones which spontaneously arise within transformant populations can be isolated by subculture under antibiotic selection.

2.5.1.1 pMTL-JH2 vector

To construct an Allele Coupled Exchange (ACE) vector pMTL-JH2 for the *C. acetobutylicum* ATCC 824 *pyrF* locus, shuttle plasmid pMTL85141 (Heap et al., 2009) was used. An internal fragment of 300bp (Short Homology Arm, SHA) of *C. acetobutylicum* ATCC 824 *pyrF* was PCR amplified using primers SbfI-Cac-*pyrF*-iF1 and Cac-*pyrF*-NotI-R1 and double digested with restriction enzymes SbfI and NotI and ligated in the same sites of pMTL85141 to make intermediate vector pMTL-JH1 which was verified by restriction digestion and nucleotide sequencing using primers SbfI-Cac-*pyrF*-iF1 and Cac-*pyrF*-NotI-R1. Another fragment of 1200bp immediately downstream of *pyrF* (Long Homology Arm, LHA) was PCR amplified with primers NheI-Cac-*pyrZ*-F1

and AscI-Cac-*pyrD*-iR1 and subsequently double digested with restriction enzymes NheI and AscI and ligated in equivalent sites of pMTL-JH1 to make the final vector pMTL-JH2. It was verified by restriction digestion and nucleotide sequencing using primers NheI-Cac-*pyrZ*-F1 and AscI-Cac-*pyrD*-iR1 and used to make *C. acetobutylicum* ATCC 824 *pyrF* mutant strain (Data obtained from John Heap).

2.5.1.2 Derivatives of pMTL-JH2 vector with lambda DNA

To insert various sizes of lambda DNA fragments at the *C. acetobutylicum* ATCC 824 *pyrF* locus, four derivatives of ACE vector pMTL-JH2 were constructed by inserting various sized fragments of phage lambda DNA into the multiple cloning sites (MCS) between two allele exchange cassettes. Phage lambda (cI857ind1 Sam 7) DNA digested to completion with HindIII was purchased from NEB BioLab. To separate the lambda cohesive ends of the fragments derived from the ends of the lambda chromosome, mixture of fragments was heated at 60°C for 5min followed by treating with T4 polymerase (NEB) to blunt-end the HindIII and lambda cohesive ends to be able to ligate them into the StuI site available in MCS between the two homology regions in pMTL-JH2 vector.

Blunt-ended lambda DNA fragments 2.0, 2.3, 4.3 and 6.5Kb were isolated by agarose gel electrophoresis, purified and cloned in pMTL-JH2 DNA linearised with StuI and treated with Antarctic Phosphatase (NEB) to prevent self-ligation. This positioned the lambda DNA in each pMTLJH2 derivative between the two regions of homology so that the lambda DNA would be delivered to the chromosome of a double-crossover clone by a successful integration event. These plasmids were designated pMTL-JH2-lambda2.0,

pMTL-JH2-lambda2.3, pMTL-JH2-lambda4.3 and pMTL-JH2-lambda6.5 accordingly. The derivatives of pMTL-JH2 were each sequenced using the M13F primer to determine the orientation of the insert (because blunt cloning is not directional) which would subsequently inform the design of the screening PCR.

2.5.1.3 pMTL-JH12 vector

To construct an Allele Coupled Exchange (ACE) vector pMTL-JH12 for *C. acetobutylicum* ATCC 824 *pyrE* locus, shuttle plasmid pMTL85141 (Heap et al., 2009) was used. An internal fragment of 300bp (Short Homology Arm, SHA) of *C. acetobutylicum* ATCC 824 *pyrE,* which lacks the 13 codons from 5` and 112 codons from 3`end, was PCR amplified using primers *pyrE*-LHAv1.0-F1 and *pyrE*-LHAv1.0short-R1and double digested with restriction enzymes SbfI and NotI and ligated in the same sites of pMTL85141 to make intermediate vector pMTL-JH11 verified by restriction digestion and nucleotide sequencing using primers *pyrE*-LHAv1.0-F1 and *pyrE*-LHAv1.0short-R1. Another fragment of 1200bp immediately downstream of *pyrE* (Long Homology Arm, LHA) was PCR amplified with primers hydA-RHAv1.0-F2 and hydA-RHAv1.0-R2 and subsequently double digested with restriction enzymes NheI and AscI and ligated in equivalent sites of pMTL-JH11 to make final vector pMTL-JH12 verified by restriction digestion and nucleotide sequencing using primers hydA-RHAv1.0-F2 and hydA-RHAv1.0-R2. pMTL-JH12 was used to make a *C. acetobutylicum* ATCC 824 *pyrE* mutant strain.

2.5.1.4 pMTL-JH14 vector

To construct an Allele Coupled Exchange (ACE) vector pMTL-JH14 to repair *C. acetobutylicum* ATCC 824 *pyrE* mutant back to wild type, ACE vector pMTL-JH12 was used which is almost identical to pMTL-JH14, except for SHA between SbfI and NotI sites. *C. acetobutylicum* ATCC 824 *pyrE* (Cac0027) was PCR amplified with primers *pyrE*-LHAv1.0-F1and *pyrE*-LHAv1.0long-R1 and double digested with SbfI and NotI and ligated in pMTL-JH12 linearised with same set of enzymes to make pMTL-JH14 which was used to repair *pyrE* mutant back to wild type. The SHA in pMTL-JH14 lacks the same initial codons as in pMTL-JH12 but includes the rest of the gene up to the stop codon. The final vector was verified by restriction digestion and nucleotide sequencing using primers *pyrE*-LHAv1.0-F1and *pyrE*-LHAv1.0long-R1 before being used to couple the two non-functional alleles to make a functional *pyrE* gene.

2.5.1.5 pMTL-ME6 vector

In the previous section, ACE plasmid pMTL-JH14 was constructed to join two non-functional fragments of *pyrE ie.*, one on plasmid and the other on host chromosome after the double crossover integration event. It also integrates sequence of *lacZalpha* downstream of *pyrE* distinguishing it from the wild type in terms of genotype. The *pyrE* mutant strain could be reverted back to wild-type in terms of phenotype, as well as genotype using ACE. This was accomplished using the vector pMTL-ME6, which was made as follows:

A fragment of 1850bp, encompassing regions of *pyrE* and *hydA* genes, was cloned by PCR using primers *pyrE*-LHAv1.0-F1 and hydA-RHAv1.0-R1 from *C. acetobutylicum* genomic DNA. This PCR product was digested with HindIII

and BglII and a 939bp fragment was purified from agarose gel and ligated in pMTL-JH14 linearised with the same restriction enzymes. Prior to ligation of the linearised vector with the isolated PCR fragment, it was treated with antarctic phosphatase to prevent the empty vector from re-circularisation. The final vector made was called pMTL-ME6.

2.5.1.6 pMTL-ME7C and pMTL-ME7X vectors

To construct chromosomal complementation and over-expression vectors for *C. acetobutylicum pyrE* locus, *C. sporogenes ferredoxin (fdx)* promoter was isolated from plasmid pMTL82253 (Heap et al., 2009) as a 214bp NotI/NdeI fragment and inserted between the equivalent sites of pMTL-JH14 (Figure 2.1). The plasmid obtained pMTL-JH14::Pcspfdx was cleaved with NheI and treated with antarctic phosphatase to prevent empty vector re-circularisation. The linearised vector was ligated to two, annealed oligonucleotides (Cpa-TT-Oligo-1 and Cpa-TT-Oligo-2), which encompassed the transcriptional terminator of the *C. pasteurianum ferredoxin,* to create pMTL-ME6X. These annealed oligonucleotides contained over-hangs of restriction site BfaI that is compatible to NheI site. The recombinant plasmids obtained were screened by sequencing, and a clone selected in which the orientation of insertion of the annealed oligonucleotide-derived fragment was such that the NheI site created, resided at the 3`end of the terminator.

A 293bp fragment was released from pMTL-ME6X with EcoRI and NheI and inserted between the equivalent sites of pMTL-JH14 to make pMTL-ME6C. These two vectors pMTL-ME6C and pMTL-ME6X could be used to complement and over-express a deleted gene respectively through their integration at the *pyrE* locus in a mutant strain. In both cases, however, the

promoter responsible for *pyrE* expression could also cause transcriptional read-through of the inserted gene.

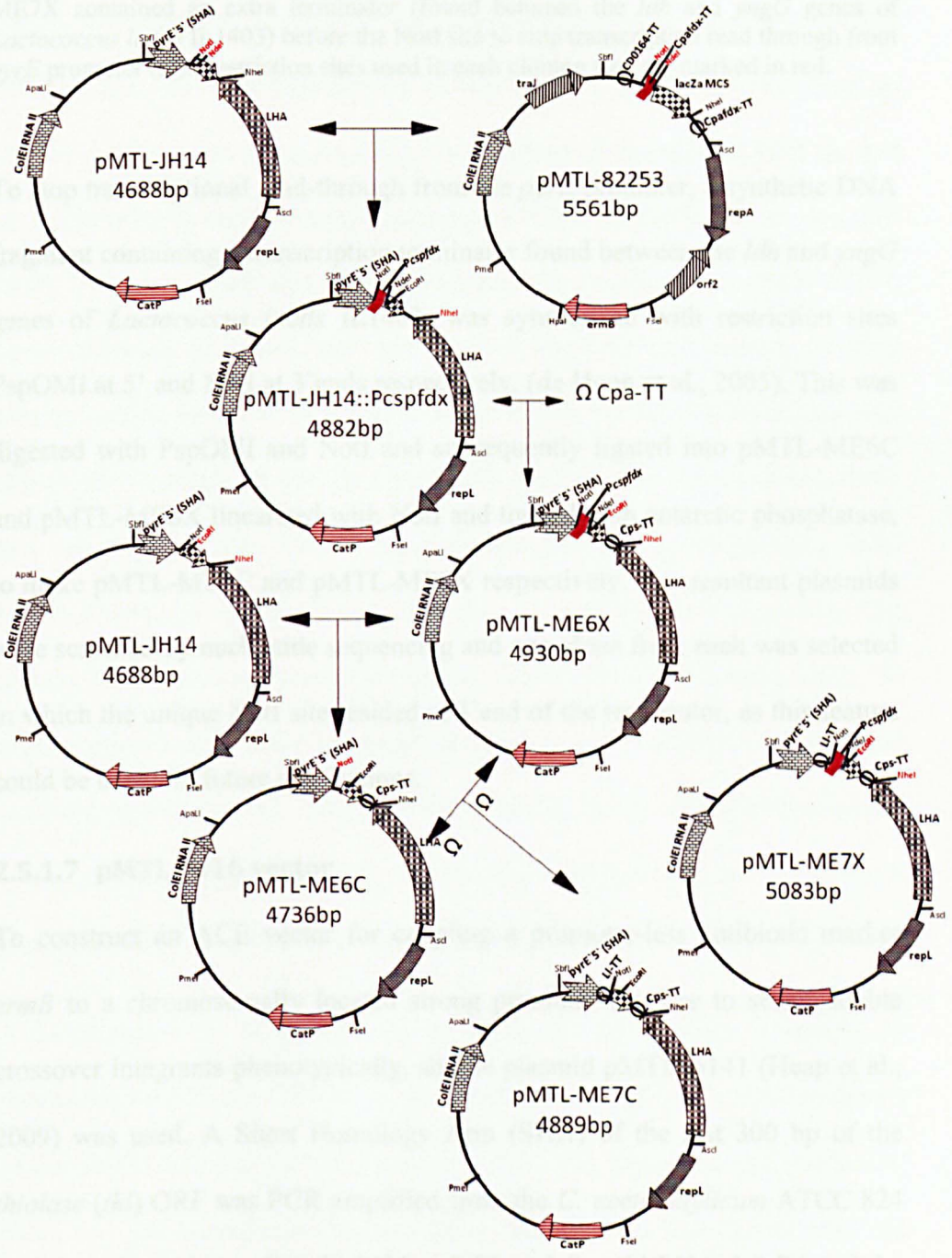


Figure 2.1 Construction of pMTL-ME7C and pMTL-ME7X vectors

The regions between NheI and SbfI in the two complementation vectors pMTL-ME6C and pMTL-ME7C and two chromosomal over-expression vectors pMTL-ME6X and pMTL-ME7X are the same consisting of long homology arm (LHA) 1200bp, homology to the downstream region of *C. acetobutylicum pyrE* locus, Gram-positive pIM13 replicon *repL*, *catP* antibiotic marker confering resistance to chloramphenicol/Thiamphenicol, Gram-negative replicon ColE1 RNA II and a

terminator from *C. pasteurianum fdx* gene before the NheI site (Ω Cpa-TT). The over-expression vectors contained the strong *fdx* promoter between NotI and NdeI sites. All four vectors carry a short homology arm (SHA) of ca. 600bp homologous to the *C. acetobutylicum pyrE* gene between SbfI and NotI while pMTL-ME7C and pMTL-ME7X contained an extra terminator (found between the *ldh* and *yngG* genes of *Lactococcus lactis* IL1403) before the NotI site to stop transcription read through from *pyrE* promoter (Ω). Restriction sites used in each cloning step are marked in red.

To stop transcriptional read-through from the *pyrE* promoter, a synthetic DNA fragment containing a transcription terminator found between the *ldh* and *yngG* genes of *Lactococcus lactis* IL1403, was synthesized with restriction sites PspOMI at 5' and NotI at 3'ends respectively. (de Hoon et al., 2005). This was digested with PspOMI and NotI and subsequently ligated into pMTL-ME6C and pMTL-ME6X linearised with NotI and treated with antarctic phosphatase, to make pMTL-ME7C and pMTL-ME7X respectively. The resultant plasmids were screened by nucleotide sequencing and one clone from each was selected in which the unique NotI site resided at 3`end of the terminator, as this feature could be useful in future applications.

2.5.1.7 pMTL-JH16 vector

To construct an ACE vector for coupling a promoter-less antibiotic marker *ermB* to a chromosomally located strong promoter *thiolase* to select double crossover integrants phenotypically, shuttle plasmid pMTL85141 (Heap et al., 2009) was used. A Short Homology Arm (SHA) of the last 300 bp of the *thiolase* (*thl*) ORF was PCR amplified from the *C. acetobutylicum* ATCC 824 genome using primers Cac-thl-LHAv1.0-F1 and Cac-thl-LHAv1.0-R1 and the *ermB* ORF with its ribosome-binding site (RBS) was PCR-amplified using pMTL21E as DNA template (Swinfield et al., 1990) with primers ClaI-*ermB*-F1 and *ermB*-AfeI-R1 and purified from an agarose gel.

Primers Cac-thi-LHAv1.0-R1 and ClaI-*ermB*-F1 were designed to produce a region of sequence identity between the 3'end of the *thl* PCR product and the 5'end of the *ermB* PCR product, which allowed fusion of these two purified PCR products by a splicing overlap extension (SOE) PCR with primers Cac-thi-LHAv1.0-F1 and *ermB*-AfeI-R1. The resultant *thl-ermB* SOE PCR product was digested with restriction endonucleases EcoRV and AfeI, (which generated blunt DNA ends) and this was ligated into the shuttle vector pMTL85141, linearised with NotI and blunt ended with T4 DNA polymerase; yielding plasmid pMTL-JH15. This strategy was designed to regenerate a single NotI site in pMTL-JH15, to the right-hand (3') side of the *thl-ermB* sequence; and also to retain the transcriptional terminator, originating from pMTL85141, to the left-hand (5') side of the *thl-ermB* sequence. The unique NotI site may prove useful in future cloning strategies, and the transcriptional terminator prevents or reduces undesirable transcriptional read-through into *ermB*, either on the plasmid, or in plasmid derivatives, or in other derivatives, such as a co-integrate molecule formed by a plasmid and the chromosome during an allelic exchange procedure.

A Long region of homology corresponding to the 1200bp immediately downstream of the *thl* ORF of *C. acetobutylicum* ATCC 824, which consists of Cac2872 ORF and part of the *atpB* ORF, was designed and purchased from the DNA synthesis company DNA 2.0 Inc, and sub-cloned using the restriction endonucleases NheI and AscI into pMTL-JH15 to generate pMTL-JH16. This vector or any derivative of it could be used to express any selectable marker at the *thl* locus.

2.5.1.8 pMTL-JH31 vector

In the preceding section ACE vector pMTL-JH16 was constructed to deliver cargo DNA downstream of the *thl* locus using ACE by selecting double crossover integrants by activating a promoter-less *ermB* marker. To test another marker, the orthologous *pyrE* gene from *C. sporogenes* ATCC 15579 was selected which is only 47.6% identical to the *C. acetobutylicum pyrE* and is therefore unlikely to recombine with the native *pyrE* gene by homologous recombination. This gene was PCR amplified from the *C. sporogenes* ATCC 15579 genome with primers ClaI-*pyrE*-F1 (contains RBS and the start codon of *pyrE*) and *pyrE*-NotI-R1 and the PCR product generated digested with ClaI and NotI and ligated into the equivalent sites in pMTL-JH15 to create pMTL-JH30. The authenticity of the plasmid made was verified by restriction digestion and nucleotide sequencing. A Long homology region of 1200bp was ligated in the same way as described for pMTL-JH16 to make pMTL-JH31. Vectors pMTL-JH30 and pMTL-JH31differ from pMTL-JH15 and pMTL-JH16 respectively only in the regions between ClaI and NotI restriction sites where the ribosome-binding site (RBS) and coding sequence of *ermB* is replaced by the RBS of *pyrE*. Plasmid pMTL-JH31 can be used in situations where ACE involving the use of pMTL-JH16 had already been implemented and *ermB* inserted into the chromosome at the *thl* locus along with cargo DNA. To insert further DNA in a second iteration of ACE, pMTL-JH31 could be used, when its promoter-less *pyrE* gene would become activated by the upstream *thl* locus. Its use was only possible in a *pyrE* mutant of *C. acetobutylicum*, created using pMTL-JH12.

2.5.1.9 Iterative ACE (iACE) vectors

To insert the entire genome of phage lambda into the chromosome of *C. acetobutylicum pyrE* showing iteration of the ACE method, three unique restriction fragments were identified in a circular form of phage lambda DNA (circularized by ligation of the cohesive ends of the linear chromosome). These fragments contained suitable overlapping regions for first homologous recombination events (*ie.*, via the long region of homology) in the second and third ACE steps during this procedure as shown in Figure 3.16 . The three restriction fragments were 28Kb (XmaI-XmaI) fragment L28, 18Kb (NheI-AscI) fragment L18 and 6.5Kb (XmaI-XhoI) fragment L6.5 and these were ligated into ACE vectors pMTL-JH16, pMTL-JH30 and pMTL-JH15, respectively to make pMTL-JH16::L28, pMTL-JH30::L12 and pMTL-JH15::L6.5 vectors respectively. Both vectors pMTL-JH30 and pMTL-JH15 contain short regions of homology for the *thiolase* locus but do not include a long region of homology and therefore are appropriate when the long region of homology depends upon a previous step of allelic exchange, and must therefore be provided as part of the insert in second or third integration step. The vectors pMTL-JH30 and pMTL-JH15 also contain promoter-less markers *C. sporogenes* ATCC 15579 *pyrE* and *ermB* respectively which are used for the selection of the double crossover integrants. The Lambda 28Kb fragment (L28) was isolated as an XmaI-XmaI fragment (between base pairs 39888-19397), and ligated into pMTL-JH16, linearised with XmaI and treated with antarctic phosphatase. The resulting plasmid pMTL-JH16::L28, was verified by restriction digestion and nucleotide sequencing. The 18Kb Lambda DNA fragment (L18) was isolated as a NheI-AscI fragment (between base pairs

16649-34679) and cloned into pMTL-JH30. Subsequent characterisation of the recombinant plasmid obtained revealed that a spontaneous deletion of ~6Kb of DNA from within the cloned lambda DNA insert had occurred. The cloned Lambda-derived region of DNA (between base pairs 21830-27732) contains genes involved in lysis of the host, which may have proven to be toxic to the *E. coli* host. However, as the region required for homologous recombination (XmaI-AscI; fragment between base pairs 16649-19397) with the L28Kb was not affected, the plasmid obtained was used in the further experiments, and the plasmid designated pMTL-JH30::L12. The last lambda fragment L6.5 was isolated as an XhoI-XmaI (between base pairs 33498-39888) and ligated in appropriate restriction sites in pMTL-JH15 to make pMTL-JH15::L6.5. This fragment included the region required for homologous recombination (between base pairs 33498-34679) with the second Lambda fragment L12Kb for the integration of the plasmid via long homology arm in the final round of the envisaged multistep ACE procedure (Data obtained from John Heap).

2.5.1.10 Attempted integration of pSOL1 into the chromosome using pMTL-ME14

To attempt to integrate pSOL1 into the *C. acetobutylicum pyrE* mutant chromosome through the use of reverse ACE (rACE), a special ACE vector pMTL-ME14 was constructed to deliver the inverse sequence of *pyrE-hydA* at the amylase locus. A 1200bp long homology arm was PCR amplified from downstream of the *amylase* gene CaP0168 including stop codon from *C. acetobutylicum* DNA using primers RH2_Cap0168_F1 and RH2_Cap0168_R1 and 300bp of internal fragment (after the 20 codons from the start of the gene) short homology arm using primers LH_Cap0168_F2 and LH_Cap0168_R2.

These PCR products were digested with restriction enzymes NheI/AscI and SbfI/NotI respectively and ligated into modular vector pMTL85141 at the equivalent restriction sites to generate pMTL85141AF. MLS marker *ermB* was PCR amplified from pMTL-JH16 using primers FRT_ErmB_F1 and FRT_ErmB_R1. FRT sites were introduced flanking the marker so that it could be removed and used later on if needed. The PCR product was digested with NotI/MluI and ligated in appropriate sites in pMTL85141AF to make pMTL85141AFB and verified by restriction analysis and nucleotide sequencing. Since larger sequences could be delivered using ACE method, an even bigger *pyrE-hydA* construct was made by SOE PCR which consists of ca. 1500bp of *hydA* and 633bp of *pyrE* gene in an inverse orientation. The PCR products:- *pyrE* from primers SOE_pyrE_F1 and SOE_pyrE_R1 and *hydA* from primers SOE_hydA_F1 and SOE_hydA_R1 were used as templates in SOE PCR with flanking primers SOE_pyrE_F1 and SOE_hydA_R1 to make *pyrE-hydA* flanked with MluI sites. It was digested with MluI and ligated in pMTL85141AFB linearised with same restriction enzymes to make the final vector pMTL-ME14.

2.5.2 pMTL007C-E2-ClosTron vectors

To make insertional mutants of various genes in *C. acetobutylicum*, ClosTron technology was used. To design re-targeted group II introns, the nucleotide sequences of the target genes were submitted into Perutka algorithm93 available at http://clostron.com/clostron2.php. The algorithm identified some possible target sites within the sequence and the re-targeted sequences were synthesised by DNA2.0 as HindIII and BsrG I fragments of ca. 353bp and cloned into same sites of ClosTron vector pMTL007C-E2 (Figure 1.3), and

delivered ready to use. The resultant re-targeted plasmids were transformed into *E. coli* strain pAN2 by electroporation and plated onto LB agar supplemented with chloramphenicol 25μg/ml and tetracycline 10μg/ml to protect DNA from the type II restriction endonuclease of *C. acetobutylicum* 824. Re-targeted introns were named in accordance with the previously published TargeTron and ClosTron nomenclature (Heap et al., 2010b; Heap et al., 2007; Karberg et al., 2001). For example, the intron on plasmid pMTL007C-E2::CaP0168-657s has been retargeted to insert in the sense orientation after base 657 of the Cap0168 gene on plasmid pSOL1 of *C. acetobutylicum*.

The *C. acetobutylicum* ATCC 824 genome contains two α-amylase genes (Cap0098 and Cap0168) which are located on the mega-plasmid pSOL1 (Nolling et al., 2001). These genes together with glucoamylase Cac2810 on the chromosome are involved in starch degradation (Jones et al., 2008) but which of these is mainly responsible for *α-amylase* production is currently unknown as loss of pSOL1 leads to severe deficiency in starch degradation (Sabathe et al., 2002). Three target sites were selected for each gene identified by algorithm available at http://clostron.com/clostron2.php which was (485a, 565s, 1039a) and (456s, 657s, 918s) for genes Cap0098 and Cap0168 respectively. Similarly, two target sites 656a and 1167s were identified for *C. acetobutylicum Polygalacturonase* (*Pectinase*) gene Cac0355 which could be responsible for pectin degradation. Intron re-targeting regions of about 353bp for each target site was DNA synthesized by DNA2.0 and ligated into pMTL007C-E2 between HindIII and BsrGI sites by replacing the *lacZalpha*

region to generate the re-targeted derivatives of pMTL007C-E2 and shipped as ready to transform in *C. acetobutylicum.*

2.5.3 pMTL-ME3 In-frame deletion vector using *pyrE*

To construct a general knockout using negative selection marker *pyrE* and a complementation vector for the *C. acetobutylicum pyrE* mutant, the *C. sporogenes* ATCC 15579 *pyrE* gene was first amplified by PCR from genomic DNA using primers Csp-*pyrE*-HpaI-sF1 and Csp-*pyrE*-HpaI-sR1. Csp-*pyrE*-HpaI-sF1 contains the *pyrE* ribosomal binding sites (RBS) but not its promoter. The resulting PCR product of 600bp was digested with HpaI and ligated in the same orientation as *catP* into pMTL83151 (Heap et al., 2009) linearized with the same enzyme and treated with antarctic phosphatase to prevent self-ligation of empty vector. Insertion was verified by restriction digestion and nucleotide sequencing. The generated *catP-pyrE* fragment was excised from the intermediate vector pMTL-ME2 as FseI/PmeI fragment and ligated into the equivalent sites in modular vector pMTL85141 (Heap et al., 2009) to construct the vector pMTL-ME3 (Figure 2.2). This plasmid was used as a complementation vector for *C. acetobutylicum* ATCC 824 *pyrE* mutant strain as well as general plasmid for making in-frame deletions in *C. acetobutylicum* using *pyrE* as a negative selection marker.

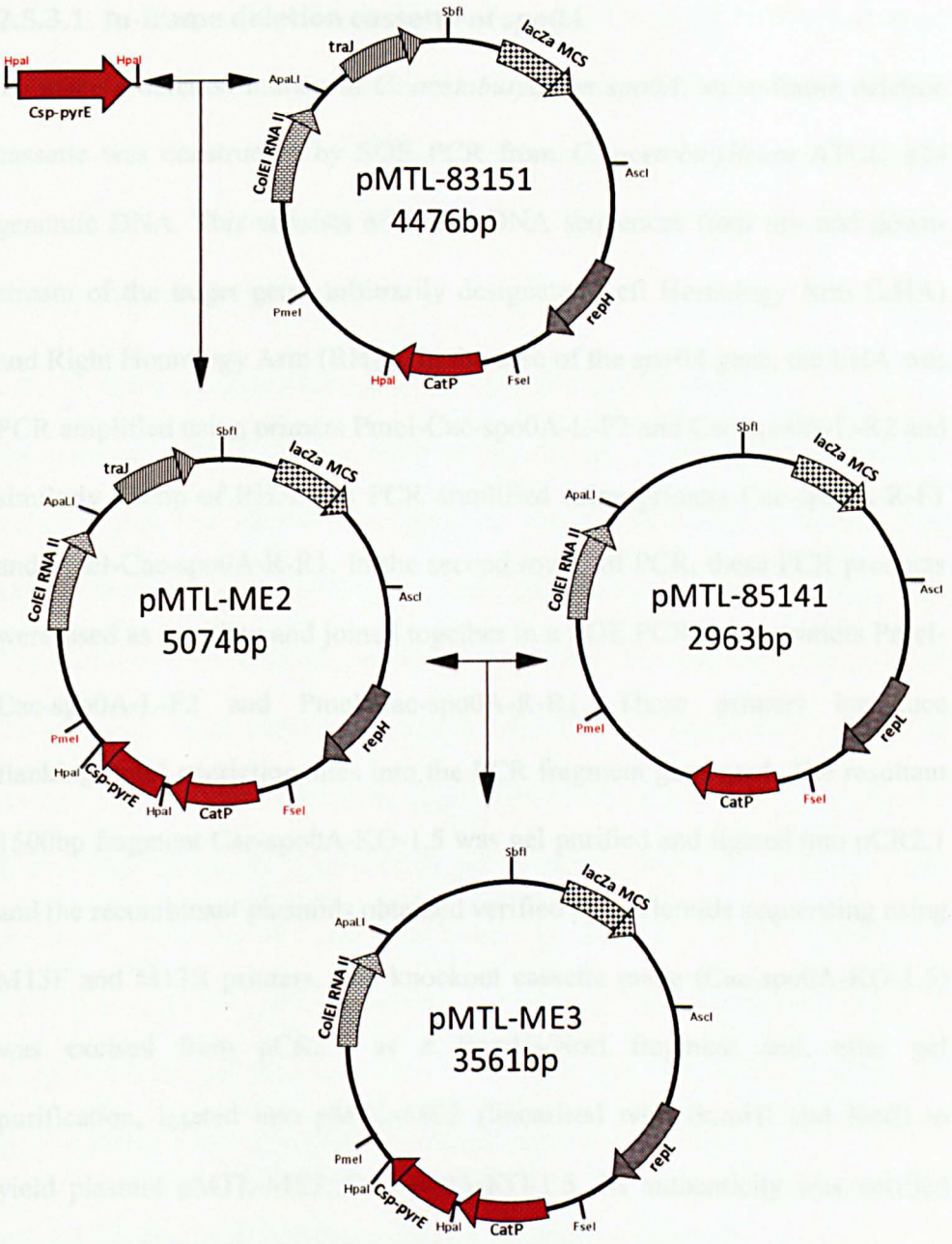


Figure 2.2 Construction of pMTL-ME3 in-frame deletion vector

This is a general complementation and deletion vector. *C. sporogenes pyrE* gene is co-expressed with *catP* antibiotic marker from the *catP* promoter. Vector pMTL-ME3 was used to complement the *C. acetobutylicum pyrE* mutant (CRG CC#1545). It also contains Gram-positive origin of replication from pIM13 vector *repL*, Gram-negative origin of replication ColEI RNA II and multiple cloning sites (MCS) between SbfI and AscI. Restriction sites used in each step of cloning are marked in red.

2.5.3.1 In-frame deletion cassette of *spo0A*

To make a deletion mutant in *C. acetobutylicum spo0A*, an in-frame deletion cassette was constructed by SOE PCR from *C. acetobutylicum* ATCC 824 genomic DNA. This consists of 750bp DNA sequences from up- and down-stream of the target gene, arbitrarily designated Left Homology Arm (LHA) and Right Homology Arm (RHA). In the case of the *spo0A* gene, the LHA was PCR amplified using primers PmeI-Cac-spo0A-L-F2 and Cac-spo0A-L-R2 and similarly 750bp of RHA was PCR amplified using primers Cac-spo0A-R-F1 and PmeI-Cac-spo0A-R-R1. In the second round of PCR, these PCR products were used as template and joined together in a SOE PCR using primers PmeI-Cac-spo0A-L-F2 and PmeI-Cac-spo0A-R-R1. These primers introduce flanking PmeI restriction sites into the PCR fragment generated. The resultant 1500bp fragment Cac-spo0A-KO-1.5 was gel purified and ligated into pCR2.1 and the recombinant plasmids obtained verified by nucleotide sequencing using M13F and M13R primers. The knockout cassette made (Cac-spo0A-KO-1.5) was excised from pCR2.1 as a BamHI/NotI fragment and, after gel purification, ligated into pMTL-ME3 (linearised with BamHI and NotI) to yield plasmid pMTL-ME3::Cac-spo0A-KO-1.5. Its authenticity was verified by restriction digestion and nucleotide sequencing.

2.5.3.2 Complementation vector for *spo0A* at *pyrE* locus

To make an ACE complementation vector for the *C. acetobutylicum spo0A* mutant, the *spo0A* gene of *C. acetobutylicum* ATCC 824 was PCR amplified with its native promoter (Harris et al., 2002) from *C. acetobutylicum* ATCC 824 genomic DNA using primers NotI-Cacspo0A-F1 and EcoRI-Cacspo0A-R1. The PCR product obtained (ca. 1211bp) was digested with restriction

enzymes NotI and EcoRI and ligated into pMTL-ME7C which had been linearised with the same restriction enzymes, to make pMTL-ME7C-PCacspo0A. This plasmid could be used to complement the in-frame deleted *spo0A* gene at the *pyrE* locus using ACE. It was verified by restriction digestion and nucleotide sequencing.

2.5.3.3 Over-expression vector for *spo0A* at *pyrE* locus

To make an ACE over-expression vector for use in the *C. acetobutylicum spo0A* mutant, an alternative plasmid to pMTL-ME7C was constructed in which the promoter of the *C. sporogenes* gene *ferredoxin* (*fdx*) was included. The *spo0A* of *C. acetobutylicum* ATCC 824 was PCR amplified from genomic DNA using primers NdeI-Cacspo0A-F1 and EcoRI-Cacspo0A-R1. The PCR product obtained (ca.859bp) was digested with NdeI and EcoRI and ligated downstream of the *fdx* promoter in pMTL-ME7X linearised with the same restriction enzymes, to make pMTL-ME7X-Cacspo0A. It was verified by restriction digestion and nucleotide sequencing. This vector was used to complement the *spo0A* mutant strain at the *pyrE* locus but also to over-express *spo0A* using the strong promoter of the *fdx* gene of *C. sporogenes*.

2.5.3.4 In-frame deletion cassette of CaP0168 (*amy B*)

To make a deletion mutant in *C. acetobutylicum amylase B* gene, an in-frame deletion cassette flanked by BamHI and NcoI was designed by taking 750bp region downstream of CaP0168 together with its stop codon as the LHA, and 750bp region upstream of CaP0168 including its start codon as the RHA. This knockout cassette was synthesised by Eurofine www.eurofinsgenomics.eu and subsequently ligated into pMTL-ME3 between the BamHI and NcoI restriction enzyme sites to make pMTL-ME3::CaP0168KO-1.5. It was verified by

restriction digestion and nucleotide sequencing. This vector was used to make an in-frame deletion of CaP0168.

2.5.3.5 Over-expression vector for CaP0168 at *pyrE* locus

To construct an over-expression vector for *amylase B*, CaP0168 gene was PCR amplified from the *C. acetobutylicum* ATCC 824 genomic DNA without its promoter using Cac_P-0168-F2 and Cac_P-0168-R1. There were several NdeI sites found in this gene, so this enzyme could not be used to produce a fragment for ligation into the NdeI site of pMTL-ME7X. AseI which is compatible with NdeI was therefore used. The 2300bp PCR product obtained was digested with AseI and BamHI and cloned in pMTL-ME7X linearised with NdeI and BamHI to make pMTL-ME7X-CaP0168 and the plasmid verified by restriction digestions and nucleotide sequencing.

2.5.4 pMTL-SC7515 In-frame deletion vector using *codA*

Recently the *E. coli codA* has been used in *C. difficile* as a heterologous negative selection marker to make precise changes to the *tcdC* gene (Cartman et al., 2012). To establish the use of this counterselection marker in *C. acetobutylicum,* vector pMTL-SC7515 was constructed (Figure 2.3) by removing the pCB102 Gram-positive replicon of pMTL-SC7315 as an AscI/FseI fragment and replacing this with pIM13 from similarly digested pMTL8514 (Data obtained from Steve Cartman).

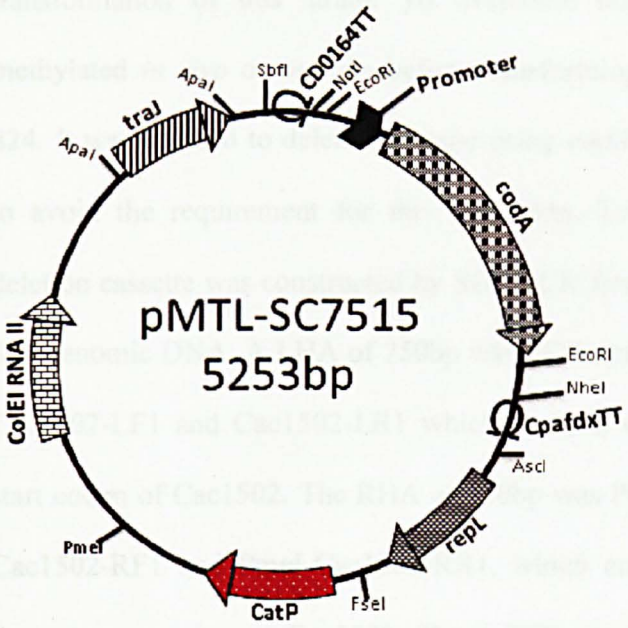


Figure 2.3 In-frame deletion vector using *codA*

This plasmid contained a codon optimised version of *codA* from *E. coli* expressed from its own promoter ligated between two EcoRI sites flanked by two terminators after the SbfI and before the AscI restriction sites. It also contains the Gram-positive replicon pIM13 (*repL*), Gram-negative replicon ColE1 RNA II, and *catP* antibiotic marker which confers resistance to chloramphenicol/thiamphenicol.

2.5.4.1 In-frame deletion cassette of *spo0A*

The in-frame deletion cassette created previously for *spo0A* was excised from pCR2.1as a BamHI and NcoI fragment and ligated into pMTL-SC7515 linearised with same set of restriction enzymes to make pMTL-SC7515::Cacspo0AKO-1.5. This vector was used to make an in-frame deletion at the *spo0A* locus in *C. acetobutylicum* ATCC 824 wild-type using *codA* as a counterselection marker.

2.5.4.2 In-frame deletion cassette of Cac1502

C. acetobutylicum ATCC 824 chromosome contains a type II restriction endonuclease (Cac1502) which is considered a strong barrier to efficient

transformation of this strain. To overcome this issue, plasmid DNA is methylated *in vivo* or *in vitro* before transforming *C. acetobutylicum* ATCC 824. It was decided to delete this gene using *codA* as counterselection marker to avoid the requirement for this later step. To achieve this, an in-frame deletion cassette was constructed by SOE PCR from *C. acetobutylicum* ATCC 824 genomic DNA. A LHA of 750bp was PCR amplified using primers PmeI-Cac1502-LF1 and Cac1502-LR1 which consists of the upstream region and start codon of Cac1502. The RHA of 750bp was PCR amplified using primers Cac1502-RF1 and PmeI-Cac1502-RR1, which contains the stop codon and downstream region of Cac1502. These PCR products were gel purified and used as templates in the second round of PCR with primers PmeI-Cac1502-LF1 and PmeI-Cac1502-RR1 to make a 1500bp KO cassette. This was then digested with PmeI and ligated in pMTL-SC7515, already linearised with PmeI and treated with antarctic phosphatase to make pMTL-SC7515::Cac1502KO-1.5. This plasmid was used to make an in-frame deletion of Cac1502.

2.5.5 Transformation efficiency of ΔCac1502

To determine the transformation efficiency of the ΔCac1502 compared to the *C. acetobutylicum* ATCC 824 WT strain, a modular shuttle plasmid pMTL85141 was used as a standard vector. A total of 1μg of unmethylated and pAN2 methylated DNA were transformed as described previously into both the ΔCac1502 and WT strain in triplicate. Cells were recovered after 5 hours of incubation and serial dilutions 10^{0}-10^{-7} were prepared. 100μl from each of the dilution was plated onto CGM Tm15μg/ml and plates were incubated for 24hrs at 37°C. The number of transformant colonies were counted and estimated as CFU/ml/μg of DNA.

Chapter Three: Allele Coupled Exchange (ACE)

3 Allele-Coupled Exchange (ACE)

3.1 Introduction

In order to improve the ability of *C. acetobutylicum* to make butanol from renewable feedstocks, it will be necessary to bring about defined changes to the organism, including the deletion, modification and addition of DNA. Crucial is the availability of the necessary gene tools with which the required changes can be made. Prior to the initiation of this project, ClosTron technology had made possible the rapid and reproducible creation of mutants for metabolic engineering purposes. However, metabolic engineering requires more than just 'knockout' systems. It also needs the ability to 'add' genetic information, *ie.*, 'knock-in'. Whilst the ClosTron can be used to introduce cargo DNA, there is a size limit of no more than 1.0 kb. Furthermore, metabolic engineering may require the replacement of existing genes with modified variants. This is not possible with ClosTron technology. A further disadvantage of the ClosTron is that, in common with any insertional mutagen, it can cause polar effects, where genes residing downstream or upstream of the intron insertion can be affected. Ideally what is required is a recombination-based method which allows allelic exchange.

The required level of genetic modification methods based on homologous recombination are well established in other organisms including *E. coli*, *B. subtilis*, yeast etc. The minimum amount of homology regions required for efficient homologous recombination process in these organism is 50 nucleotides (nt) (Datsenko & Wanner, 2000; Fabret et al., 2002; Nikawa & Kawabata, 1998; Wach, 1996) while in mammalian cells recombination

frequency was severely affected when the length of homology was dropped from 214 to 163 base pairs (Rubnitz & Subramani, 1984). Such procedures were successfully used in *E. coli* and *Saccharomyces cerevisiae* where linear PCR fragments were used to alter the desired allele. These PCR products contained a FRT flanked antibiotic marker, flanked by less than 50nt long oligonucleotides that have homology to the 3` and 5`ends of the target gene to be inactivated (Datsenko & Wanner, 2000; Wach, 1996). In the case of *B. subtilis* a negative selection marker, *upp* was used for the selection of the second crossover events (Fabret et al., 2002). Transfer of linear DNA to C. *acetobutylicum* ATCC 824 and related species is not considered feasible given the very low transformation efficiency with replicative plasmids, even after protecting them against the strong type II restriction endonuclease system *Cac*824*I*, a defence of the host against foreign DNA incursion. Classical allelic exchange has been proven difficult in clostridia. The minimum length of the host nucleotide sequence that can efficiently promote homologous recombination in *Clostridium* is not yet known. Therefore only a handful of genes have been mutated by single crossover integration of plasmid into *C. acetobutylicum* ATCC 824. These include butyrate kinase *buk*, Cac3075; phosphate acetyltransferase *pta*, Cac1742; alcohol aldehyde dehydrogenase *aad*, Cap0162, and *solR*. Positive selection using resistance to antibiotic was used in all the cases. These mutants were made using at least 650bp or more of an internal fragment of the respective gene on the plasmid which then recombined with the host chromosome by Campbell-like integration (Green & Bennett, 1996; Green et al., 1996; Nair et al., 1999; Sarker et al., 1999). Moreover, these mutants were unstable due to the presence of two copies of

repeated sequence in the chromosome which can lead to plasmid excision if antibiotic selection is removed. Experience shows that sequence length below 300bp severely affect the recombination process in *Clostridium* and ability to detect single crossover integrants (S. T. Cartman and N. P. Minton, personal communication).

To explore the possible development of allelic exchange methods, a system was devised which set out to promote the integration of DNA into the clostridial chromosome by using antibiotic resistance-encoding plasmid that were defective for replication ('pseudo-suicide' plasmids) (Heap et al., 2009), and which were endowed with two host-derived regions of homology to the chromosome of asymmetric size. The one, represented an extensive region of homology (*ie.*, 1200bp) and was termed the Long Homology Arm (LHA). The other region of DNA was considerably smaller, and was called the Short Homology Arm (SHA) and was composed of only 300bp. In general, the antibiotic resistance gene carried by the plasmid was *catP*, which specifies resistance to thiamphenicol (Tm). When cells carry such a plasmid are grown in the presence of Tm, their growth rate is limited by the rate at which the replication defective plasmid is partitioned to the daughter cells at cell division. However, when the plasmid integrates into chromosome, the cell that arises has a growth advantage as *catP* is now received by all daughter cells at cell division as the antibiotic resistance gene is now in the chromosome. These cells are readily identified as they form large colonies due to their faster growth in the presence of Tm.

As the LHA is 4X greater in size than the SHA, initial integration of the plasmid invariably will take place at the LHA. Once integrated, the cell will

carry two copies of each of the homology arms. It follows that it can excise through recombination between either pair, or will predominately take place between the two copies of LHA as opposed to the two copies of SHA. This will lead to the generation of a wild-type cell that carries the original plasmid. However, the system was designed such that when recombination takes place between the two copies of SHA, a new allele is formed that is selectable. This allows this relatively rare (compared to recombination between the two copies of LHA) event to be isolated. In essence, a chromosomal allele is coupled to a plasmid-borne allele to create a new allele that can be selected, hence Allele-Coupled Exchange (ACE) technology. Using this method, it is clear that if a piece of DNA is inserted between the SHA and LHA on the plasmid, then it will be incorporated into the genome, providing a facility for 'knock-in'.

Pivotal to the system, was the identification of alleles which when 'coupled' (combined) would create a new selectable allele. Two basic systems were explored:

(i) The first system is based on genes such as *pyrF* or *pyrE* which are both positively and negatively selectable. The *pyrF* gene encodes orotidine 5-phosphate decarboxylase and is involved in *de novo* pyrimidine biosynthesis pathway as is orotate phosphoribosyltransferase, encoded by *pyrE* (Boeke et al., 1984; Haas et al., 1990; Yamagishi et al., 1996). Cells which contain *pyrF* and *pyrE* genes can be positively selected on growth medium lacking uracil, as they are required for uracil biosynthesis but cells that lack either *pyrF* or *pyrE* can be negatively selected on growth medium supplemented with antimetabolite 5-fluoroorotic acid (5-FOA). The above mentioned enzymes can convert 5-FOA into 5-fluorouracil (5-FU) in the

form of 5-fluorouridylic acid (5-FUMP). 5-FU is a pyrimidine analogue that can be misincorporated into DNA and RNA in place of thymine or uracil, leading to interference in replication and transcription and thus to cell death (Jones & Fink, 1982). Mutants of these genes require exogenous uracil for growth and are viable in medium containing 5-FOA (Heap et al., 2007).

(ii) In the second strategy, a promoter-less selectable marker on the plasmid is coupled to a chromosomally located *thiolase* (*thl*) promoter, which is known to be strongly expressed throughout growth (Harris et al., 2002; Tummala et al., 1999). Two promoter-less markers were investigated: a macrolide–lincosamide–streptogramin (MLS) antibiotic resistance marker *ermB* and a *C. sporogenes* ATCC 15579 *pyrE* gene. This latter gene was chosen because it is only 47.6% identical to the *C. acetobutylicum pyrE*, and should, therefore, not recombine at the host's *pyrE* locus.

3.2 Aim of this study

The aim of this study was to develop a reproducible method of bringing about allelic exchange in *C. acetobutylicum* using a novel approach based on the use of asymmetric homology arms, termed Allele-Coupled Exchange Technology (ACE). The use of two different types of alleles were explored, those based on genes involved in uracil metabolism (*pyrE* and *pyrF*), and those based on promoter-less copies of genes encoding either antibiotic resistance (*ermB*) or genes that restore uracil prototrophy (*pyrE*). Once derived, the utility of the system for introducing large fragments of DNA was explored through either the iterative use of the method (phage lambda) or in a single step using an inverse application of the method (pSOL1). The specific aims were:

1. To exemplify ACE using *pyrF* and *pyrE* alleles
2. To exemplify ACE using promoter-less marker genes (*ermB* and *pyrE*)
3. To demonstrate the iterative use of the method
4. To introduce large fragments of DNA (phage lambda and pSOL1) into the chromosome

In the following results section, some of the studies involving *pyrF* (Sections 2.5.1.1 and 3.3.1) and construction of plasmids for iterative ACE exemplification (Section 2.5.1.9 only) were undertaken by John Heap. All the other sections (unless indicated) were performed by the PhD candidate.

3.3 Results

3.3.1 Inactivation of *pyrF* by ACE using pMTL-JH2

To exemplify the allelic exchange method, the *C. acetobutylicum pyrF* gene was selected as the first target site. The *pyrF* gene encodes orotidine 5-phosphate decarboxylase and is involved in the *de novo* pyrimidine biosynthesis pathway. It was demonstrated that inactivation of this gene using ClosTron technology enables such mutants in a number of different clostridial species, to be selected on media supplemented with antimetabolite 5-fluoroorotic acid (FOA) (Heap et al., 2007).

As discussed in the introduction to this chapter, a special form of allele exchange cassettes was designed (Figure 3.1) using asymmetric homology regions in an attempt to control the order of the recombination events that take place. A long homology arm (LHA) of 1200bp immediately downstream of *pyrF* containing *pyrZ* and part of the adjacent *pyrD* gene was intended to direct the first recombination event so that a large majority of single-crossover clones

that arise are the result of recombination in this region, which would not inactivate *pyrF*.

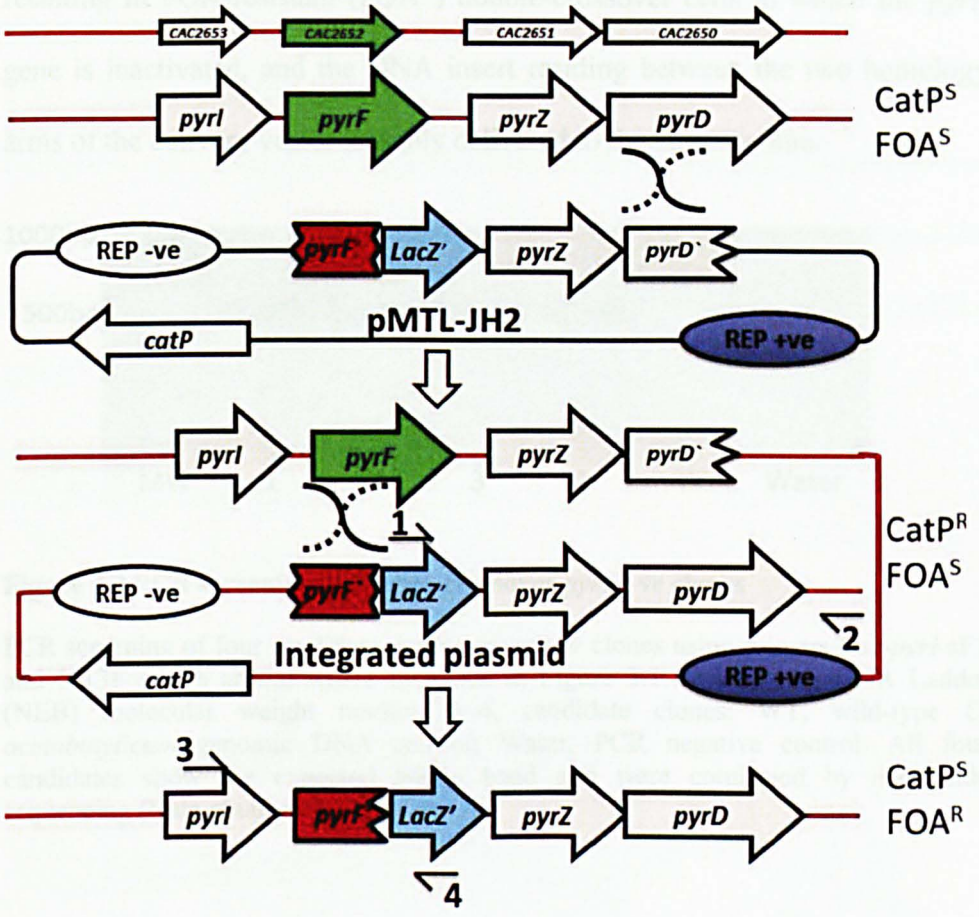


Figure 3.1 Allelic exchange using pMTL-JH2

The first recombination event (plasmid integration) is mediated by the long homology arm (LHA) between pMTL-JH2 and *pyrZ/pyrD* regions on chromosome. Single crossover clones are obtained on medium supplemented thiamphenicol due to the presence of antibiotic marker (*catP)* that confers resistance to thiamphenicol/chloramphenicol ($CatP^R$). Single crossover could be screened by PCR using primers lacZalpha-sF2 (1) and Cac-pyrD-sR1 (2). The second recombination event (plasmid excision) is mediated by the short homology arm (SHA) of plasmid pMTL-JH2 and an internal fragment of *pyrF*. Double crossover clones are selected on media supplemented with Uracil 1μg/ml and 5-fluoroorotic acid (FOA) 400ug/ml by acquiring 5-fluoroorotic acid resistance (FOA^R) and become thiamphenicol/chloramphenicol-sensitive ($CatP^S$). Double crossover integrants were screened by PCR using primers Cac-pyrI-sF1 (3) and M13F (4). Primer binding sites are indicated by arrows (data obtained from John Heap).

A subsequent recombination event at a short homology arm (SHA) corresponding to a 300bp internal portion of *pyrF* would excise the plasmid, resulting in FOA-resistant (FOA^R) double-crossover cells in which the *pyrF* gene is inactivated, and the DNA insert residing between the two homology arms of the delivery vector is stably delivered to the chromosome.

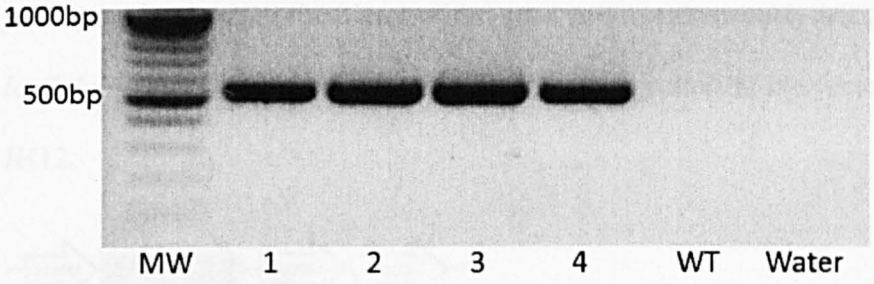


Figure 3.2 PCR screening of double-crossover *pyrF* -ve clones

PCR screening of four candidate double-crossover clones using primers Cac-*pyrI*-sF1 and M13F which anneal where indicated in Figure 3.1. MW, 2-Log DNA Ladder (NEB) molecular weight marker; 1–4, candidate clones; WT, wild-type *C. acetobutylicum* genomic DNA control; Water, PCR negative control. All four candidates show the expected 549bp band and were confirmed by nucleotide sequencing (Data obtained by John Heap).

To construct the above stated recombinant strain, plasmid pMTL-JH2 DNA was transformed into *C. acetobutylicum* and cells were selected on CGM supplemented with 15µg/ml Tm. Four transformants were subcultured twice on the same medium (passages P1 and P2) and followed by streaking onto CGM supplemented with 400µg/ml FOA and 1µg/ml Uracil to select the desired *pyrF*-minus double-crossover clones. FOA^R colonies were obtained from each of the four independent subcultures (passage P3), and one clone from each was purified by subculturing again on the same medium (passage P4). Overnight liquid cultures were set up supplemented with 20µg/ml uracil and the following day genomic DNA was extracted and PCR screening showed the expected

double-crossover genotype for all four clones (Figure 3.2). These clones were then confirmed for ACE plasmid loss by replica-plating on medium with and without thiamphenicol (Heap J).

3.3.2 Integration of lambda-DNA cargo at *pyrF*

In Section 3.3.1, allelic exchange was used to inactivate the *C. acetobutylicum pyrF* gene by deleting the 3`end of the gene and simultaneously integrated the *lacZalpha* region present between the two homology arms in the vector pMTL-JH12.

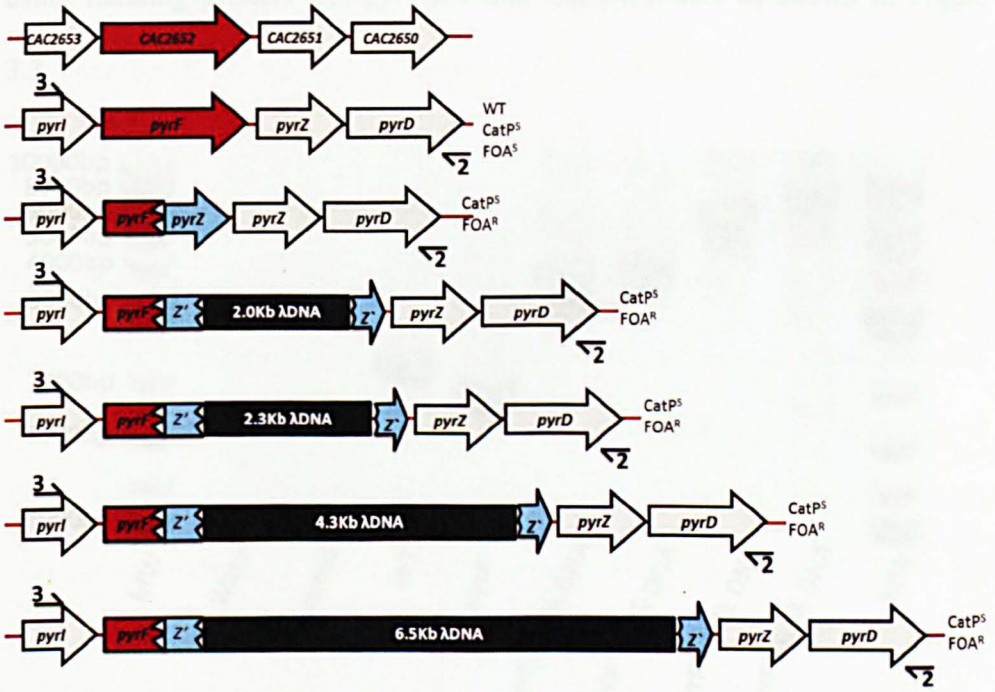


Figure 3.3 Strains with integrated lambda DNA fragments using ACE

This figure shows the WT; wild-type gene arrangement at the *C. acetobutylicum pyrF* locus (1 and 2), *pyrF* mutant (3), and integration of various lambda DNA fragments ranged from 2.0 kb to 6.5 kb shown to scale by black shading (4-7). Primers Cac-pyrI-sF1 (3) and Cac-pyrD-sR1 (2) were used for screening of double crossover integrants. $CatP^S$; indicates sensitivity to thiamphenicol/chloramphenicol; and FOA^R indicates 5-fluoroorotic acid-resistance. Primer binding sites are indicated by arrow.

To test the feasibility of the method for the integration of large DNA sequences, various HindIII lambda DNA fragments ranging from 2.0Kb to 6.5Kb were cloned in pMTL-JH2 between the two asymmetric homology arms to construct derivative vectors of pMTL-JH2.

The procedure previously used to integrate *lacZalpha* regions of pMTL-JH2 into the chromosome was repeated using the derivative plasmids carrying various DNA fragments of phage lambda. Following the selection of the desired double crossover clones, they were screened for authenticity by PCR using flanking primers Cac-*pyrI*-sF1 and Cac-*pyrD*-sR1 as shown in Figure 3.3.

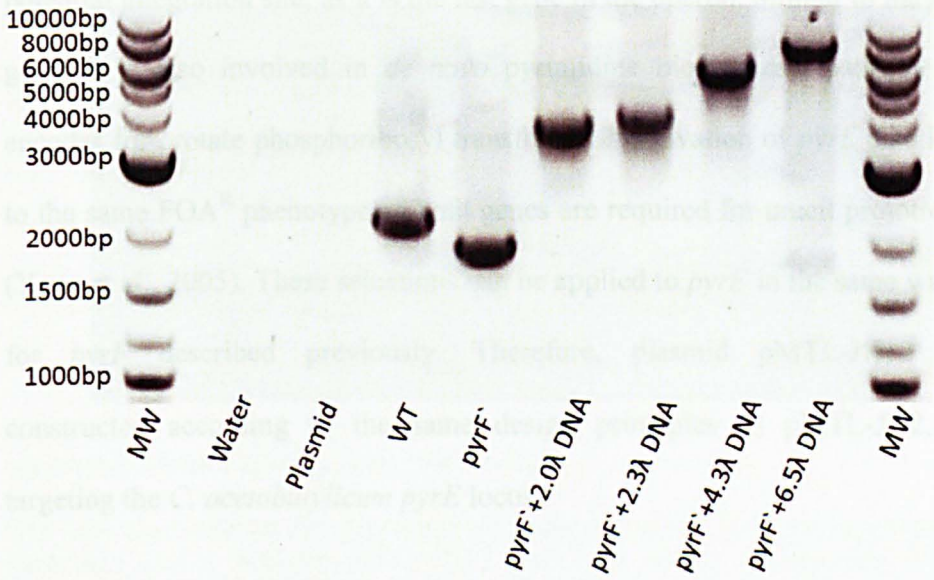


Figure 3.4 Verification of lambda fragment integration by PCR

PCR screening of double-crossover clones shows integration of *lacZalpha* and various sizes of lambda DNA fragment at the *pyrF* locus obtained using pMTL-JH2 and its derivative vectors. Primers Cac-pyrI-sF1 and Cac-pyrD-sR1 were used for screening and annealing sites are indicated in Figure 3.3. MW, 2-Log DNA Ladder (NEB) molecular weight marker; Water, PCR negative control; Plasmid, pMTL-JH2 DNA control; WT, wild-type *C. acetobutylicum* ATCC 824 genomic DNA control (2195bp); *pyrF'*, *C. acetobutylicum* ATCC 824 *pyrF*-ve genomic DNA control (1967bp); and all four derivatives strains showed the expected products of (3997bp), (4292bp), (6336bp) and (8523bp) after integration of Lambda2.0, Lambda2.3, Lambda4.3, and Lambda6.5 fragments respectively.

As shown in Figure 3.4 DNA fragments of the expected size were obtained in all cases. In each case the DNA fragment generated by PCR was subjected to nucleotide sequencing and the sequence shown to conform to the lambda sequence expected (Data not shown).

3.3.3 Inactivation of *pyrE* locus using pMTL-JH12

In Section 3.3.2, it was demonstrated that the *C. acetobutylicum pyrF* locus could be used to integrate upto 6.5Kb of DNA sequence. However, the *pyrF* gene lies in the middle of the *pyr* operon and integration of any synthetic gene or operon may, therefore, affect the expression of downstream genes. To overcome such issues, the *C. acetobutylicum pyrE* locus was chosen as a potential integration site, as it is the last gene in the operon. Similar to the *pyrF* gene, it is also involved in *de novo* pyrimidine biosynthesis pathway and encodes for orotate phosphoribosyl transferase. Inactivation of *pyrE* also leads to the same FOA^R phenotype as both genes are required for uracil prototrophy (Yano et al., 2005). These selections can be applied to *pyrE* in the same way as for *pyrF* described previously. Therefore, plasmid pMTL-JH12 was constructed according to the same design principles as pMTL-JH2, but targeting the *C. acetobutylicum pyrE* locus.

Plasmid pMTL-JH12 DNA was transformed into *C. acetobutylicum* and transformants were selected on medium supplemented with 15μg/ml Tm. Eight transformants were subcultured twice on the same medium (passage P1 and P2) before overnight cultures from each were set up in 2xYTG medium supplemented with 7.5μg/ml Tm. Genomic DNA was extracted and all clones were screened by PCR using primers lacZa-sF2 (1) and Cac-hydA-sR2 (2) (as shown in Figure 3.5) to show that they were single-crossover integrants in

which plasmid pMTLJH12 had integrated via the long region of homology (LHA), into the chromosome by homologous recombination by generating a fragment of 1428bp (Figure 3.6).

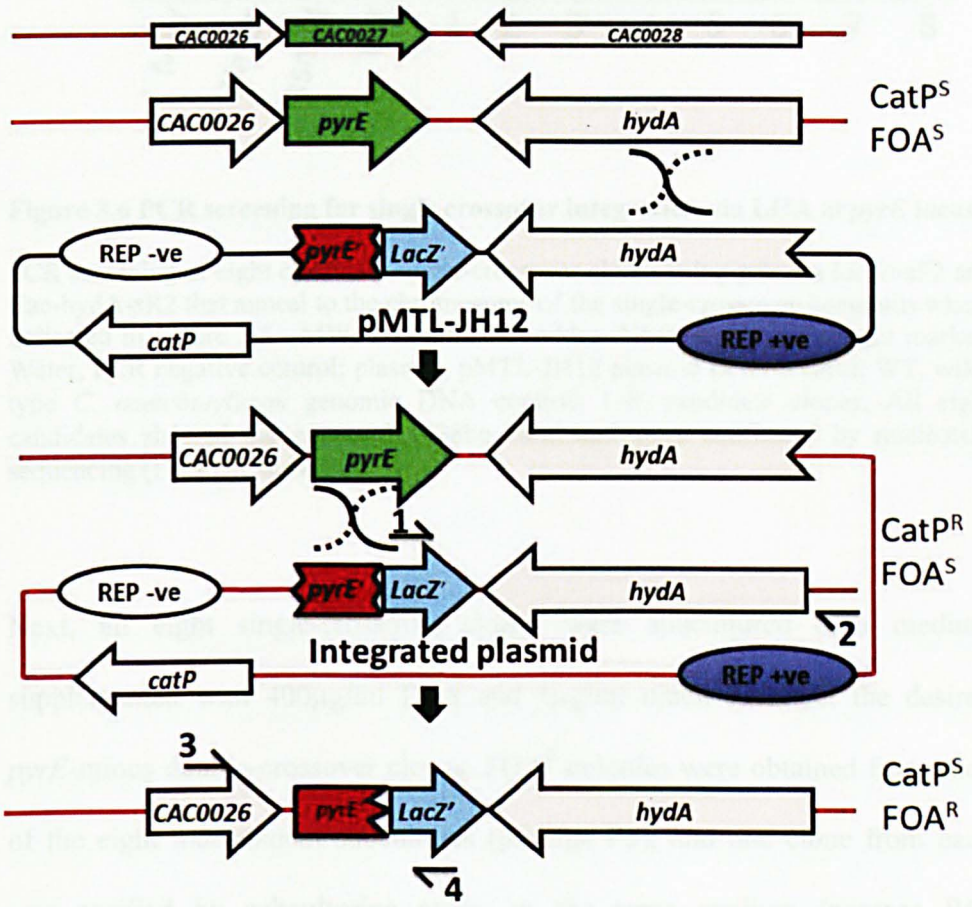


Figure 3.5 Allelic exchange using pMTL-JH12

The first recombination event (plasmid integration) is mediated by the long region of homology (LHA) between pMTL-JH12 and *hydA* in the chromosome. Single-crossover clones are obtained on medium containing thiamphenicol where resistance is conferred by *catP* on the plasmid. Integration is detected by PCR using primers *lacZa*-sF2 (1) and Cac-hydA-sR2 (2). The second recombination event (plasmid excision) is mediated by the short region of homology (SHA) between pMTL-JH12 and an internal portion of *pyrE* on the chromosome. Double-crossover clones are selected on medium supplemented with 5-fluoroorotic acid (FOA) and Uracil and screened by PCR using primers CAC0026-sF2 (3) and M13F (4). Thiamphenicol-sensitive ($CatP^S$); Thiamphenicol-resistant ($CatP^R$); 5-fluoroorotic acid-resistant (FOA^R); 5-fluoroorotic acid-sensitive (FOA^S). Primer binding sites are indicated by arrows.

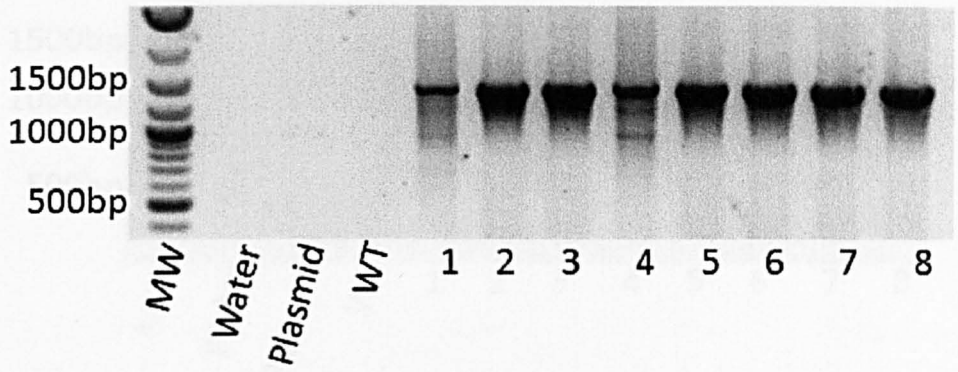


Figure 3.6 PCR screening for single crossover integration via LHA at *pyrE* locus

PCR screening of eight candidate single-crossover clones using primers *lacZa*-sF2 and Cac-hydA-sR2 that anneal to the chromosome of the single-crossover integrants where indicated in Figure 3.5 . MW, 2-Log DNA Ladder (NEB) molecular weight marker; Water, PCR negative control; plasmid, pMTL-JH12 plasmid DNA control; WT, wild-type *C. acetobutylicum* genomic DNA control; 1–8, candidate clones. All eight candidates showed the expected 1428bp band and were confirmed by nucleotide sequencing (Data not shown).

Next, all eight single-crossover clones were subcultured onto medium supplemented with 400µg/ml FOA and 1µg/ml uracil to select the desired *pyrE*-minus double-crossover clones. FOA^R colonies were obtained from each of the eight independent subcultures (passage P3), and one clone from each was purified by subculturing again on the same medium (passage P4). Overnight liquid cultures were set up in 2xYTG supplemented with 20µg/ml uracil and the following day genomic DNA was extracted and screened by PCR. This analysis demonstrated the expected double-crossover genotype for all eight clones in Figure 3.5 and Figure 3.7 . Loss of the ACE plasmid in all of the clones was confirmed by replica-plating on medium with and without thiamphenicol.

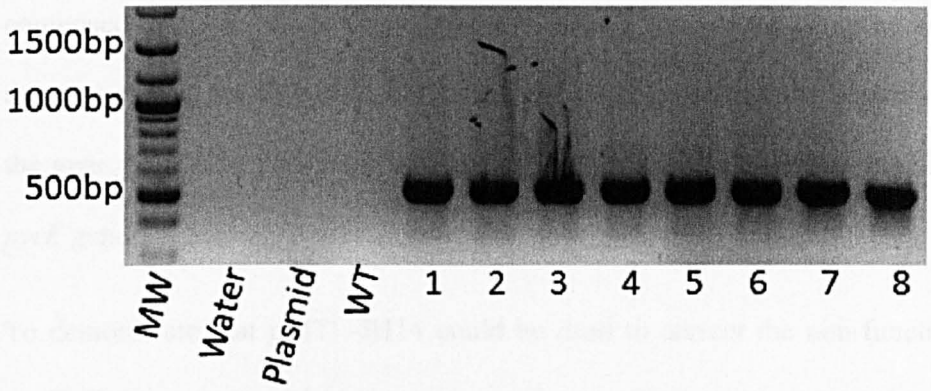


Figure 3.7 PCR screening for Double crossovers integration via SHA at the *pyrE* locus

PCR screening of eight candidate double-crossover clones using primers CAC0026-sF2 and M13F that anneal to the double-crossover chromosome where indicated in Figure 3.5 . MW, 2-Log DNA Ladder (NEB) molecular weight marker; Water, PCR negative control; Plasmid, pMTL-JH12 plasmid DNA control; WT, wild-type *C. acetobutylicum* genomic DNA control; 1–8, candidate clones. All eight candidates showed the expected 558 bp band.

One of the strains generated in this section (CRG CC# 1545)) was stored at -80 °C and was selected to be used as the host strain for making in-frame deletion using *C. sporogenes pyrE* as a heterologous negative selection marker and for complementation and over-expression of deleted genes (Chapter Four).

3.3.4 Repairing of *pyrE* using pMTL-JH14

The *pyrE* allele in the vector pMTL-JH12 (used to make the *pyrE* mutant strain described in Section 3.3.3), was specially configured to facilitate subsequent genetic modification using pMTL-JH14 or its derivative vectors. To achieve repair of the *pyrE* mutation, two non-functional alleles; one on the plasmid and the other counter-part on the chromosome, were fused together to make one functional allele of *pyrE* which enabled the cells to grow on minimal medium lacking uracil. To achieve this, plasmid pMTL-JH14 was designed and constructed exactly as was the case with pMTL-JH12 except that the SHA was

composed of a *pyrE* allele that carried the entire 3'-end of the gene (including the stop codon) but as before still lacked the 13 codons from the beginning of the gene. This new plasmid-borne allele was, therefore, also a non-functional *pyrE* gene.

To demonstrate that pMTL-JH14 could be used to correct the non-functional *pyrE* allele present on the chromosome of a uracil minus mutant made using pMTL-JH12, the following experiment was undertaken. A previously constructed *pyrE* mutant clone of *C. acetobutylicum* (CRG CC#1545 Table 2.1) was transformed with methylated pMTL-JH14 DNA. This strain lacks a functional *pyrE* gene and hence exogenous 20μg/ml uracil was added to the CGM media employed, together with Tm to select for transformants. Seven transformants were selected and each one was sub-cultured twice on the same medium (passages P1 and P2) to isolate highly pure single crossover integrants clones. They were not screened by PCR to check for integration of plasmid, as experience had shown that after just two passages on media supplemented with Tm the majority of the clones are composed of cells in which the plasmid has integrated via recombination between the LHA and complementary region in the chromosome. Such single crossover integrants are still uracil auxotrophs as they do not yet carry a functional copy of *pyrE*; rather they possess two non-functional alleles.

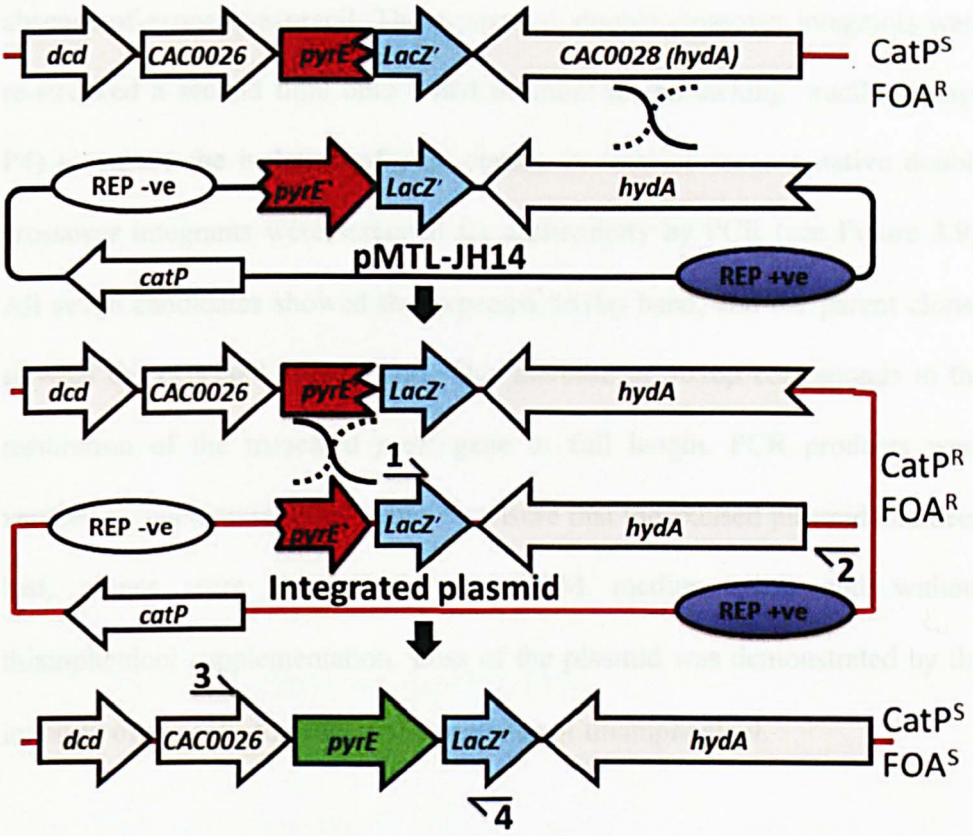


Figure 3.8 Homologous recombination events using pMTL-JH14

The first recombination event (plasmid integration) is mediated by the long homology arm (LHA) of vector pMTL-JH14 and the corresponding region *lacZalpha/hydA* on the chromosome. Single-crossover clones were enriched on medium containing Tm and uracil due to the presence of pIM13 replicon and could be screened by PCR using primers lacZa-sF2 (1) and Cac-hydA-sR2 (2). The second recombination event (plasmid excision) is mediated by the short homology arm (SHA) of vector pMTL-JH14 and the corresponding portion of *pyrE* on the chromosome. Double-crossover clones are selected using growth medium lacking uracil and screened by PCR using primers CAC0026-sF2 (3) and M13F (4) (Figure 3.8). Tm-sensitive ($CatP^S$); Tm-resistant ($CatP^R$); 5-fluoroorotic acid-resistant (FOA^R); 5-fluoroorotic acid-sensitive (FOA^S). Primer binding sites are indicated by arrows.

To select for cells in which these alleles have combined to form a functional *pyrE* allele, cells were sub-cultured onto CBM minimal medium lacking uracil (passage P3). At this stage, they grew like the wild-type, indicating that the desired plasmid element had excised, resulting in fusion (coupling) of the two dysfunctional *pyrE* alleles to form a functional gene that allowed growth in the

absence of exogenous uracil. These putative, double crossover integrants were re-streaked a second time onto CBM minimal media lacking uracil (passage P4) to ensure the isolation of pure clones. A total of seven putative double crossover integrants were screened for authenticity by PCR (see Figure 3.9). All seven candidates showed the expected 861bp band, and the parent clones showed the expected 558bp band. This increase of 303bp corresponds to the restoration of the truncated *pyrE* gene to full length. PCR products were verified by nucleotide sequencing. To ensure that the excised plasmid had been lost, clones were re-streaked onto CBM medium with and without thiamphenicol supplementation. Loss of the plasmid was demonstrated by the inability of the cells to grow in the presence of thiamphenicol.

Figure 3.9 PCR screening of Double cross-over using pMTL-JH14

PCR screening of seven candidate double-crossover clones using primers CAC0026-sF2 and M13F that anneal to the double-crossover chromosome where indicated in Figure 3.8. MW, 2-Log DNA Ladder (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL-JH14 plasmid DNA control; Cac824-*pyrE*-ve Parental clones, (CRG CC#1545, Table 2.1) genomic DNA control Figure 3.7 ; 1–7, candidate clones.

The strategy described above effectively replaces a negatively selectable gene with a sequence of interest, which could be achieved using a conventional allele exchange cassette (Knipfer et al., 1997). The fundamental difference is that the *pyrE* locus of a recombinant strain constructed using pMTLJH12 is specially designed to enable subsequent genetic modification (Figure 3.5). The non-functional *pyrE* in these cells is not completely deleted, it is only truncated at the 3`-end. Homologous recombination between this partial *pyrE* gene and an appropriate counterpart partial *pyrE* allele on plasmid pMTL-JH14, foreshortened at the 5`-end, would result in a full-length, functional gene.

3.3.5 Repairing of *pyrE* using pMTL-ME6

As shown in Section 3.3.4, fusion of two non-functional fragments of *pyrE* using pMTL-JH14 resulted in the creation of a functional *pyrE* gene. In plasmid pMTL-JH14, as the *lacZalpha* region is inserted between the two asymmetric homology arms, it also becomes integrated into the genome, downstream of the corrected *pyrE* gene. Its presence can be used as a distinguishing feature compared to the wild type. In the main, however, it is preferable that the *pyrE* locus is reverted back to exactly match the situation in the wild-type organism, with no additional sequences present. Therefore, a plasmid called pMTL-ME6 was designed and constructed. This plasmid is exactly the same as pMTL-JH14, but lacks the *lacZalpha* region downstream of *pyrE*. In this case, the *pyrE* mutant strain could be reverted back to wild-type in terms of phenotype as well as genotype using ACE technology.

To achieve this, the *C. acetobutylicum pyrE* mutant strain was transformed with methylated pMTL-ME6 vector DNA and transformants were selected on CGM medium supplemented with 15μg/ml Tm and 20μg/ml uracil. The allelic

exchange procedure was followed exactly as described for pMTL-JH14 to fuse the two non- functional *pyrE* alleles to make one functional *pyrE* gene (Figure 3.10). After the P4 passage on CBM medium, overnight cultures were set up in 2xYTG and following genomic DNA extraction screened by PCR using flanking primers Cac0026-sF2 and Cac-hydA-sR2 (Figure 3.10).

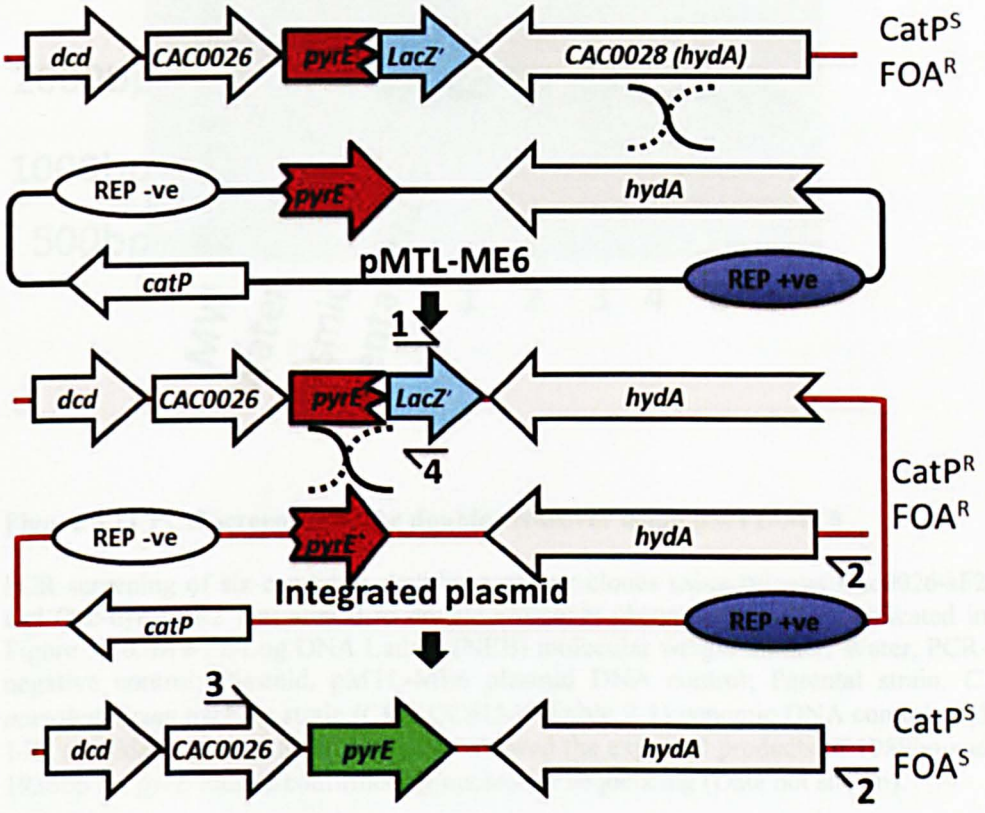


Figure 3.10 Allelic exchange using pMTL-ME6

Plasmid integration and selection of single crossover was achieved by long region of homology similar to pMTL-JH14. Plasmid excision via the short homology arm and selection of double crossover integrants was attained on medium lacking uracil and six clones were passaged independently till the last step. Double-crossover clones are selected using growth medium lacking uracil and screened by PCR using primers CAC0026-sF2 (3) and Cac-hydA-sR2 (2) (Figure 3.10). Tm-sensitive ($CatP^S$); Tm-resistant ($CatP^R$); 5-fluoroorotic acid-resistant FOA^R; 5-fluoroorotic acid-sensitive FOA^S.

Use of these primers in PCR produced ca. the same size of DNA fragment as for *pyrE* mutant made by pMTL-JH12 (Figure 3.11) but nucleotide sequence of the fragment showed that the fragment was indeed equivalent to that expected for a cell that had been reverted back to wild-type in the *pyrE* region (Data not shown).

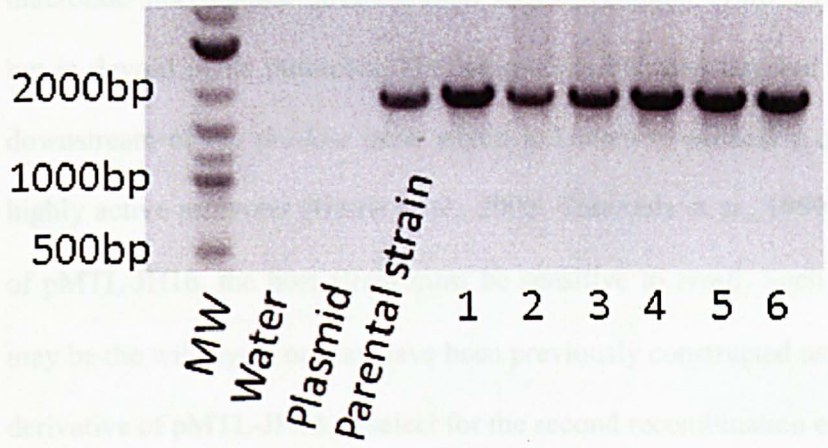


Figure 3.11 PCR screening of the double crossover using pMTL-ME6

PCR screening of six candidate double-crossover clones using primers Cac0026-sF2 and Cac-hydA-sR2 that anneal to double-crossover chromosome where indicated in Figure 3.10. MW, 2-Log DNA Ladder (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL-ME6 plasmid DNA control; Parental strain, *C. acetobutylicum pyrE*-ve strain (CRG CC#1545 Table 2.1) genomic DNA control; and 1–6, candidate clone. All six candidates showed the expected products of 1989bp and 1936bp for *pyrE* mutant confirmed by nucleotide sequencing (Data not shown).

3.3.6 Exemplification of ACE technology at the *thiolase* locus as the target and plasmid pMTL-JH16

In the previous Sections 3.3.4 and 3.3.5, the utility of ACE technology was demonstrated using the *pyrE* locus as the target by deployment of two different dysfunctional *pyrE* alleles on the plasmid used to create the *pyrE* mutant (pMTL-JH12) and thereafter, a plasmid used to correct the mutation (pMTL-JH14 or pMTL-ME6). In this section an alternative approach is demonstrated,

where the plasmid-borne allele is a promoter-less antibiotic resistance gene and the chromosomal allele is merely an active promoter. By coupling the chromosomal promoter to the promoter-less antibiotic resistance gene, the latter is activated conferring resistance to the antibiotic on the cell. To test this approach, plasmid pMTL-JH16 (section 2.5.1.7) was designed. It carries a macrolide–lincosamide–streptogramin (MLS)-resistant *ermB* structural gene, but is devoid of its promoter. The integration site was targeted to a position downstream of the *thiolase* gene which is known to possess a constitutively, highly active promoter (Harris et al., 2002; Tummala et al., 1999). In the case of pMTL-JH16, the host strain must be sensitive to *ermB*. Such a host strain may be the wild-type, or may have been previously constructed using a suitable derivative of pMTL-JH15 to select for the second recombination event.

To test the approach, plasmid pMTL-JH16 was transformed into *C. acetobutylicum* wild-type and transformants were selected on CGM supplemented with Tm. The preferred integration of plasmid DNA into the chromosome is mediated by the recombination event between the long homology arm (LHA) on the plasmid and the same corresponding region on the chromosome downstream of the *thiolase* gene (Figure 3.12). During this process, cells in which the first homologous recombination event has occurred are enriched on medium supplemented with Tm comparative to cells carrying autonomous plasmid due to the replication defective nature of the pIM13 replicon on the plasmid. Under these circumstances the growth rate of cells carrying the *catP* marker on the plasmid is restricted compared to those cells in which an integration event has introduced the *catP* gene into the chromosome.

This growth rate difference provides the basis for enrichment of cells in which the plasmid integration has occurred.

Six of the largest thiamphenicol resistant isolated colonies were sub-cultured onto fresh plates of the same growth medium (passage P1) after forty-eight hours of incubation. All these clones were treated independently in all subsequent steps of the procedure.

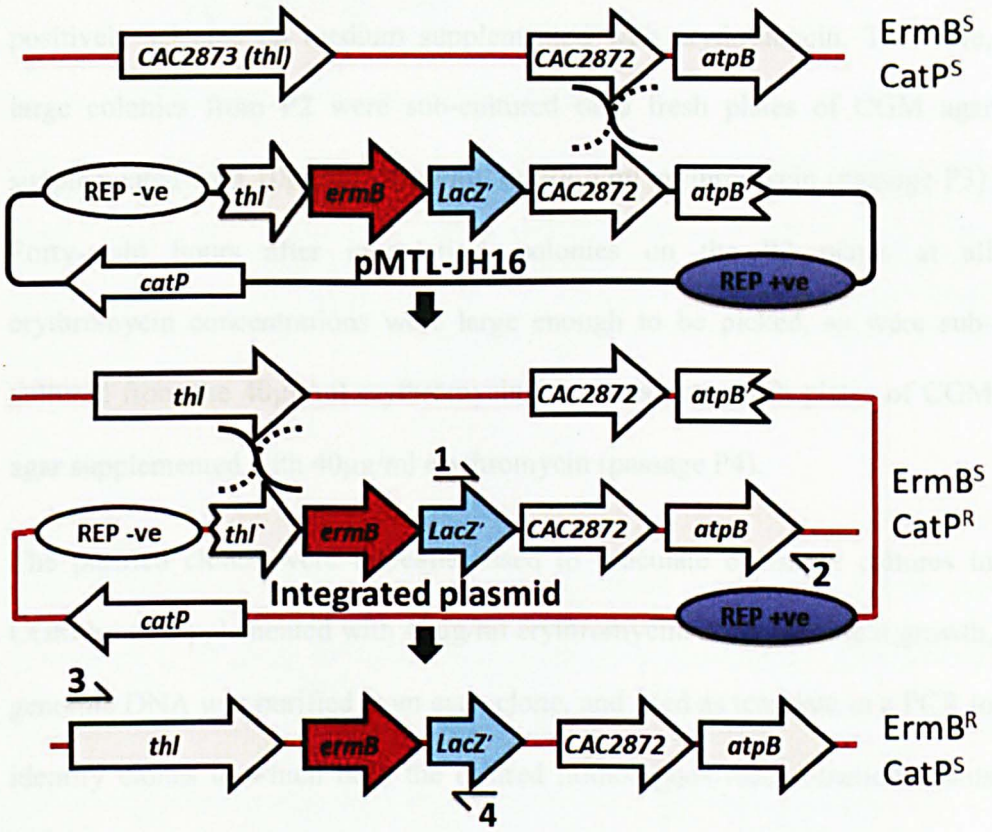


Figure 3.12 Allelic exchange at promoter level using pMTL-JH16

Plasmid integration is mediated via the long homology arm (LHA) on the plasmid by recombining with the corresponding region on the chromosome downstream of the *thiolase* gene and was confirmed by PCR using primers *lacZalpha*-sF2 (1) and Cac-atpB-sR1 (2) Plasmid excision is directed by the recombination between the short homology arm (SHA) corresponding to the last 300bp of *thl* ORF on the plasmid, and the equivalent region of *thl* ORF on the host cell chromosome and double cross-over integrants were screened by PCR using primers Cac-thl-sF1 (3) and M13F (4). This second allelic exchange event activates the marker to be selected positively on medium supplemented with erythromycin. ErmB^R, erythromycin-resistance; ErmB^S, erythromycin-sensitive; CatP^R, thiamphenicol/chloramphenicol-resistance; CatP^S, thiamphenicol/chloramphenicol-sensitive

Twenty-four hours later, colonies on the P1 plates were large enough to be picked, and were sub-cultured onto fresh plates of the same growth medium (passage P2). The second homologous recombination event is directed by the short homology arm (SHA) corresponding to the last 300bp of the *thl* ORF, and the equivalent region of the full-length the *thl* ORF originating from the host cell chromosome. During a second phase of the allelic exchange procedure, cells in which a second homologous recombination event has occurred are positively selected on medium supplemented with erythromycin. Therefore, large colonies from P2 were sub-cultured onto fresh plates of CGM agar supplemented with 10μg/ml, 20μg/ml, or 40μg/ml erythromycin (passage P3). Forty-eight hours after inoculation, colonies on the P3 plates at all erythromycin concentrations were large enough to be picked, so were sub-cultured from the 40μg/ml erythromycin P3 plates onto fresh plates of CGM agar supplemented with 40μg/ml erythromycin (passage P4).

The purified clones were thereafter used to inoculate overnight cultures in CGM broth supplemented with 40μg/ml erythromycin. After overnight growth, genomic DNA was purified from each clone, and used as template in a PCR to identify clones in which both the desired homologous recombination events had occurred. In a first "Long Homology Arm crossover" PCR the primers used were *lacZalpha*-sF2 which binds to the plasmid *lacZalpha* sequence in the forward orientation; and Cac-atpB-sR1 which binds to the *C. acetobutylicum* ATCC 824 chromosome downstream of the long homology arm in the reverse orientation (as shown in Figure 3.12). This pair of primers amplified as expected a PCR product of 1450bp from genomic DNA of cells in which a

homologous recombination event has occurred at the long homology arm (Figure 3.13).

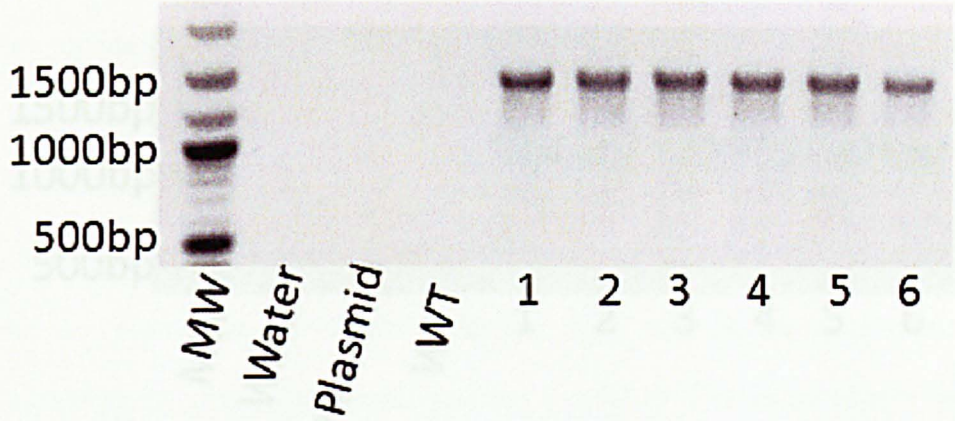


Figure 3.13 LHA PCR screening using pMTL-JH16

PCR screening of six candidate double-crossover clones using primers *lacZalpha*-sF2 and Cac-atpB-sR1 that anneal to single-crossover chromosome where indicated in (Figure 3.12). MW, 2-Log DNA Ladder (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL-JH16 plasmid DNA control; WT, wild-type *C. acetobutylicum* genomic DNA control; 1–6, candidate clone. All six candidates showed the expected band of 1450bp verified by nucleotide sequencing (Data not shown).

In a second "Short Homology Arm (SHA) crossover" PCR , the primers used were M13F which binds to the plasmid *lacZalpha* sequence in the reverse orientation; and Cac-thl-sF1 which binds to the *C. acetobutylicum* chromosome upstream of the short homology arm in the forward orientation. This pair of primers amplified as expected a PCR product of 1250bp from genomic DNA of cells in which a homologous recombination event has occurred at the short homology arm (Figure 3.14). The PCR products of one double-crossover clone were verified by nucleotide sequencing which confirmed the allele exchange. Finally, the same clone was shown to be thiamphenicol-sensitive by replica-plating single colonies onto plates of CGM agar supplemented with 40μg/ml erythromycin and plates of CGM agar supplemented with 15μg/ml Tm. Thiamphenicol-sensitive phenotype suggested that the plasmid sequence

containing the *catP* gene is not present in the cell and were confirmed by a PCR using primers CatP-F and CatP-R (Data not shown).

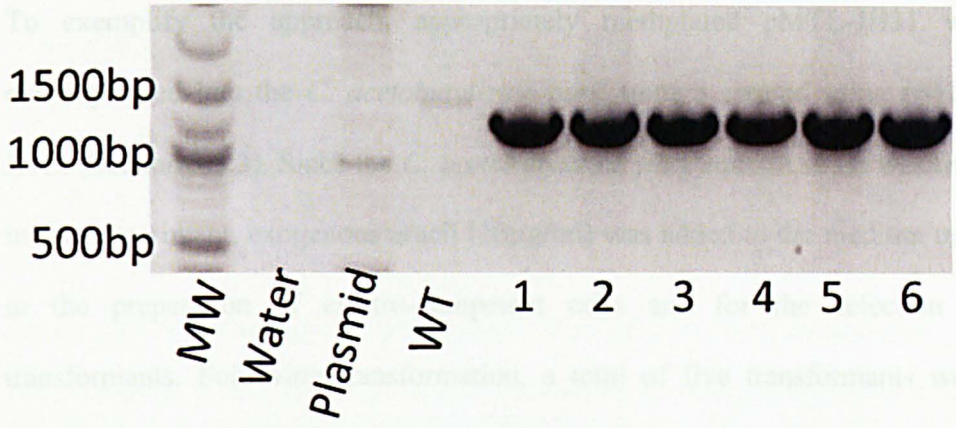


Figure 3.14 SHA PCR screening using pMTL-JH16

PCR screening of six candidate double-crossover clones using primers M13F and Cac-thl-sF1 that anneal to double-crossover chromosome as indicated in (Figure 3.12). MW, 2-Log DNA Ladder (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL-JH16 plasmid DNA control; WT, wild-type *C. acetobutylicum* ATCC 824 genomic DNA control, 1–6, candidate clone. All six candidates showed the expected band of 1250bp verified by nucleotide sequencing (data not shown).

3.3.7 Exemplification of ACE technology at the *thiolase* locus as the target and plasmid pMTL-JH31

In the preceding study, a promoter-less macrolide–lincosamide–streptogramin (MLS) antibiotic resistance marker *ermB* was coupled to a chromosomally located *thl* promoter using the vector pMTL-JH16 via homologous recombination events. To test another marker, an orthologous *pyrE* gene from *C. sporogenes* ATCC 15579 was selected. It shares only 47.6% identity to the *C. acetobutylicum pyrE* gene, and is, therefore, unlikely to undergo homologous recombination with the native gene when present on an introduced ACE plasmid. The *pyrE* gene (CLOSPO_00385) from *C. sporogenes* was

ligated between the two homology regions in pMTL-JH16 and replaced the *ermB* marker. The plasmid made was designated pMTL-JH31.

To exemplify the approach, appropriately methylated pMTL-JH31 was electroporated into the *C. acetobutylicum pyrE* mutant created using pMTL-JH12 (Section 3.3.3). Since the *C. acetobutylicum pyrE* mutant strain was used in this experiment, exogenous uracil (20μg/ml) was added to the medium used in the preparation of electro-competent cells and for the selection of transformants. Following transformation, a total of five transformants were selected from those that grew on CGM agar supplemented with 15μg/ml Tm and 20μg/ml uracil.

To enrich single crossover integrants, these were passaged twice on CGM supplemented with 15μg/ml Tm and 20μg/ml uracil. It was assumed that the majority of the clones represented cells in which integration had taken place via the long region of homology without affecting the *pyrE* mutant. The short region of homology (*ie.*, the last 300bp of the *thiolase* ORF) would be expected to mediate the excision of the plasmid and place the 'silent' (lacking a promoter and therefore not expressed when on the plasmid) *pyrE* gene immediately downstream of the strong *thl* promoter located on the chromosome, leading to its co-expression with *thl* and allowing these double-crossover clones to be selected on CBM lacking uracil. All five candidate clones were authenticated by showing that an expected 1101bp DNA fragment was generated in a diagnostic PCR (Figure 3.15), and that the nucleotide sequence of the fragment was exactly as predicted.

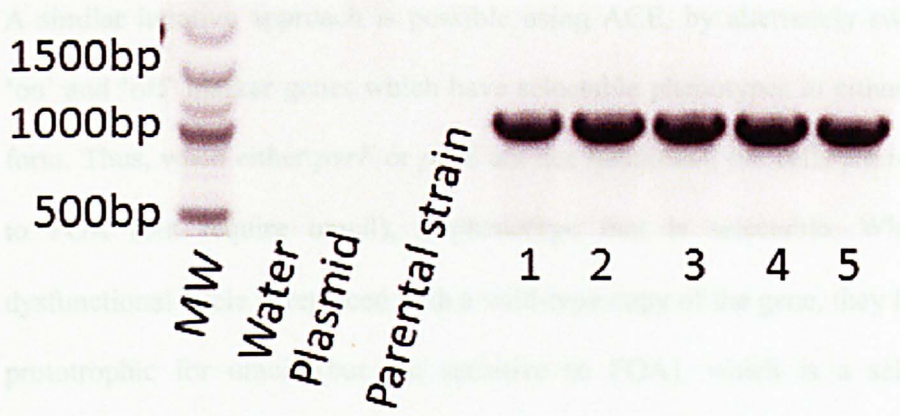


Figure 3.15 PCR screening for pMTL-JH31-derived double crossover events

PCR screening of five candidate double-crossover clones using primers Cac-thl-sF1 and M13F that anneal as indicated in (Figure 3.12). MW, 2-Log DNA (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL-JH31 plasmid DNA control; Parental control, *C. acetobutylicum pyrE* mutant (CRG CC#1545) genomic DNA control; 1–5, candidate clones.

3.3.8 Integration of the phage lambda genome into the chromosome

The experiments till this point had demonstrated that integration of ca. 800bp of heterologous DNA, as well as larger fragments of ca. 6.5kb into the *C. acetobutylicum* ATCC 824 chromosome was possible, but there must be limits to the size of DNA that can be integrated into chromosome in a single step. These limitations would be imposed by many factors and, for *C. acetobutylicum*, could be frequency of transformation which is low even for small plasmids (Heap et al., 2009). Sequences larger than the optimum length can be integrated in multi-steps using overlapping sub-fragments of the desired sequence. Such an approach has previously been reported, with the introduction of the genomes of two different organelles (the 16.3kb mouse mitochondrial genome and the 134.5kb rice chloroplast genome) into the chromosome of *B. subtilis* using the so-called 'domino' method (Itaya et al., 2008).

A similar iterative approach is possible using ACE, by alternately switching 'on' and 'off' marker genes which have selectable phenotypes in either allelic form. Thus, when either *pyrF* or *pyrE* are not functional, the cells are resistant to FOA (but require uracil), a phenotype that is selectable. When the dysfunctional allele is replaced with a wild-type copy of the gene, they become prototrophic for uracil (but not sensitive to FOA), which is a selectable phenotype (Figure 3.1 and Figure 3.5).

Alternatively, it is possible to alternate between two different promoter-less selectable markers, first activating the one marker through its insertion downstream of a strong promoter, and then replacing it with a second promoter-less marker (Figure 3.14 and Figure 3.15). In either scheme, the short region of homology would remain the same in every step, whereas the long region of homology would be identical to the end of the previous insert in the second and subsequent steps. Therefore, to exemplify the use of ACE in the iterative insertion or large amounts of DNA, it was decided to integrate the whole lambda DNA in three steps into the chromosome of the *C. acetobutylicum pyrE* mutant. Three unique restriction fragments were identified in a circular form of phage lambda genome and ligated into ACE vectors as described in section 2.5.1.9.

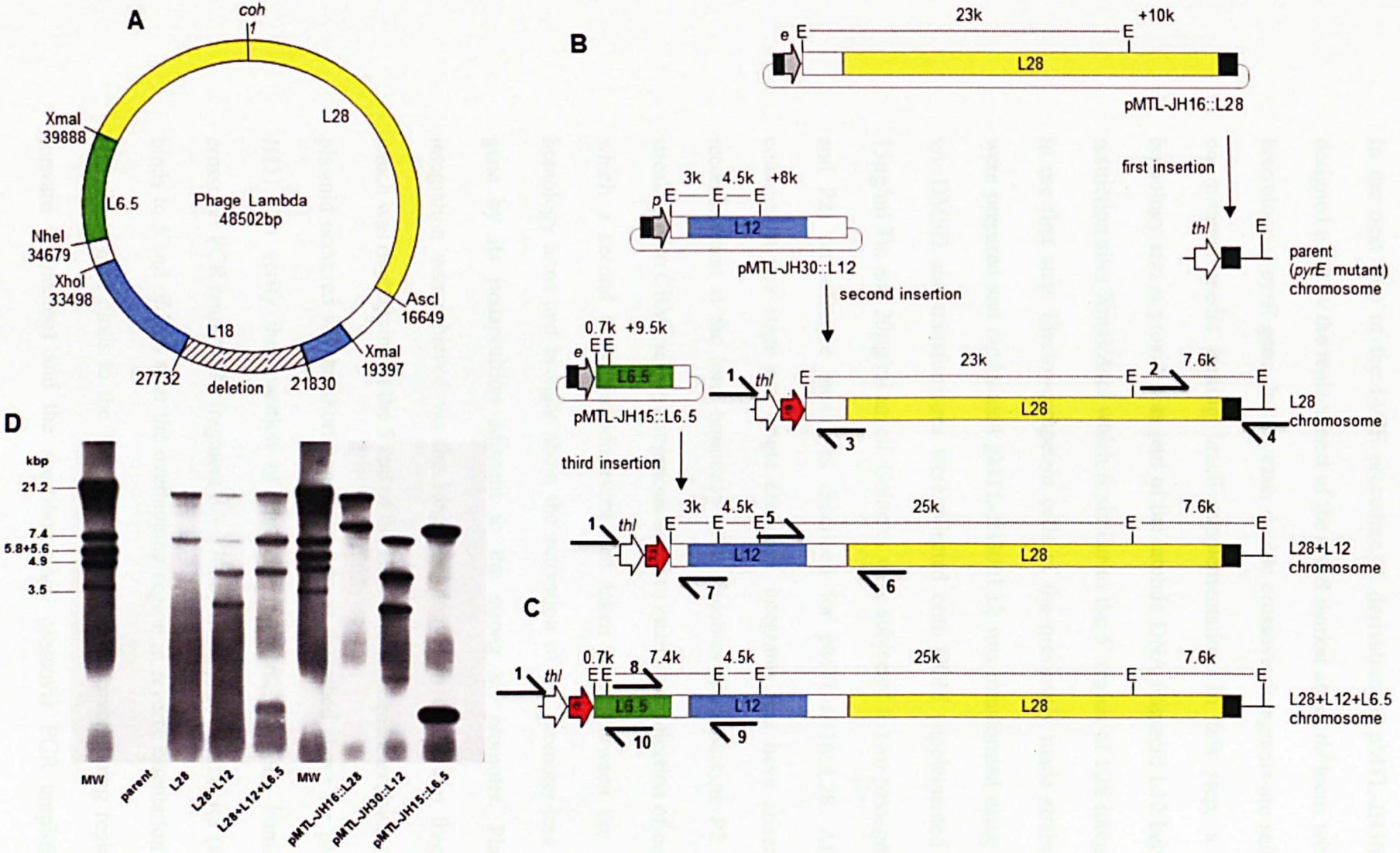
For the integration of the first lambda DNA fragment (L28), electro-competent cells of the *C. acetobutylicum pyrE* mutant strains were prepared as described above and just before electroporation of methylated pMTL-JH16::L28 DNA, they were incubated with 10% v/v Dimethyl sulfoxide (DMSO) to enhance the permeability of the cell membrane (de Menorval et al., 2012) and facilitate the electroporation of this larger DNA molecule. Transformants were recovered

with overnight incubation in 2xYTG supplemented with 7.5μg/ml Tm and 20μg/ml uracil and plated on CGM media supplemented with 7.5μg/ml Tm and 20μg/ml uracil. After seventy-two hours, colonies were large enough to re-streak on the same medium (passage P1). After forty-eight hours, colonies were re-streaked again on the same medium (passage P2). At this stage, genomic DNAs were extracted from overnight cultures and integration of the plasmid into the host chromosome via the long homology arm was confirmed using PCR primers DNA L28_F1 and Cac-atpB-sR1 as shown in Figure 3.16 (data not shown). After forty-eight hours, clones were streaked onto CGM supplemented with 40μg/ml erythromycin and 20μg/ml uracil (passage P3), and then again forty-eight hours later (passage 4, P4). Putative integrants were then inoculated in 1ml of 2xYTG supplemented with 40μg/ml erythromycin and 20μg/ml uracil and incubated anaerobically overnight. Cell growth was harvested by centrifugation and genomic DNA extracted from cell pellets and the integration of the 28Kb lambda DNA fragment confirmed by PCRs using primers DNA L28_R1 and Cac-thl-sF1 as shown in Figure 3.16 that flanked the predicted junction points of the inserted lambda DNA fragment. Accordingly, a fragment of ~1900bp was amplified using a forward primer from 3`end of the lambda DNA L28_F1 and a reverse primer form the chromosome Cac-atpB-sR1 to confirm integration of the plasmid into the chromosome via the long homology arm. Excision of the plasmid via the short homology arm was confirmed by the generation of a PCR fragment of ~1500bp using a reverse primer at 5`end of the lambda DNA L28_R1 and a forward primer from the chromosome Cac-thl-sF1 (data not shown). The recombinant

strain obtained was named L28, and became the parental strain used in the next step of the *iterative* ACE (iACE) procedure.

Figure 3.16 Integration of Phage lambda genome using iACE

Multistep insertion of the chromosome of phage lambda into the chromosome of *C. acetobutylicum*. (A) Chromosome of phage lambda showing restriction sites used to excise the three overlapping fragments. Yellow, 28 kb XmaI-XmaI fragment L28; blue, 18 kb NheI-AscI fragment L18; green, 6.5 kb XmaI-XhoI fragment L6.5; white, regions of overlap between fragments; cross-shaded deletion, 6 kb region of L18 absent from L12; coh, ligated cohesive ends. (B) Integration plasmids pMTL-JH16::L28, pMTL-JH30::L12 and pMTL-JH15::L6.5. The lambda sequences are coloured as in (A). Grey arrows e and p, inactive (non-expressed) *ermB* and *pyrE*, respectively; black, homology to *thl* locus; E, EcoRI sites; dashed lines, EcoRI fragments (except those spanning the plasmid backbone); numeric labels, EcoRI fragment lengths in kilo base pair. (C) The *thl* locus of *C. acetobutylicum* before and after 1, 2 and 3 insertions of lambda DNA. Elements are labelled as in (B). The first recombination event at each step, indicated, is directed by a long region of homology. A short region of homology mediates plasmid excision and simultaneously activates *ermB* or *pyrE* in alternate steps, shown by red arrows e and p, respectively, by positioning them under the control of the chromosomal *thl* promoter. (D) Southern blot of EcoRI digests of the plasmids and chromosomes shown in (B) and (C), using lambda DNA as probe. MW, EcoRI-digested lambda DNA molecular weight marker; Parent, *C. acetobutylicum pyrE* mutant (CRG CC# 1545) Primers Cac-thl-sF1 (1), DNA L28_F1 (2), DNA L28_R1 (3), Cac-atpB-sR1 (4), L12_F2 (5), L28_R3 (6), L12_R2 (7), L6.5_F2 (8), L12_R4 (9), L6.5_R2 (10)


A
coh
1
L28
XmaI
39888
L6.5
Phage Lambda
48502bp
NheI
34679
XhoI
33498
AscI
16649
L18
27732
deletion
21830
XmaI
19397
B
23k
+10k
e
E
E
L28
pMTL-JH16::L28
first insertion
3k
4.5k
+8k
p
E
E
E
L12
pMTL-JH30::L12
thl
E
parent
(pyrE mutant)
chromosome
second insertion
0.7k
+9.5k
e
E
E
L6.5
pMTL-JH15::L6.5
1
thl
E
23k
E
2
7.6k
E
L28
chromosome
3
4
third insertion
3k
4.5k
E
E
E
5
25k
E
7.6k
E
1
thl
E
L12
L28
L28+L12
chromosome
7
6
C
0.7k
7.4k
4.5k
25k
7.6k
1
thl
E
E
8
E
E
E
E
L6.5
L12
L28
L28+L12+L6.5
chromosome
10
9
D
kbp
21.2
7.4
5.8+5.6
4.9
3.5
MW
parent
L28
L28+L12
L28+L12+L6.5
MW
pMTL-JH16::L28
pMTL-JH30::L12
pMTL-JH15::L6.5

In the next step of the iACE procedure, a derivative of pMTL-JH30 was designed to allow the replacement of the *ermB* marker at the *thl* locus with the heterologous *pyrE* gene. In this case, double crossover integrants are selected on minimal media lacking uracil supplementation. In this step, a long homology arm is provided as part of the lambda DNA fragment L12 between restriction sites XmaI/AscI which is similar to the 5` region of L28 integrated in the first step. Electro-competent cells of the previously made strain L28 were prepared and methylated pMTL-JH30::L12 was transformed using 10% v/v DMSO and transformants were selected onto CGM supplemented with 15μg/ml Tm and 20μg/ml uracil. Colonies were subjected to two passages (P1 and P2) on selective media, as described for pMTL-JH16::L28. All the colonies at this stage are single crossover integrants that have arisen via recombination at the long homology arm. Colonies from passage P2 were streaked onto CBM lacking exogenous uracil to enable the selection of cells in which a second recombination event had taken place between the short homology arms and brought about the activation of the promoter-less *pyrE* gene by its juxtaposition adjacent to the strong *thl* promoter. Plasmid integration was achieved via the long homology arm of 2700bp fragment which was overlapping to the 5`end of the L28. The subsequent excision of the plasmid occurred via the short homology arm as described above for pMTL-JH31. To verify the insertion of the L12 fragment, the long homology crossover PCR amplified a fragment of ~3700bp using primers L28_R3 (which binds to 5`end of L28 after the overlapping region in reverse orientation) and L12_F2 (which binds to the 3`end of L12 before the overlapping region in forward orientation) and the short-homology crossover PCR amplified a

fragment of ~1500bp using Cac-thl-sF1 (binds to upstream of the integration site on the chromosome) and L12_R2 (which binds to 5`end of the L12 fragment in the reverse orientation) and were verified by nucleotide sequencing (data not shown). The resultant strain was called L28-L12 and was confirmed to be sensitive to erythromycin (due to replacement of *ermB* by *pyrE*) and thiamphenicol sensitive, due to loss of the ACE plasmid that carried the *catP* marker gene. Sensitivity of the strain L28-L12 to both antibiotics erythromycin and thiamphenicol was crucial to be able to use it as a parental strain in the final iteration of ACE.

In the final step of the iACE procedure, the third lambda DNA fragment L6.5 was cloned in pMTL-JH15 which carries a copy of *ermB* that is destined to replace the heterologous *pyrE* in strain L28-L12 located downstream of the *thiolase* locus. L6.5 carried a long homology arm between the restriction sites NheI/XhoI which is homologous to the 5`end of the integrated fragment L12. Electro-competent cells were prepared and methylated pMTL-JH15::L6.5 was transformed into L28-L12 strain and transformants were selected on CGM supplemented with 15μg/ml Tm. Since this strain, L28-L12, now has an active *pyrE* gene, no exogenous uracil was required for selection of transformants. After forty-eight hours, colonies were big enough to re-streak allowing passages P1 and P2 onto CGM supplemented with 15μg/ml Tm. At this stage the majority of the colonies were likely to be single crossover integrants due to recombination at the long homology arm (~1200bp) of L6.5 which overlaps with the 5`end of the L12 Lambda fragment. Colonies from P2 were streaked onto CGM supplemented with 40μg/ml erythromycin and 20μg/ml uracil to select for those cells in which a second recombination event has occurred

between the short homology arms leading to activation of the *ermB*. As the chromosomally located *pyrE* is replaced with *ermB*, exogenous uracil was added to the medium as the desired recombinant strain will now be auxotrophic for uracil. Colonies were streaked once more on the same medium and genomic DNA was extracted from overnight cultures and the integration of the L6.5 fragment was confirmed by appropriate PCRs that amplified the junction points in the predicted integrants, as described below.

To demonstrate that integration of plasmid via the long homology arm had occurred, a fragment of ~2300bp was amplified using primers L6.5_F2 (which binds to the 3`end of the L6.5 fragment before the overlapping region in the forward orientation) and L12_R4 (which binds to the 5`end of the L12 fragment after the overlapping region in the reverse orientation). To confirm the excision of plasmid via the short homology arm, a fragment of ~1500bp was amplified using Cac-thl-sF1 (binds to upstream of the integration site on the chromosome) and L6.5_R2 (which binds to the 5`end of the L6.5 in the reverse orientation) (data not shown). These PCR products were confirmed by nucleotide sequencing and the resultant strain was named L28-L12-L6.5.

In summary, using a uracil auxotroph, the first lambda fragment L28 was integrated by selecting double crossover clones that had acquired resistance to erythromycin (*ermB*), the second L12 lambda fragment was integrated by exchanging *ermB* with a heterologous *pyrE* gene that made the cell prototrophic for uracil and the third L6.5 Lambda fragment inserted by swapping the *pyrE* gene with *ermB* and selecting for erythromycin resistance. These experiments demonstrated that iACE could be used to insert the entire phage DNA genome into the *C. acetobutylicum* chromosome at the *thiolase*

locus in three iterative cycles. The authenticity of the final clones were confirmed by undertaking a Southern blot of EcoRI-digested genomic DNA from the three recombinant strains (L28, L28-L12 and L28-L12-L6.5) using lambda DNA as the probe as shown in Figure 3.16 . For comparative purposes, the parental *pyrE* mutant strain and plasmid DNA used for three integrations were also included. All samples showed the expected pattern of fragments (none in the case of the *pyrE* strain control). The authenticity of each insert was also verified by nucleotide sequencing Figure 3.16 .

3.3.9 Attempted integration of pSOL1 into the *C. acetobutylicum* chromosome by reverse ACE (rACE)

Degeneration is a common feature of solventogenic bacterial strains, *ie.*, they stop producing solvents and lose features associated with the normal cell life-cycle, such as sporulation. In the case of *C. acetobutylicum,* degeneration is caused through the loss of the pSOL1 megaplasmid that carries the genes encoding the majority of the enzymes involved in solvent production. Having successfully demonstrated the delivery of the phage genome into the chromosome, it was reasoned that it should be possible to integrate the entire pSOL1 plasmid into the genome, thereby preventing strain degeneration when used in an industrial fermentation process. Theoretically, this could be achieved by cloning pSOL1 into an ACE vector, and then introducing the whole plasmid into the chromosome as described in section 3.3.8. However, it is practically impossible to isolate an intact pSOL1 from *C. acetobutylicum,* clone such a large piece of DNA (192Kb) into an ACE vector, and transform *C. acetobutylicum* to integrate this into the chromosome. An alternative strategy would be to introduce the Long Homology Arm (LHA) and Short

Homology Arm (SHA) into the pSOL1 plasmid in an inverted orientation. The net result would be in effect being an ACE vector carrying pSOL1. In the initial experiments, insertion of the inverted LHA and SHA was attempted using the ClosTron.

Given the size limitation of the cargo DNA that can be delivered using the ClosTron system (Chen et al., 2007; Heap et al., 2010b; Plante & Cousineau, 2006), it was decided to both use a ClosTron variant lacking the *ermB* RAM and to reduce the overall size of the homology arms to a total of ca. 1500bp, by reducing the size of LHA to 900bp.

3.3.9.1 Identification of a suitable ClosTron target

As envisaged the ClosTron would not carry the selectable *ermB* marker, a target was needed that would result in a screenable phenotype when the group II intron, together with the inverted LHA and SHA, had inserted. A gene encoding amylase activity was chosen, as inactivation would result in colonies that failed to form halos in the presence of iodine on media containing starch.

3.3.9.2 Amylases CaP0098 and CaP0168

C. acetobutylicum pSOL1 contains two genes annotated as encoding α-amylase genes [Cap0098 (1686bp) and Cap0168 (2283bp)] (Nolling et al., 2001). Loss of the megaplasmid leads to strain degeneration and loss of the ability to form clear zones on media containing starch (Sabathe et al., 2002).

To determine which of the two amylase genes is principally responsible for starch metabolism, and zone formation on media containing starch, three ClosTron target sites (484a, 564s and 1039a) were selected for Cap0098. The necessary re-targeted pMTL007C-E2 plasmids were ordered from DNA2.0 and

used to make ClosTron mutants according to the previously described procedure (Heap et al., 2010b; Heap et al., 2007). PCR primers AM1_2-Fw and AM3_Rev flanking the insertion sites of CaP0098 gene were employed to screen for the mutants as shown in Figure 3.17 .

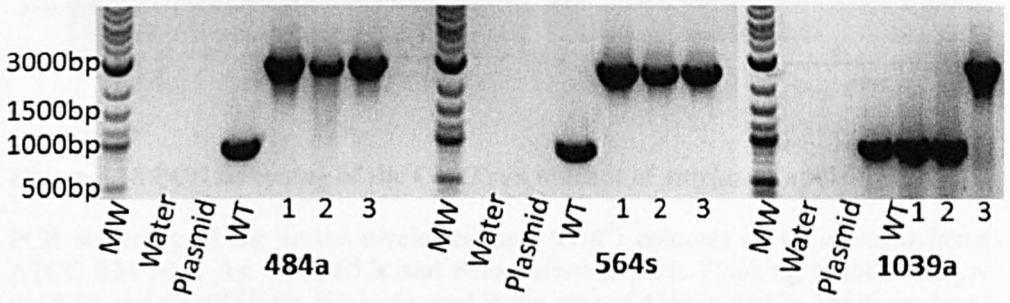


Figure 3.17 PCR screening of the ClosTron mutant of *amylase* Cap0098

PCR screening of the three *ermB* resistant colonies of *C. cetobutylicum* ATCC 824 each for Cap0098-484a, Cap0098-564s and Cap0098-1039a insertion sites using flanking primers AM1_2-Fw and AM3_Rev. PCR produced 897bp and 2697bp fragments for wild type and intron insertion respectively and verified by nucleotide sequencing. MW, 2-log DNA marker (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL007C-E2-Cap0098-484a, pMTL007C-E2-Cap0098-564s and pMTL007C-E2-Cap0098-1039a plasmid DNA control; WT, wild-type *C. cetobutylicum* ATCC 824 genomic DNA control and 1-3 clones for each target sites.

The intron insertional mutants of Cap0098 as well as wild-type strain were grown overnight in 2xYTG and 10µl from each were spotted on CBM 2% w/v starch and 0.5% w/v glucose and incubated for 48hrs in anaerobic conditions and followed by staining with iodine (Sabathe et al., 2002). No difference was observed among mutants and wild-type in terms of halo formation (data not shown).

Similar to CaP0098, three target sites were selected for Cap0168 as 456s, 657s and 918s and the appropriate re-targeted pMTL007C-E2 vectors ordered from DNA2.0 and ClosTron mutants were made (Heap et al., 2010b; Heap et al.,

2007). In the case of Cap0168, PCR primers flanking the insertion sites were employed to screen for the mutants as show in Figure 3.18 .

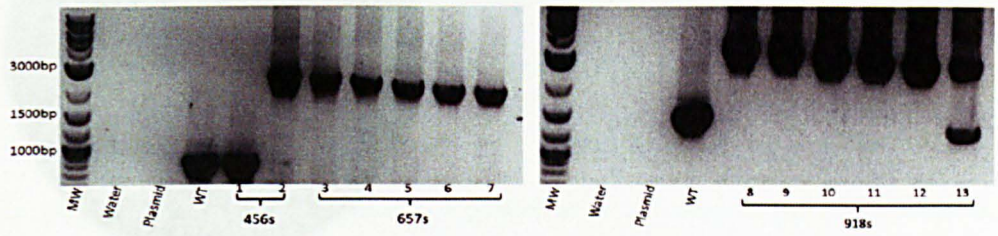


Figure 3.18 PCR screening of the ClosTron mutant of *amylase* Cap0168

PCR screening of the erythromycin-resistant (*erm*^{R}) colonies of *C. acetobutylicum* ATCC 824 *pyrE* for 456s, 657s and 918s insertion sites. Flanking primers Cac_p-0168-F1 and Cap0168-Int-sF2 were used in the case of 456s and 657s which produced 800bp for WT and 2600bp for ClosTron mutants while primers Cac_p-0168-F1 and Cap0168-Int-sF1 were used to screen for 918s that amplified 1478bp and 3278bp fragments for wild type and intron insertion respectively. PCR products were confirmed by nucleotide sequencing (data not shown). MW, 2-log DNA marker (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL007C-E2-Cap0168-456s, pMTL007C-E2-Cap0168-657s and pMTL007C-E2-Cap0168-918s plasmid DNA control; WT, wild-type *C. acetobutylicum* ATCC 824 genomic DNA control.

The intron insertional mutants of Cap0168 as well as *C. acetobutylicum* ATCC 824 *pyrE* strain (CRG CC #1545) were grown overnight in 2xYTG supplemented with 20µg/ml uracil 10µl from each culture was then spotted onto CBM agar containing 2% w/v starch, 0.5% w/v glucose and supplemented with 20µg/ml uracil following incubation for 48hrs under anaerobic conditions before staining with iodine (Sabathe et al., 2002).

A clearance of zones *ie.*, halo formation was observed in the case of *C. acetobutylicum* 824 *pyrE* mutant strain (CRG CC#1545) but was significantly reduced in the ClosTron mutants (Figure 3.19) which confirmed that Cap0168 is responsible for starch degradation in *C. acetobutylicum* ATCC 824.

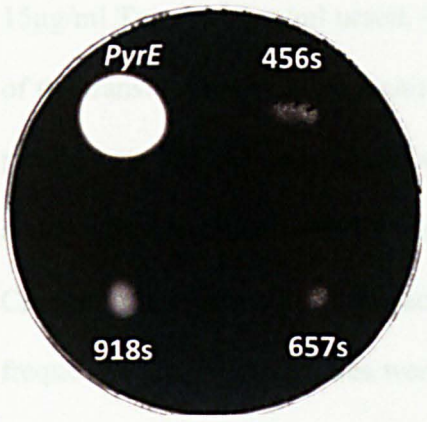


Figure 3.19 Phenotype of ClosTron mutants of *amylase* Cap0168

PyrE; *C. acetobutylicum* 824 *pyrE* (CRG CC#1545); ClosTron mutant of Cap0168 at 456s, 657s, and 918s sites spotted on CBM containing 2% w/v starch, 0.5% w/v glucose and supplemented with 20μg/ml uracil and incubated for 48 hrs at 37 °C in anaerobic conditions before staining with iodine.

3.3.9.3 Identification of the most effective ClosTron target against Cap0168

As the ClosTron to be used to deliver the inverted ACE homology arms would not possess an *ermB* gene, and therefore cannot be selected for, it was important to identify the most effective retargeted ClosTron so as to maximise the chances of finding an amylase minus ClosTron insertion in the subsequent mutant screen. Therefore, the effectiveness of the three re-targeted plasmids in generating an *amylase* minus mutant was calculated in terms of the number of colonies acquiring *ermB* resistance and confirmation of intron insertions by PCR (Figure 3.20).

The three re-targeted ClosTron plasmids pMTL007C-E2-Cap0168-456s, pMTL007C-E2-Cap0168-657s and pMTL007C-E2-Cap0168-918s were transformed in *C. acetobutylicum* 824 *pyrE* mutant (CRG CC#1545 Table 2.1) and transformation mixtures were plated onto CGM supplemented with

15µg/ml Tm and 20µg/ml uracil. Three transformants were selected from each of the transformation mixture and dilutions were prepared in PBS followed by plating onto CGM supplemented with 15µg/ml Tm and 20µg/ml uracil as well as onto CGM supplemented with 10µg/ml erythromycin and 20µg/ml uracil. CFU/ml were calculated in each case (data not shown) and integration frequencies of the target sites were calculated as per cent of the total CFU/ml of thiamphenicol-resistance colonies. The highest number of *erm*^{R} colonies was obtained for target site 918s as 0.12% of total CFU/ml and was selected for intron insertion using RAM-less ClosTron plasmid to integrate the inverse homologous sequence at the amylase locus Cap0168 in Section 3.3.9.4.

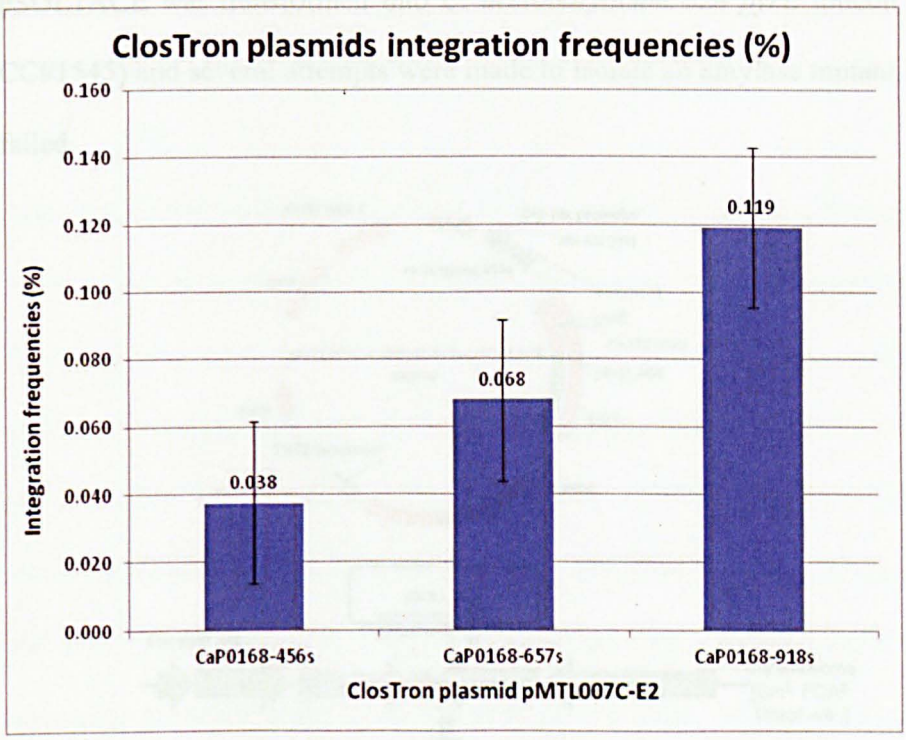


Figure 3.20 ClosTron plasmids integration frequencies (%)

The three target sites 456s, 657s and 918s of Cap0168 for intron insertions are shown on x-axis and number of *ermB*^{R} colonies are shown on y-axis as percent CFU/ml. Error bar are mentioned in percent as Standard deviation divided by the number of samples *ie.*, 3 for each target site denoted on each bar.

3.3.9.4 Attempted integration of inverse sequence into pSOL1 using ClosTron

Having identified the most effective targeting sequence for the insertion of the group II intron into *amyB*, attempts were made to use it to introduce the inverted LHA and SHA regions into pSOL1. The retargeting region selected was 918s. An inverse SHA:LHA of ca.1500bp for the *pyrE-hydA* loci flanked by MluI sites was designed and synthesized by DNA2.0 Inc, and inserted into ClosTron plasmid pMTL007C-E2-Cap0168-918s digested with MluI to liberate the *ermB* RAM marker to allow replacement with the inverse homology arm sequence. The resultant vector pMTL007C-Cap0168-918s-pSOL1ACE was transformed into *C. acetobutylicum* 824 *pyrE* mutant (CRG CC#1545) and several attempts were made to isolate an amylase mutant but all failed.

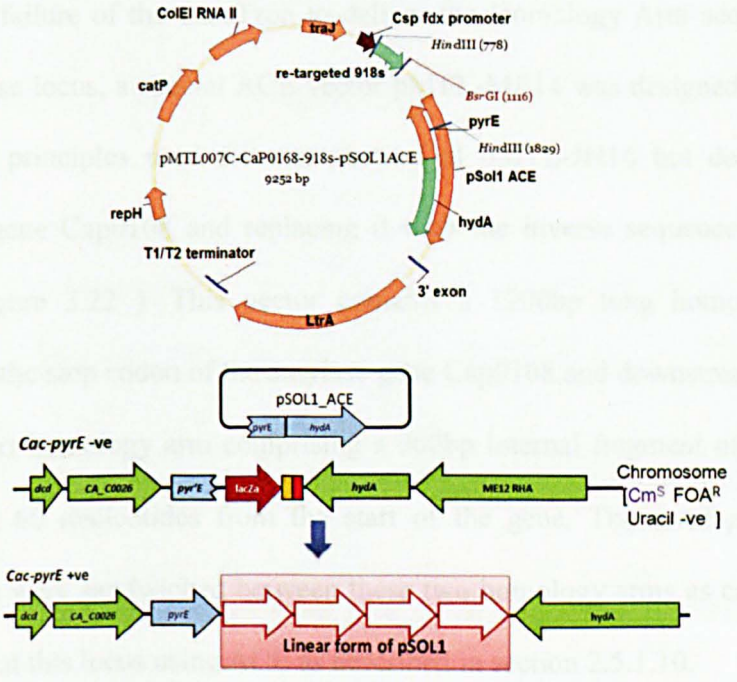


Figure 3.21 Attempted integration of pSOL1 into the chromosome by ClosTron and ACE

A RAM-less ClosTron plasmid was re-targeted to Cap0168-918s site carrying the SHA:LHA inverse cargo DNA to convert the pSOL1 into an ACE like vector. This will then potentially integrates into the *C. acetobutylicum* 824 *pyrE* mutant chromosome at the *pyrE* locus, repair the *pyrE* mutation and integrate the megaplasmid pSOL1 into the chromosome. Cm^S, Chloramphenicol-sensitive; FOA^R, 5-Fluoroorotic Acid-resistant

During screening, individual transformant colonies and diluted cell suspensions were plated onto CBM containing 2% w/v starch and 0.5% w/v glucose to detect the *amyB* mutant phenotype. Individual colonies which showed loss of amylase activity on plates were screened by PCR for the insertion of LHA:SHA inverse sequence but PCR failed to detect the megaplasmid pSOL1 in these cells. Several pool samples were also taken to screen for the mutation by PCR, but no evidence for its presence in the cell population was evident.

3.3.9.5 Attempted integration of inverse sequence into pSOL1 using ACE

After the failure of the ClosTron to deliver the Homology Arm sequences at the amylase locus, a special ACE vector pMTL-ME14 was designed based on the same principles used for construction of pMTL-JH16 but deleting the amylase gene Cap0168 and replacing it with the inverse sequence of *pyrE-hydA* (Figure 3.22). This vector contains a 1200bp long homology arm including the stop codon of the amylase gene Cap0168 and downstream region, and a short homology arm comprising a 300bp internal fragment of Cap0168 beginning 60 nucleotides from the start of the gene. The *ermB-pyrE-hydA* sequences were sandwiched between these two homology arms as cargo to be delivered at this locus using ACE as described in section 2.5.1.10.

Appropriately methylated pMTL-ME14 plasmid DNA was transformed into the *C. acetobutylicum* 824 *pyrE* mutant strain constructed in Chapter 3.3.3

(CRG CC# 1545). Electro-competent cells of *C. acetobutylicum pyrE* mutant strain were prepared transformed with pAN2 methylated pMTL-ME14, and transformants selected on CGM supplemented with 15µg/ml Tm and 20µg/ml uracil. Twenty transformants were selected and streaked twice on the same medium for passages P1 and P2. Plasmid pMTL-ME14 could integrate into the chromosome either into *hydA* or downstream of the *amylase* Cap0168 gene due to the presence of homology arms for both sequences.

To screen for single crossover integrants at the anticipated location *ie.*, amylase Cap0168, overnight cultures were prepared in 1ml of 2xYTG supplemented with 7.5µg/ml Tm, 20µg/ml uracil and after genomic DNA extraction, screened by PCR for pure single cross over integration at the *amylase* locus.

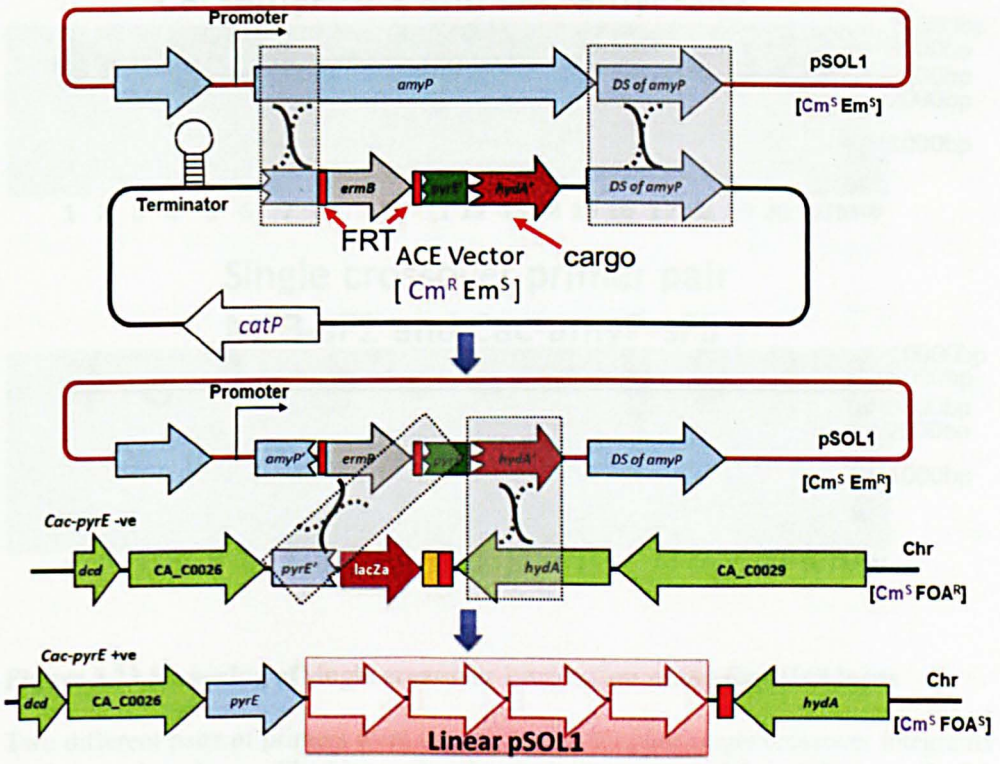


Figure 3.22 Attempted Integration of pSOL1 into chromosome by double ACE procedure

ACE vector pMTL-ME14 targeted to delete the *amyP* gene by activating *ermB* gene and at the same time delivering the inverse cargo DNA *pyrE:hydA* which will convert the megaplasmid pSOL1 into an ACE like vector to be integrated at the *C. acetobutylicum pyrE* locus and revert the cells back to *pyrE* positive by selecting on the minimal medium CBM lacking uracil. FRT, flippase recognition target sites, FOA^R, 5-fluoroorotic acid resistance, FOA^S, 5-fluoroorotic acid-sensitive, Cm^S, thiamphenicol/chloramphenicol-sensitive, Cm^R, thiamphenicol/chloramphenicol-resistance, Em^R, erythromycin-resistance, Em^S, erythromycin-sensitive, pSOL1, *C. acetobutylicum* ATCC 824 megaplasmid pSOL1, Chr, *C. acetobutylicum* ATCC 824 chromosome

Six clones were single crossovers at the *amylase* locus (Cap0168) but only one of them was a pure single crossover clone (clone 14) as shown in Figure 3.23 . The remainder did not produce specific bands and may be single crossover integrants at the *hydA* locus. Clone 14 was stored at -80 °C for the time being and work will resume after submission of this thesis.

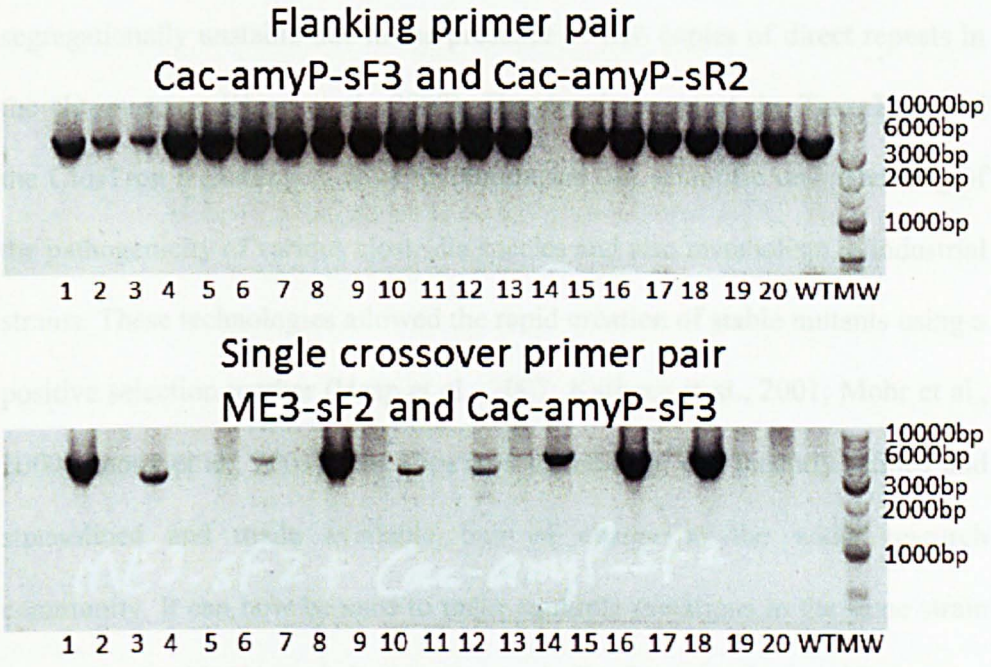


Figure 3.23 Screening of single crossover integration at the Cap0168 locus

Two different pairs of primers were used to screen for pure single crossover integrants at the *amylase* locus. Flanking pair of primer Cac-amyP-sF3 and Cac-amyP-sR2 produced ca. 4500bp DNA bands equal to Wild-type control (*C. acetobutylicum* 824 *pyrE)* for all the clones 1-20 except 14. Primer pair ME3-sF2 and Cac-amyP-sF3 was used to confirm the integration at *amylase* locus. Some of the clones did not produce a band of the appropriate size for this pair of primers ca.5000bp. These could be

integrants at the *hydA* locus as the ACE vector has two allelic exchange cassettes *ie.*, *pyrE/hydA* and *amylase* loci.

3.4 Discussion

Integration of foreign DNA has been a problem in *Clostridium* species due to the lack of availability of sophisticated genetic tools. Therefore recombinant strains were only made by single crossover integration of replicating and non-replicating vectors (Green & Bennett, 1996; Green et al., 1996; Liyanage et al., 2001; Wilkinson & Young, 1994). These methods were very tedious and required laborious screening with a very low reproducibility of results for identification of the mutants. Moreover, the mutants obtained were segregationally unstable due to the presence of two copies of direct repeats in the chromosome (Heap et al., 2007). The development of the TargeTron and the ClosTron technologies have revolutionised our scientific understanding of the pathogenicity of various clostridia species and also metabolism in industrial strains. These technologies allowed the rapid creation of stable mutants using a positive selection marker (Heap et al., 2007; Karberg et al., 2001; Mohr et al., 2000; Zhong et al., 2003). The ClosTron technology was recently refined and streamlined and made available free of charge to the wider research community. It can now be used to make multiple mutations in the same strain (Heap et al., 2010b) and the creation and distribution of a modular vector system (Heap et al., 2009), for general use means that the pool of available mutants is likely to increase in future.

Whilst the ClosTron can be used to deliver synthetic gene (s) to the chromosome, there is a severe limitation on the size of DNA that can be

delivered, equating to no more than 1Kb (Heap et al., 2010b). More effective methods which can deliver large segments of DNA are clearly required, particularly for the successful application of synthetic biology approaches to enhance the properties of commercially useful strains for biofuel production. Linear DNA could be delivered into the bacterial chromosome in one-step where the bacteria are naturally competent and the transformation frequency is very high, such as is the case with *B. subtilis* (Itaya & Tanaka, 1990). Larger amounts of genetic material could be integrated into the chromosome in a multistep procedure through the use of iterative selection markers as in the "domino" method for *Bacillus* (Itaya et al., 2008). Clostridia belong to a group of bacteria which cannot be naturally transformed, and those artificial systems that have been developed (electroporation) do not allow the transfer of linear DNA and in general exhibit a low frequency of plasmid transfer. This precludes the use of non-replicative 'suicide' plasmids, and has necessitated the use of replicative plasmids carrying the desired homologous sequences. All of these factors reduce the frequency of isolation of the desired double crossover integrants. The efficiency can however be improved through the use of negative selection markers located on the vector backbone (Tripathi et al., 2010).

In this study a generally applicable method was developed to deliver DNA to organisms where the required sophisticated gene tools are not available particularly industrially important strains of solventogenic clostridia. Double crossover clones are selected at certain types of genomic loci without using negative/counter selection markers. Crucially, the creation of a novel, selectable marker is brought about by the coupling within the genome of the

two non-functional alleles through a recombination event. This approach was called Allele-Coupled Exchange (ACE). One of its most characteristic features is the use of asymmetric homology arms which allows the sequence of homologous recombination events to be controlled.

For the proof of principle experiments, genes such as *pyrE* and *pyrF* were chosen as selectable phenotypes are available for both the functional and non-functional allele variants of each gene. Both genes are involved in the *de novo* pyrimidine biosynthesis pathway. Functional copies of *pyrE* and *pyrF* can be selected in *pyrE/pyrF* mutant backgrounds by plating on minimal media lacking uracil as their presence restores them to prototrophy. Conversely, the inactivation of either *pyrE* or *pyrF*, through their replacement with a non-functional allele, confers on the cell resistance to FOA 400ug/ml and has been used in various organisms (Boeke et al., 1984; Deng et al., 2009; Husson et al., 1990; Knipfer et al., 1997) including very recently in *C. thermocellum* (Tripathi et al., 2010).

Asymmetric homology arms are used (in contrast to general allelic exchange where equal length homology arms are used that have equal chances of mediating integration into the chromosome) to control the recombination events. A long region of homology of 1200bp directs the integration of plasmid into the chromosome producing the majority of the single crossover integrants initially obtained. Whilst, the long homology arm will be the favoured way by which the integrated plasmid excises, excision via the smaller 300bp homology arm can be detected, as the allele that will result following this excision event can be selected. Thus, when a 300bp internal fragment from *pyrE* or *pyrF* is used, the resulting double cross-over mutant becomes FOA-resistant. Although

either *pyrE* or *pyrF* can be used, the former is preferred as it is the last gene in the *pyrE* operon, and so insertions downstream of the gene are unlikely to have any substantial effects on uracil metabolism. In contrast, the *pyrF* gene is in the middle of an operon. Avoiding downstream polar effects is especially important to later work in this thesis, where *pyrE* is used as the site for complementation studies (Chapter 4).

The created *pyrE* mutant will be used as a parental host to employ a heterologous *pyrE* from *C. sporogenes* ATCC 15579 as a negative selection marker for the creation of allelic exchange mutants elsewhere in the genome. This type of allelic exchange, in which a negatively selectable marker is replaced could be achieved using a conventional allelic exchange cassette (Knipfer et al., 1997) but the *pyrE* allele was specially configured in this present study to facilitate subsequent genetic modification (Figure 3.8). The *pyrE* in these recombinant strains was not deleted completely but only truncated at the 3`end. In a subsequent step of the allelic exchange process using vector pMTL-JH14, this truncated *pyrE* and its partial counterpart were coupled together to achieve a full-length functional gene. Experience has shown that after streaking transformants with ACE vectors twice on medium supplemented with antibiotic, the majority of the clones are single crossover integrants which have always integrated via the long region of homology. In the second step of the recombination event using vector pMTL-JH14, plasmid is excised via the short homology arm resulting in coupling of the two non-functional parts into one functional *pyrE*, and strains carrying *pyrE* can be selected on minimal medium lacking uracil as the cells are now prototrophic for pyrimidine biosynthesis (pages 93 and 96).

The *pyrE* genes in other clostridia, for example *C. difficile* and *C. sporogenes,* have also been exploited during the course of this study and mutants generated by the ACE method could serve as hosts for making in-frame deletions using heterologous *pyrE* gene as a negative selection. In the case of *C. sporogenes*, a stable strain could be constructed by integrating various pro-drug converting enzymes (*eg.*, nitroreductase, NTR) at the *pyrE* locus to create strains with application in cancer therapy (Heap et al., 2012). Moreover, this approach has now been extended to non-clostridial organism such as *GeoBacillus thermoleovorans* which shows that this method has great potential and is widely applicable to other organisms (Lili Sheng unpublished data).

As described above two non-functional *pyrE* sequences were coupled together to make one fully functional gene via homologous recombination events. Here another approach was also used in which a promoter-less selectable marker was fused to a chromosomally located strong promoter to drive its expression and allow phenotypic selection of the double crossover integrants on the basis of acquisition of new phenotypic trait. In the first instance, the use of a macrolide–lincosamide–streptogramin (MLS) antibiotic resistance *ermB* marker was explored and was integrated downstream of constitutively expressed *thiolase* promoter. The second marker used was a promoter-less heterologous *pyrE* gene from *C. sporogenes* ATCC 15579. When used in a *C. acetobutylicum pyrE* mutant, its introduction into the chromosome downstream of the *thl* promoter led to restoration of prototrophy.

Positive selection markers can be re-cycled and have the potential to be used to introduce large amounts of DNA into the chromosome in a multi-step process. In the proof-of-principle experiments, over ~42.5 kb of lambda DNA was

introduced into the *C. acetobutylicum* ATCC 824 chromosome which is the largest arbitrary sequence integrated in any *Clostridium* by an order of magnitude (Chen et al., 2007; Heap et al., 2010b). After each successful transformation of the lambda constructs, double-crossover clones were obtained just as easily as in the earlier experiments using the empty vectors or small cargo DNA less than 10kb. However, several electroporation attempts were made with large quantities of DNA for both pMTL-JH16::L28 and pMTL-JH30::L12 before transformant colonies were observed. Apparently, inserts of this size substantially decrease transformation frequency to a level that reduces the practicality of the procedure. This observation suggests that the size of inserts used in multistep strategies, at least in *C. acetobutylicum* ATCC 824, should strike a balance between the number of steps and the ease of each step. Inserts L28 and L12 appear to be larger than the optimum.

C. acetobutylicum ATCC 824 and other solventogenic strain tend to lose the ability to produce solvents and to form spores if grown in continuous culture. This is known as strain degeneration. In the case of *C. acetobutylicum,* loss of the pSOL1 megaplasmid, that contains all the necessary genes of solvent formation, causes this degeneration. To overcome this issue, and make a stable strain of *C. acetobutylicum,* attempts were made to integrate pSOL1 into the chromosome in a similar way as described for the phage lambda genome. However, in this strategy, rather than clone pSOL1, or parts of it, into an ACE vector, an attempt was made to introduce an inverted copy of the LHA and SHA (*pyrE-hydA* region of the chromosome) into the pSOL1 plasmid itself. Initial efforts focussed on the use of the ClosTron, and targeted the amylase gene CaP0168. However, despite screening many transformants, it did not

prove possible to find a cell in which the ClosTron had inserted. The main issue was lack of selection for the desired intron insertion, because the *ermB* RAM had been removed. This was considered necessary as the efficiency of delivery of DNA cargo is significantly affected by size. Indeed in past work, the PhD candidate showed that retargeting of the ClosTron group II intron could not be detected when the cargo was above 1,000bp (Heap et al., 2010), in addition to the *ermB* RAM. In the present strategy, it was necessary to deliver 1500bp of DNA, necessitating that the *ermB* RAM was removed from the construct. As there was no selection, a gene with an easily screenable phenotype was targeted (amylase, where knockout prevented the formation of a clear halo on starch-iodine plates). However, despite screening more than fifty colonies on starch-iodine plates, no putative knockout clones could be detected. Moreover, even when all transformants were pooled, and DNA prepared, it was not possible to detect the desired intron insertion by PCR. This suggests the efficiency of insertion of the group II intron carrying the homology arms was very low.

Currently another approach is in progress in which a vector with two sets of homology arms is being used. The first set of homology arms is designed to insert the second set of homology arms (those needed to integrate the pSOL1 plasmid into the genome) at the *amylase* locus CaP0168 as shown in Figure 3.22 .

3.5 Key Outcomes

The following are the key outcomes of this chapter:-

1. Allele-Coupled Exchange (ACE) technology was exemplified using either *pyrF* or *pyrE* locus.
2. ACE proved equally effective when targeting the *thiolase* locus, and activating either a promoter-less antibiotic resistance gene (*ermB*), or a heterologous *pyrE* gene in a *pyrE* mutant background.
3. Using ACE, it proved possible to introduce the entire lambda genome into the chromosome of *C. acetobutylicum* in three iterative cycles of the method.
4. It did not prove possible to introduce the pSOL1 megaplasmid into the chromosome using a novel adaptation of the method, but experiments are on-going.
5. The *pyrE* mutant strain generated using ACE laid the foundation for a novel method of making mutants by allelic exchange using a heterologous *pyrE* allele.

Chapter Four: In-frame deletions using negative selection markers

4 In-frame deletions using negative selection markers

4.1 Introduction

Construction of mutants made by allelic exchange using negative selection markers are preferred compared to intron mutagenesis methods because of the fact that no antibiotic marker is used to select the double crossover mutant and multiple mutants can be made in a single strain (Nariya et al., 2011; Zhang et al., 2006). One of the most useful tools available for implementing the necessary procedures used to generate such mutants are negative selection, or counterselection, markers, as discussed in the Introduction (1.10.3). *Pyr* genes have been relatively widely used in microbial systems. The equivalent of *pyrE* in *Saccharomyces cerevisiae* uracil biosynthesis pathway is *URA3/5*. These have been widely used as counterselection markers, and encode orotate phosphoribosyl transferase and orotidine-5'-phosphate decarboxylase respectively, which catalyze the conversion of 5-fluoroorotic acid (FOA, a uracil analogue) into a highly toxic derivative (Boeke et al., 1984). *URA* based counterselection systems have been widely adopted in other organisms using *pyrE* as a negative selection marker for making gene deletions by homologous recombination (Bitan-Banin et al., 2003; Knipfer et al., 1997; Sato et al., 2003) and most recently *pyrF* was used in *C. thermocellum* to knock out a gene encoding phosphotransacetylase (*pta*) (Tripathi et al., 2010).

As discussed in Chapter 3, it proved relatively simple to create a *pyrE* mutant in *C. acetobutylicum* using ACE technology. The construction of the *pyrE*

mutant provided the opportunity to use *pyrE* as a negative selection marker for making in-frame deletions. It was the objective of this chapter to explore the possibility of using a heterologous *pyrE* gene as a negative selection marker to make mutants in *C. acetobutylicum* by allelic exchange. Additionally, during the course of this study, the Clostridia Research Group (CRG) at Nottingham demonstrated that the *codA* gene from *E. coli* could be used as a counterselection marker to make allelic exchange mutants in *C. difficile* (Cartman et al., 2012). The *codA* converts innocuous pyrimidine analogue 5-fluorocytosine (5-FC) to 5-fluorouracil (5-FU) which in the presence of *upp* (uracil phosphoribosyl transferease) is toxic to cells. The development of this system presented the opportunity to perform experiments in parallel with *pyrE* to compare the utility of *codA* as a negative selection marker in gene knockout. In the case of the *pyrE* system, the ability to rapidly restore the chromosomal *pyrE* allele to wild-type, allows the specific in-frame deletion mutant to be characterised in a clean, wild-type background. Moreover, variants of the same vector may be used to deliver the wild-type allele of the deleted gene into the genome. This would allow complementation studies to be performed at an appropriate gene dosage, a considerable advantage over the use of high copy number autonomous plasmids. Moreover, a functional copy of the gene could additionally be delivered under the control of a strong promoter (such as *fdx*), potentially allowing an assessment of the effect of over-expressing the gene.

4.2 Aim of this study

The aim of this study was to utilise the *C. acetobutylicum* 824 *pyrE* mutant (CRG CC#1545) created in Section 3.3.3 using ACE, as the host for making in-frame deletions elsewhere in the chromosome through the use of a

heterologous *pyrE* allele as a negative selection marker. Having demonstrated the feasibility of this approach, a further aim was to demonstrate that in-frame deletions could be complemented by inserting a functional copy of the gene into the chromosome concomitant with correction of the *pyrE* mutation back to wild-type. Over and above complementation, it was also considered useful in parallel to generate a complemented strain in which the gene being complemented was over-expressed, in order to assess the effects of such a manoeuvre on phenotype. To exemplify the latter system, it was desirable to select gene(s) for knockout and over-expression that could be easily assayed both qualitatively and quantitatively. Accordingly, two candidate genes were explored encoding either amylase or pectinase. Finally, it was of interest to explore the utility of the *codA* system that had been developed in the Clostridia Research Group for allelic exchange in *C. difficile.* The specific aims were:-

1. To derive a procedure for making in-frame deletions based on the use of a heterologous *pyrE* allele
2. To demonstrate chromosomal complementation of knockout mutants using ACE-mediated correction of the host *pyrE* mutant allele.
3. To demonstrate chromosomal over-expression of a complemented gene
4. To adopt the method to other clostridia such as *C. difficile*
5. To exemplify gene knockout using *codA* as a negative selection marker

4.3 Results

To demonstrate the above mentioned methods of mutant generation, complementation and over-expression, target sites or genes were required which encode a phenotype which was easily quantified. The genes Cac2071

(*spo0A*), Cac0355 *(polygalacturonase),* Cap0098 (*amyA*), Cap0168 (*amyB*) and Cac1502 (restriction endonuclease *Cac*824*I*) were selected. The *spo0A* is the master regulator of spore formation and plays a major role in the switch of *C. acetobutylicum* from acidogenesis to solventogenesis and has been knocked out previously with ClosTron technology (Heap et al., 2010b). The restriction endonuclease *Cac*824*I* encoded by Cac1502 and a second restriction modification enzyme encoded by Cac3535, were deleted by homologous recombination process using MLS as a positive selection and *upp* as a negative selection marker (Soucaille et al., 2006). *C. acetobutylicum* ATCC 824 megaplasmid pSOL1 contains two amylase genes Cap0168 and Cap0098 responsible for degradation of starch, but to determine their specific role, rapid insertional mutants were made using ClosTron as described in Section 3.3.9.2. Cap0168 was shown to encode amylase activity and was therefore selected for in-frame deletion using the *pyrE* system. *C. acetobutylicum* ATCC 824 genome contains Cac0355 encoding a putative *polygalacturonase* responsible for pectin metabolism (Nolling et al., 2001). The ClosTron mutagenesis tool was used to make insertional mutants of Cac0355 as described in Figure 4.7 , but analysis showed no effects of this mutation on halo formation in CBM medium supplemented with pectin as a sole source of carbon (Figure 4.8). So this gene was not examined further.

4.3.1 In-frame deletion using *pyrE* as a negative selection marker

4.3.1.1 Complementation of *pyrE* mutant with *C. sporogenes pyrE*

To derive an allelic exchange method based on the use of *pyrE* as a negative selection marker, it was first necessary to identify a heterologous *pyrE* gene that could be used. The use of a heterologous gene was considered preferable

to using a copy of the *C. acetobutylicum* ATCC 824 gene, as this would avoid recombination between the functional *pyrE* gene being used, and the remnants of the mutant *pyrE* allele in the chromosome.

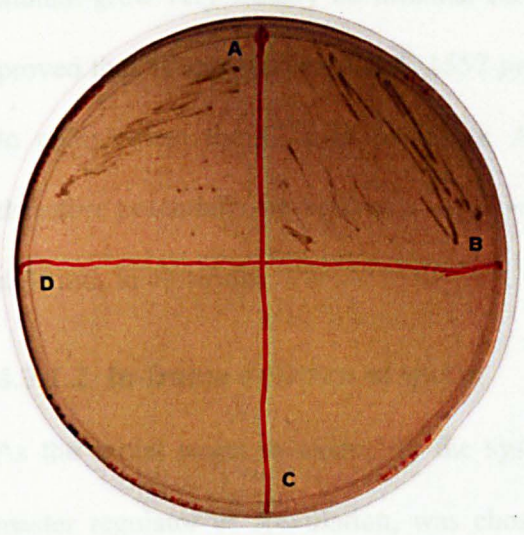


Figure 4.1 Complementation assay of the *pyrE* mutant on CBM agar.

A, *C. acetobutylicum* ATCC 824 *pyrE* mutant (pMTL-ME3); B, *C. acetobutylicum* ATCC 824 wild-type; C, *C. acetobutylicum* ATCC 824 *pyrE* mutant and D, *C. acetobutylicum* ATCC 824 *pyrE* mutant (pMTL-85141) incubated for 24 hrs at 37° C under anaerobic conditions.

To achieve the above, the ability of the *pyrE* gene of *C. sporogenes* ATCC 15579 to restore the *C. acetobutylicum* 824 *pyrE* mutant strain to uracil prototrophy was tested. Accordingly, a plasmid pMTL-ME3 was constructed carrying *C. sporogenes* ATCC 15579 *pyrE* downstream of *catP* gene as described in section 2.5.3, and transformed into the *C. acetobutylicum* ATCC 824 *pyrE* mutant along with the empty vector control, pMTL-85141. Transformants were selected on CGM supplemented with 15µg/ml Tm and 20µg/ml uracil and incubated for 24hrs. Single colonies were re-streaked onto the same medium followed by streaking onto minimal medium CBM agar,

without Tm. All the cells transformed with pMTL-ME3 showed growth similar to the *C. acetobutylicum* ATCC 824 wild-type but the cells transformed with the control vector pMTL-85141 and the *C. acetobutylicum* ATCC 824 *pyrE* mutant grew very weakly on minimal medium as shown in Figure 4.1 . This proved that *C. sporogenes* ATCC 1557 *pyrE* gene could be used on a plasmid to complement the *C. acetobutylicum* ATCC 824 *pyrE* mutant and could therefore potentially be used as a negative selection marker to make in-frame deletions in clostridia.

4.3.1.2 In-frame deletion of *spo0A*

As the initial target to exemplify the system, the *spo0A* gene, encoding the master regulator of sporulation, was chosen as the sporulation phenotype is easily scored. The *C. acetobutylicum* ATCC 824 *pyrE* mutant (CRG CC#1545) made using pMTL-JH12 (Section 3.3.3) was therefore transformed with pMTL-ME3::Cacspo0A-KO-1.5 which contains the chloramphenicol/ thiamphenicol-resistance marker *catP*, the heterologous *pyrE* gene from *C. sporogenes* ATCC 15579 and Gram positive origin of replication from *B. subtilis* plasmid pIM13 (Monod et al., 1986). The pIM13 replicon exhibits segregational instability in *C. acetobutylicum* (Heap et al., 2009) and single-crossover clones which spontaneously arise within transformant populations can be isolated by subculture under antibiotic selection. Transformants were selected on CGM supplemented with 15µg/ml Tm.

After forty-eight hours of incubation, faster growing colonies were apparent, which were streaked onto the same medium (passage P1) and incubated for forty-eight hours. Single colonies from these plates were then re-streaked again onto the same medium (passage P2). As the homology arms are of equal size,

plasmid integration can occur between DNA at either side of the in-frame deletion, as shown in Figure 4.2 . At this stage, several independent colonies were inoculated in separate 1ml 2xYTG broth cultures supplemented with 7.5µg/ml Tm and after extracting genomic DNA, screening PCRs were performed to identify single crossover integrants.

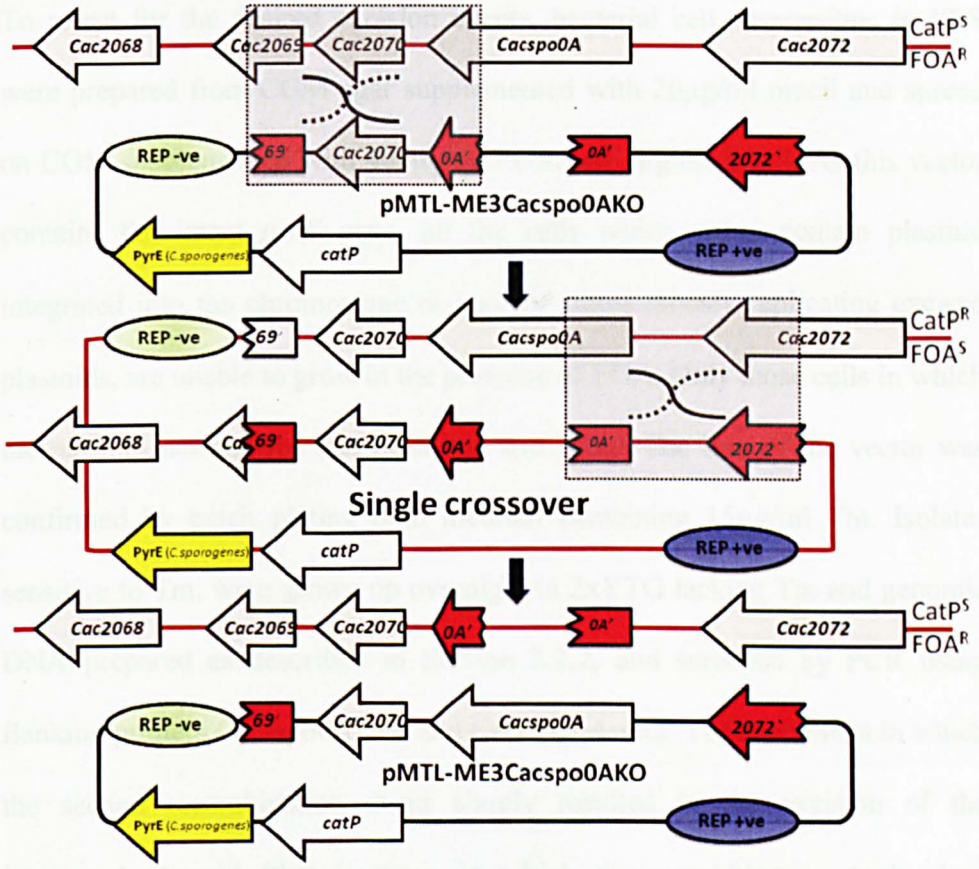


Figure 4.2 Schematic representation of allelic exchange for *spo0A* using *pyrE*

Knockout vector pMTL-ME3 containing allelic exchange cassette can integrate either right or left homology arm due to symmetric homology arms to select pure single crossover integrant in the presence of replication defective Gram-positive replicon pIM13 under antibiotic selection pressure confers by *catP*. Single crossover integrants are then propagated under no selection to allow cells to undergo second recombination resulting in either wild type or deletion which could be detected using FOA in the media followed by screening by PCR using flanking primers Cac-spo0A-sF2 and Cac-spo0A-sR2 encompassing the homology arms. $CatP^R$, chloramphenicol/thiamphenicol -resistance; $CatP^S$, chloramphenicol/thiamphenicol-sensitive; FOA^R, 5-Fluoroorotic Acid-resistance; FOA^S, 5-Fluoroorotic Acid-sensitive.

These integrants were then streaked onto CGM supplemented with 20μg/ml uracil and incubated for 2-3 days to allow cells to undergo a second recombination event. Excision of the plasmid would result either in wild-type cells, or cells in which the wild-type allele had been exchanged for the in-frame deletion mutant allele carried by the plasmid as shown in Figure 4.2 .

To select for the desired excision events, bacterial cell suspensions in PBS were prepared from CGM agar supplemented with 20μg/ml uracil and spread on CGM supplemented with 400μg/ml FOA and 1μg/ml uracil. As this vector contains the intact *pyrE* gene, all the cells which either contain plasmid integrated into the chromosome or contain autonomously replicating excised plasmids, are unable to grow in the presence of FOA. Only those cells in which the plasmid has excised and been lost will grow. The loss of the vector was confirmed by patch plating onto medium containing 15μg/ml Tm. Isolates sensitive to Tm, were grown up overnight in 2xYTG lacking Tm and genomic DNA prepared as described in Section 2.2.2, and screened by PCR using flanking primers Cac-spo0A-sF2 and Cac-spo0A-sR2. Those colonies in which the second recombination event simply resulted in the excision of the integrated plasmid (that is those in which the recombination took place between the same homology arms that had 'driven' integration) generated a PCR product of 2114bp, synonymous with the parental strain *C. acetobutylicum* ATCC 824 *pyrE* mutant as shown in Figure 4.2 .

Those in which the excised plasmid carried the previously located parental strain *spo0A* allele (that is those in which the recombination took place between the homology arms that were not involved in the original integration event) led to the generation of a PCR fragment of 1679bp. In these clones an

in-frame deletion was created in the chromosomal *spo0A* gene as shown in Figure 4.2 . These PCR products were then confirmed by nucleotide sequencing. Out of the nine clones screened, one showed reversion to parental genotype at *spo0A,* whereas five contained the desired in-frame deletion and the rest generated PCR products suggesting presence of both wild-type and in-frame deletion (Figure 4.3).

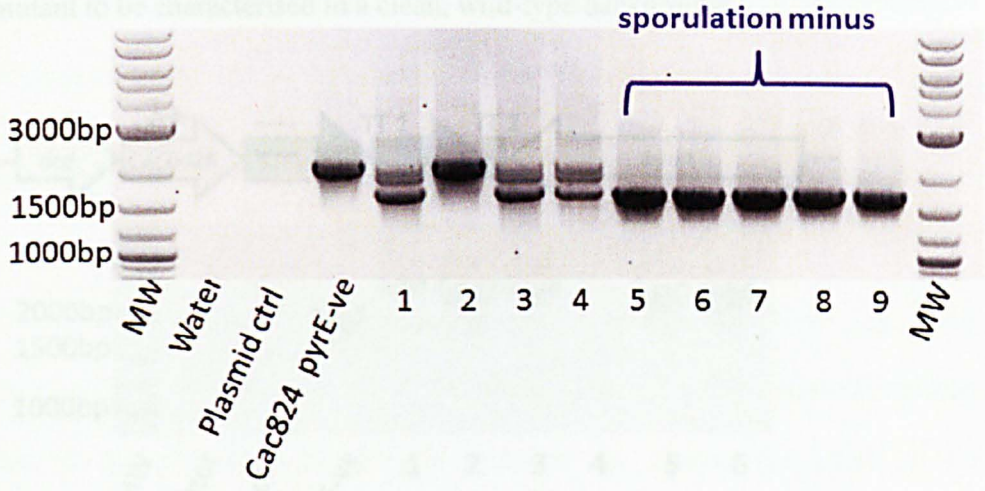


Figure 4.3 PCR screening of the in-frame deletion of *spo0A*

PCR screening of nine FOA^R colonies using flanking primers Cac-spo0A-sF2 and Cac-spo0A-sR2, PCR produced 1679bp and 2114bp fragments for double crossover deletion mutant and parental strain *C. acetobutylicum* ATCC 824 *pyrE* mutant respectively and confirmed by nucleotide sequencing these fragments (data not shown). MW, 2-log DNA marker (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL-ME3::Cacspo0A-KO-1.5 plasmid DNA control; Cac824 *pyrE*-ve; Parental strain *C. acetobutylicum* 824 *pyrE* mutant (CRG CC #1545) and 1-9 clones.

The sporulation minus phenotype of the mutants was confirmed by phase contrast microscopy of 2-5 day old broth cultures showing that unlike the parental strain, phase-bright spores were absent in the mutant cultures. Following heat shock at 80 °C for 10min, and plating on CGM medium

supplemented with 20µg/ml uracil, colonies were obtained from broth cultures of the parental strain but not for those of the *spo0A* mutant (data not shown).

4.3.1.3 Repair of *pyrE* locus and assessment of phenotype

Having created a *spo0A* mutant, an ACE-vector, pMTL-ME7C (MATERIALS AND METHODS, Section 2.5.1.6) was used to rapidly restore the chromosomal *pyrE* allele to wild-type, allowing the specific in-frame deletion mutant to be characterised in a clean, wild-type background.

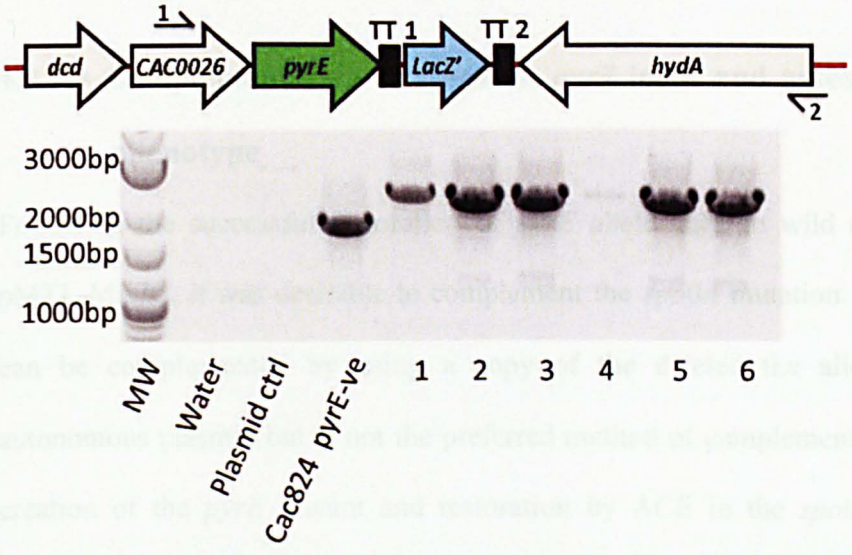


Figure 4.4 Repair of *pyrE* locus using pMTL-ME7C in *spo0A* minus strains

PCR screening of the six colonies using flanking primers Cac0026-sF2 (1) and Cac-hydA-sR2 (2). PCR produced 1936bp and 24754bp fragments for parental strain *pyrE* negative and *C. acetobutylicum* 824 *pyrE* repaired with pMTL-ME7C. MW, 2-log DNA marker (NEB) molecular weight marker; Water, PCR-negative control; plasmid, pMTL-ME7C plasmid DNA control; Cac824 *pyrE*-ve, *C. acetobutylicum* 824 *pyrE* mutant (CRG CC#1545) genomic DNA control and 1-6 clones. TT1, Transcriptional terminator from *Lactococcus lactis* IL403 between *ldh* and *yngG* genes; TT2, Transcriptional terminator found between *C. acetobutylicum* ATCC 824 *pyrE* and *hydA* genes.

The *C. acetobutylicum* 824 *pyrE* mutant strain containing the in-frame deletion of *spo0A* was transformed with pMTL-ME7C following the same procedure as

described previously for pMTL-JH14 in Section 3.3.4. Colonies isolated on minimal medium were expected to be double crossover integrants and were screened by PCR using flanking primers Cac0026-sF2 and Cac-hydA-sR2 as shown in Figure 4.4 . These PCR products were confirmed by nucleotide sequencing (data not shown).

The sporulation minus phenotype of the *pyrE*+ve, *spo0A* mutants (CRG CC# ME19) was confirmed by phase contrast microscopy and plating of heat treated cultures on CGM agar as described in Section 4.3.1.2.

4.3.1.4 Complementation of *spo0A* at *pyrE* locus and assessment of phenotype

Following the successful restoration of *pyrE* allele back to wild type using pMTL-ME7C, it was desirable to complement the *spo0A* mutation. Mutations can be complemented by using a copy of the deleted the allele on an autonomous plasmid but is not the preferred method of complementation. The creation of the *pyrE* mutant and restoration by ACE in the *spo0A* mutant, provided the opportunity to deliver the wild-type allele of the deleted gene to the chromosome at an appropriate gene dosage under the control of its native promoter using an ACE-vector.

To achieve this, the *C. acetobutylicum* ATCC 824 (*pyrE*-ve, *spo0A*-ve) (CRG CC# 3520) mutant was transformed with ACE vector pMTL-ME7C-PCacspo0A (complementation) following the same procedure as described for pMTL-JH14 (Section 3.3.4). Colonies isolated on minimal medium CBM, were screened for complementation by PCR using flanking primers Cac0026-sF2 and Cac-hydA-sR2 as shown in Figure 4.5.

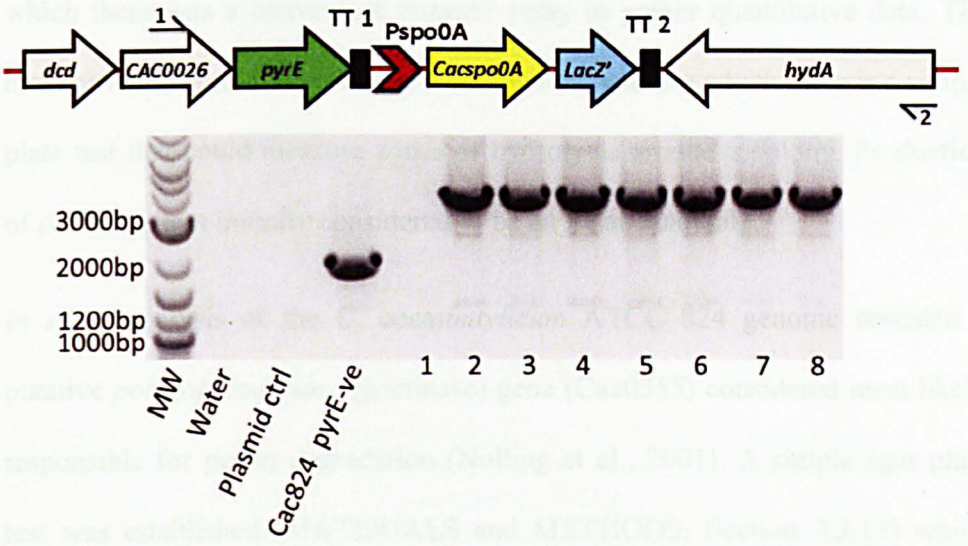


Figure 4.5 PCR screening for integration of *spo0A* at *pyrE* locus for complementation studies

PCR screening of eight colonies using flanking primers Cac0026-sF2 and Cac-hydA-sR2, PCR produced 1936bp and 3659bp fragments for *pyrE*-ve and *C. acetobutylicum* 824 *pyrE*+ve repaired and complemented with pMTL-ME7C-PCacspo0A respectively. MW, 2-log DNA marker (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL-ME7C-PCacspo0A plasmid DNA control; Cac824 *pyrE*-ve, *C. acetobutylicum* 824 *pyrE*-ve genomic DNA control and 1-8 clones. PCR failed for clone 1. TT1, Transcriptional terminator from *Lactococcus lactis* IL403 between *ldh* and *yngG* genes; TT2, Transcriptional terminator found between *C. acetobutylicum* ATCC 824 *pyrE* and *hydA* genes.

All the colonies in which a functional copy of *spo0A* with native promoter (Harris et al., 2002) was integrated at the *pyrE* locus were confirmed for sporulation phenotype as described in Section 4.3.1.2 (data not shown).

4.3.1.5 Over-expression of complementing genes at the *pyrE* locus

Pectinase

In order to exemplify the utility of over-expressing complementing genes at the *pyrE* locus in mutant strains, it was desirable to use a mutant which had an easily 'scoreable' phenotype, at both a qualitative and quantitative level. It was considered that the most appropriate target would be a gene encoding a protein, the production of which could be measured by a simple agar plate test, and for

which there was a convenient enzyme assay to gather quantitative data. The most obvious candidate would be a secreted enzyme for which there is a simple plate test that could measure zones of hydrolysis around a colony. Production of pectinase was initially considered to be an ideal candidate.

In silico analysis of the *C. acetobutylicum* ATCC 824 genome revealed a putative *polygalacturonase* (pectinase) gene (Cac0355) considered most likely responsible for pectin degradation (Nolling et al., 2001). A simple agar plate test was established (MATERIALS and METHODS, Section 2.2.17) which demonstrated that the organism did indeed produces pectinase activity, and furthermore that this could be easily detected (Figure 4.6).

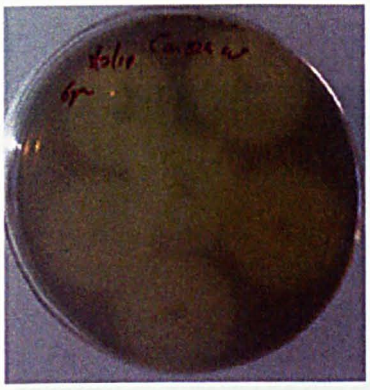


Figure 4.6 Pectinase assay on CBM containing 0.5% w/v pectin

C. acetobutylicum ATCC 824 was grown overnight in 2xYTG and 10µl of culture spotted on CBM agar containing 0.5% w/v pectin and incubated for 48 hours followed by staining with iodine showing *pectinase* activity as zones of clearing

Rather than investing the time and effort required to make an in-frame deletion mutant, the more rapid mutagenic ClosTron (pMTL007C-E2-Cac0355-1167s and pMTL007C-E2-Cac0355-656a) was used to make insertional mutants in the encoding gene. PCR primers (Cac0355-NdeI-sF1 and Cac0355-EcoRI-sR1) flanking the insertion sites were employed to screen for the ClosTron mutants distinguished from wild-type as show in (Figure 4.7).

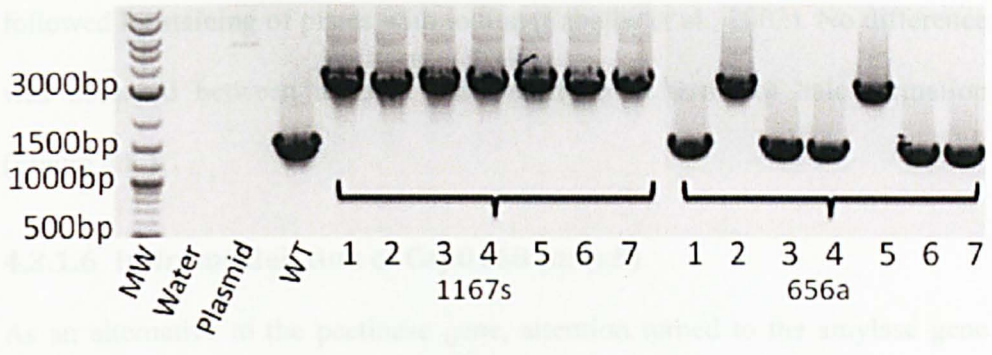


Figure 4.7 PCR screening of ClosTron mutant of *Polygalacturonase*

PCR screening of seven erythromycin-resistant colonies of *C. acetobutylicum* ATCC 824 obtained for insertion sites 1167s and 656a using flanking primers Cac0355-NdeI-sF1 and Cac0355-EcoRI-sR1. PCR produced 1542bp and 3342bp fragments for wild-type and intron insertion respectively. Insertion sites were confirmed by sequencing (Data not shown). MW, 2-log DNA marker (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL007C-E2-Cac0355-1167s and pMTL007C-E2-Cac0355-656a plasmid DNA control; WT, wild-type *C. acetobutylicum* ATCC 824 genomic DNA control and 1-7 clones for each target sites. All the clones for Cac0355-1167s were mutants while for Cac0355-656a only two out of seven were correct.

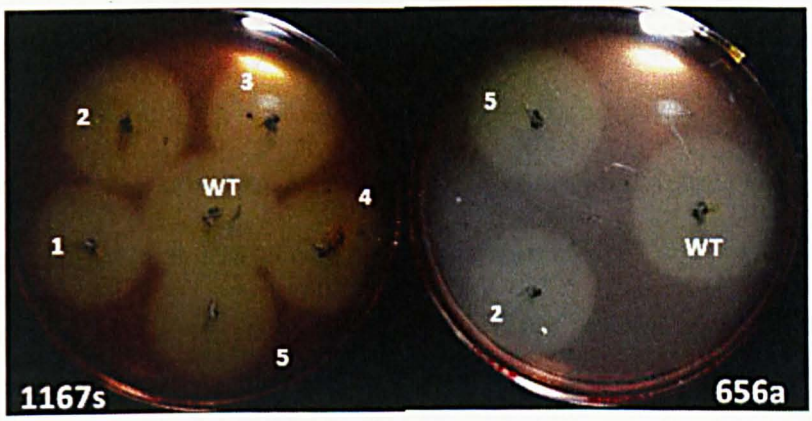


Figure 4.8 Phenotype of the ClosTron mutants of *polygalacturonase*

C. acetobutylicum ATCC 824 wild type and ClosTron mutants of Cac0355 for both target sites 1167s and 656a were grown overnight in 2xYTG and 10µl of culture spotted on CBM plate containing 0.5% w/v pectin and incubated for 48hrs followed by staining of plates with iodine to check the clearing of zones.

These intron insertional mutants of Cac0355 as well as the wild-type strain were grown overnight in 2xYTG and 10µl from each were spotted on CBM 0.5% w/v pectin and incubated for 72hours under anaerobic conditions

followed by staining of plates with iodine (Sabathe et al., 2002). No difference was observed between mutants and wild-type in terms of halo formation (Figure 4.8).

4.3.1.6 In-frame deletion of Cap0168 (*amyB*)

As an alternative to the pectinase gene, attention turned to the amylase gene Cap0168, as parallel studies (see Section 3.3.9.1) had established that the expression of this gene could be easily assessed using starch-iodine plates. As the ClosTron mutagenesis had already established that this gene was responsible for starch degradation and halo phenotype (Sections 3.3.9.1 and 3.3.9.2), an appropriate mutant was generated by allelic exchange as shown in (Figure 4.9).

Accordingly, the *C. acetobutylicum* 824 *pyrE* mutant was transformed with pMTL-ME3::Cap0168KO-1.5, and the in-frame deletion procedure followed exactly the same as described for *spo0A* (see Figure 4.2). The mutant phenotype was evaluated on plates containing 2% w/v starch and 0.5% w/v glucose and found to be similar to ClosTron mutants as described in Sections 3.3.9.1 and 3.3.9.2 (data not shown).

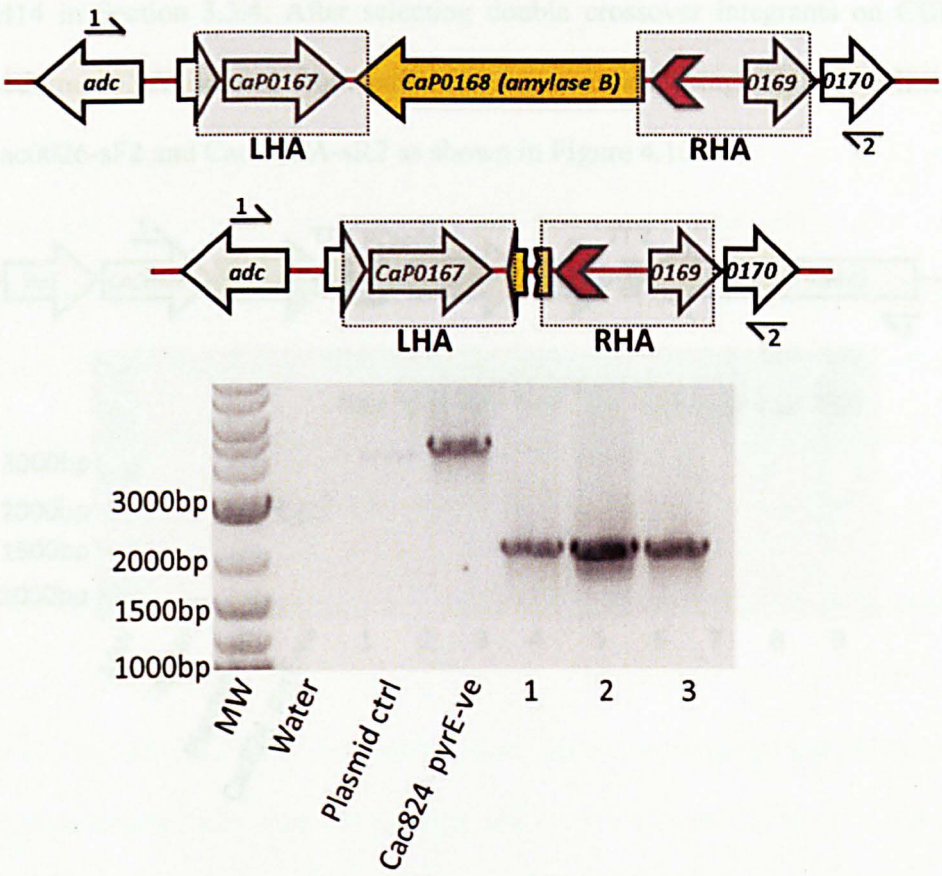


Figure 4.9 PCR screening of the in-frame deletion of Cap0168 using *pyrE*

PCR screening of three clones using flanking primers cac-amyP-sF2 (1) and cac-amyP-sR2 (2), PCR produced 2078bp and 4355bp fragments for in-frame deletion mutant *C. acetobutylicum* ATCC 824 *pyrE*-veCap0168-ve and parental strain *C. acetobutylicum* ATCC 824 *pyrE*-ve, respectively. MW, 2-log DNA marker (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL-ME3-CaP0168-KO-1.5 plasmid DNA control; Cac824 *pyrE*-ve, *C. acetobutylicum* ATCC 824 *pyrE*-ve genomic DNA control and 1-3 clones.

4.3.1.7 Over-expression of complementing gene at the *pyrE* locus Cap0168 (*amyB*)

To utilize the system of over-expressing a complemented gene, the amylase gene Cap0168 was cloned in pMTL-ME7X downstream of the strong *fdx* promoter to make pMTL-ME7X-Cap0168. This was transformed into the *C. acetobutylicum* ATCC 824 *pyrE* mutant (CRG CC#1545) and screening for integration performed using the same procedure as described above for pMTL-

JH14 in Section 3.3.4. After selecting double crossover integrants on CBM medium, DNA isolated from colonies was screened using flanking primers Cac0026-sF2 and Cac-hydA-sR2 as shown in Figure 4.10 .

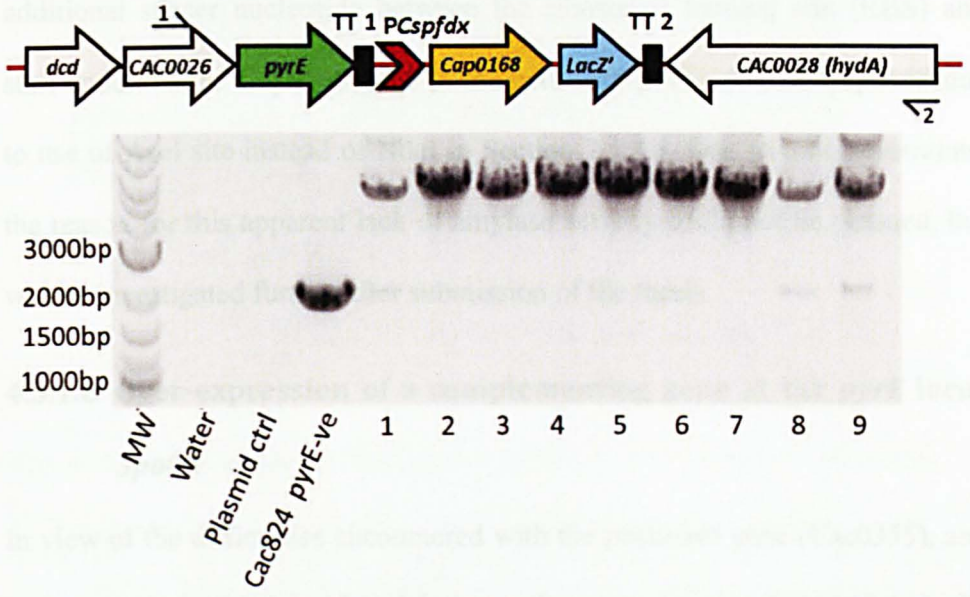


Figure 4.10 PCR screening of over-expression constructs for Cap0168 inserted at the *pyreE* locus

PCR screening of nine isolated clones using flanking primers Cac0026-sF2 (1) and Cac-hydA-sR2 (2). PCR produced 1936bp and 4918bp fragments for parental *pyrE*-ve strain (CRG CC#1545) and Cac824-*pyrE*+ve, over-expression strain (CRG CC# ME30 Table 2.1) with pMTL-ME7X-Cap0168 respectively. MW, 2-log DNA marker (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL-ME7X-Cap0168 plasmid DNA control; Cac824 *pyrE*-ve, *C. acetobutylicum* ATCC 824 *pyrE*-ve genomic DNA control and 1-9 clones. TT1, Transcriptional terminator from *Lactococcus lactis* IL403 between *ldh* and *yngG* genes; TT2, Transcriptional terminator found between *C. acetobutylicum* ATCC 824 *pyrE* and *hydA* genes.

The in-frame deletion mutant of Cap0168, the over-expressed strains and wild-type controls were analysed for amylase activity on CBM medium containing 2% w/v starch 0.5% w/v glucose, but the over-expression strain (CRG CC# ME30 Table 2.1) did not show any difference in amylase activity when compared to the parental strain (CRG CC#1545 Table 2.1) (data not shown). It

may be because the *fdx* promoter from *C. sporogenes fdx* used in this case was not strong enough compared to its natural promoter (Nolling et al., 2001). Another possible reason for the deficiency of amylase activity could be the additional spacer nucleotide between the ribosomal binding site (RBS) and start codon of the Cap0168 gene in the construct pMTL-ME7X-Cap0168 due to use of AseI site instead of NdeI in Section 2.5.3.5. Due to time constraints, the reason for this apparent lack of amylase activity could not be pursued, but will be investigated further after submission of the thesis.

4.3.1.8 Over-expression of a complementing gene at the *pyrE* locus

Spo0A

In view of the difficulties encountered with the pectinase gene (Cac0355), and because experiments aimed at deleting and over-expressing Cap0168 (*amy* B) initially ran into difficulties such as deletion of Cap0168 took longer than expected and later on over-expression of Cap0168 under the control of *fdx* promoter did not produced the expected results as described in sections 4.3.1.6 and 4.3.1.7. It was decided to assess the effect of over-expressing the complementing *spo0A* gene.

The *C. acetobutylicum* ATCC 824 *spo0A* gene was cloned into the ACE vector pMTL-ME7X under the control of the strong *fdx* promoter to make pMTL-ME7X-Cacspo0A which was transformed into the *C. acetobutylicum* 824 *pyrE* mutant (CRG CC#1545 Table 2.1). Revertants were selected on CBM medium for the restoration of *pyrE* gene as described for pMTL-JH14 in Section 3.3.4. Integrants were verified by PCR screening using flanking primers Cac0026-sF2 and Cac-hydA-sR2 as shown in Figure 4.11

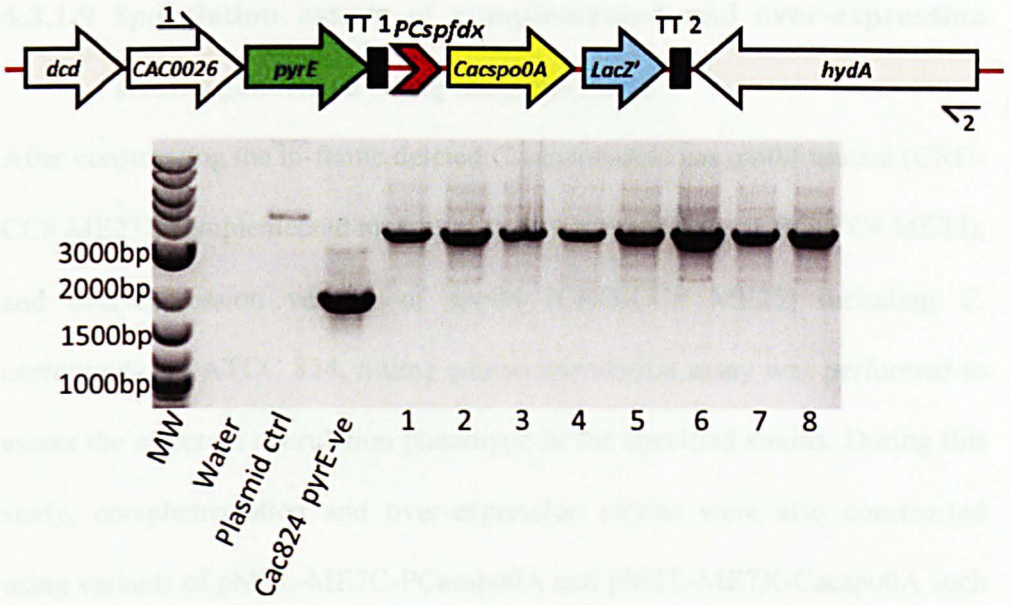


Figure 4.11 PCR screening for insertion of *spo0A* for Over-expression studies

PCR screening of eight colonies using flanking primers Cac0026-sF2 (1) and Cac-hydA-sR2 (2), PCR produced 1936bp and 3505bp fragments for *C. acetobutylicum* 824 *pyrE*-ve (CRG CC#1545) and *C. acetobutylicum* 824 *pyrE*+ve, complemented with over-expression vector pMTL-ME6X-Cacspo0A respectively. MW, 2-log DNA marker (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL-ME6X-Cacspo0A plasmid DNA control; Cac824 *pyrE*-ve, *C. acetobutylicum* 824 *pyrE*-ve genomic DNA control and 1-8 clones. TT1, Transcriptional terminator from *Lactococcus lactis* IL403 between *ldh* and *yngG* genes; TT2, Transcriptional terminator found between *C. acetobutylicum* ATCC 824 *pyrE* and *hydA* genes.

The phenotype of the over-expressed strain (CRG CC# ME25 Table 2.1) was checked for sporulation by phase contrast microscopy and was found similar to that of wild-type and *spo0A* complemented strains. Similar results for spore formation of the *spo0A* complemented strains were shown on CGM agar after heat treatment at 80 °C for 10min and incubation for 48hrs anaerobically (data not shown).

4.3.1.9 Sporulation assays of complemented and over-expression strains generated using the *pyrE* locus

After constructing the in-frame deleted *C. acetobutylicum spo0A* mutant (CRG-CC# ME23), complemented mutant of *spo0A* at *pyrE* locus (CRG-CC# ME24), and over-expression version of *spo0A* (CRG-CC# ME25) including *C. acetobutylicum* ATCC 824, a time course sporulation assay was performed to assess the effect on sporulation phenotype in the specified strains. During this study, complementation and over-expression strains were also constructed using variants of pMTL-ME7C-PCacspo0A and pMTL-ME7X-Cacspo0A such as pMTL-ME6C-PCacspo0A and pMTL-ME6X-Cacspo0A respectively. These variants do not contain the extra terminator TT2 found between the *ldh* and *yngG* of *Lactococcus lactis* IL1403 as shown in Figure 4.11 (de Hoon et al., 2005). This was incorporated downstream of the *C. acetobutylicum* 824 *pyrE* gene to stop the transcription through from the *pyr* operon running into the complementation and over-expression genes. Due to lack of time, in this sporulation assay, only the *C. acetobutylicum spo0A* mutant (CRG CC# ME23), complemented mutant of *spo0A* at *pyrE* locus (CRG CC# ME24), and over-expression version of *spo0A* (CRG CC# ME25) and *C. acetobutylicum* ATCC 824 wild type were tested. After 24hrs incubation, a slightly higher number of spores were found to be produced by the *spo0A* complemented and over-expression strain compared to *C. acetobutylicum* ATCC 824 wild-type but this difference was not apparent after this time point (Figure 4.12).

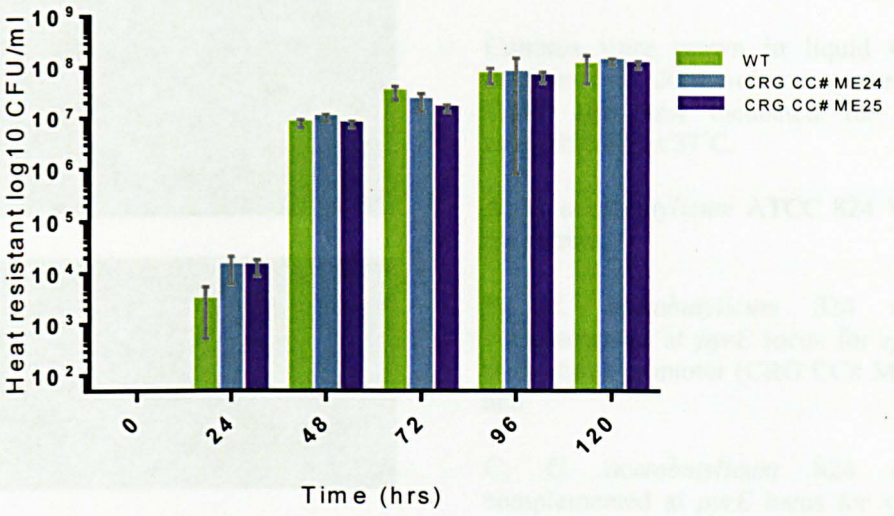
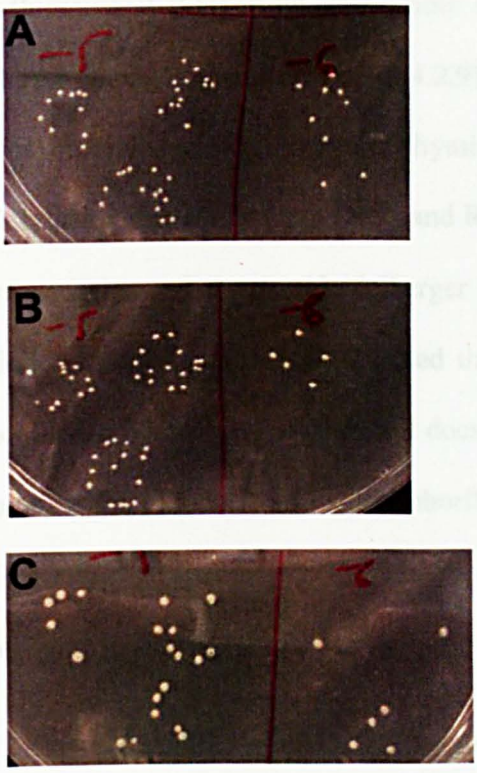


Figure 4.12 Comparison of sporulation of *C. acetobutylicum* strains

Sporulation count is expressed as log10 of heat resistant colony forming unit per ml (CFU/ml) on y-axis with error bars as Mean+SD Error and 24hours time points on x-axis for WT, *C. acetobutylicum 824* wild-type, *C. acetobutylicum* 824 *pyrE* mutant repaired with plasmid pMTL-ME6C-Cacspo0A (CRG CC# ME24),, *C. acetobutylicum* 824 *pyrE* mutant repaired with plasmid pMTL-ME6X-*Cacspo0A* (CRG CC# ME25. Since *C. acetobutylicum spo0A* mutant (CRG-CC# ME23) is a null mutant of *spo0A* gene therefore data is not shown).

Although over-expression of *spo0A* using *fdx* promoter did not show any change in the sporulation assay, a clear phenotypic change was observed as colonies of the over-expression strain, were visibly bigger compared to complemented only and wild-type strains when grown on CBM medium (Figure 4.13)



Cultures were grown in liquid CBM medium and 20µl volume spotted on CBM agar and incubated for 24hr anaerobically at 37°C.

A; *C. acetobutylicum* ATCC 824 Wild-type strain,

B; *C. acetobutylicum* 824 strain complemented at *pyrE* locus for *spo0A* with native promoter (CRG CC# ME24) and

C; *C. acetobutylicum* 824 strain complemented at *pyrE* locus for *spo0A* with *fdx* promoter (CRG CC# ME25).

Figure 4.13 Effect of over-expression of *spo0A* on colony size

4.3.2 In-frame deletion using counter-selection marker *codA*

4.3.2.1 In-frame deletion of *spo0A*

Recently *codA* has been successfully utilized in the pathogenic clostridial species *C. difficile* to make precise changes in the *tcdC* genotype (Cartman et al., 2012). The allelic exchange cassettes and methodological procedures are the same as for *pyrE*, except counter-selection is based on differential resistance to FC (CodA) rather than FOA (PyrE). In contrast to the *pyrE*-based system, *codA*-based mutagenesis can be undertaken in a wild type background (provided the host does not carry a native *codA* gene). The *codA* gene encodes for cytosine deaminase (EC 3.5.4.1), an enzyme that catalyzes the conversion of innocuous pyrimidine analogue 5-fluorocytosine (FC) into the highly toxic

5-fluorouracil (FU) in an irreversible reaction. FU toxicity occurs via uracil phosphoribosyltransferase (EC 2.4.2.9) leads to inhibition of the nucleotide biosynthesis pathway enzyme thymidylate synthase and misincorporates fluorinated nucleotides into DNA and RNA hence affecting the replication and transcription processes (Heidelberger et al., 1983; Longley et al., 2003). Genome sequence analysis revealed that *C. acetobutylicum* ATCC 824 does not have a homologue of *codA* but does have a homologue of a *upp* (Cac2879) gene, which encodes uracil phosphoribosyltransferase (Nolling et al., 2001). This indicated that *C. acetobutylicum* would be inherently resistant to FC but sensitive to FU, a prerequisite for using *codA* as a counterselection marker.

To extend the utility of *codA* to *C. acetobutylicum*, *codA* vector pMTL-SC7515 was constructed by Dr. Steve Cartman (CRG, Nottingham) as shown in Figure 2.3 by replacing the origin of replication pCB102 in pMTL-SC7315 with *repL* from *Bacillus subtilis*. plasmid pIM13 (Monod et al., 1986) and contains the plasmid selection marker *catP*. The first target gene selected for inactivation was *spo0A*, which is the master regulator of sporulation and solvent production in *C. acetobutylicum*. The knockout cassette created previously (Section 2.5.3.1) was cloned into pMTL-SC7515 to make pMTL-SC7515::Cacspo0A-KO-1.5 and transformed into *C. acetobutylicum* ATCC 824 wild-type. This plasmid is segregationally unstable in *C. acetobutylicum* (Heap et al., 2009) and single-crossover clones can be isolated by subculture under antibiotic selection. Transformants were selected on CGM supplemented with 15μg/ml Tm. After forty-eight hours of incubation, faster growing colonies were streaked onto the same medium (passage P1) and incubated for twenty-four hours and then re-streaked again onto the same medium (passage P2). As the

homology arms are of equal size, integration of the plasmid can occur through recombination at either homology arm (Figure 4.2). At this stage, a few colonies were inoculated into each 1ml 2xYTG broth culture supplemented with 7.5μg/ml Tm and grown overnight. Genomic DNA extracted from harvested pellets was subjected to PCR screening using one primer annealing to the chromosome (Cac-spo0A-sF2 or Cac-spo0A-sR2) and a second to the plasmid (SC7-R or SC7-F) in order to determine whether transformants were single crossover integrants (data not shown). Cells of those clones that were determined to be pure single crossover integrants were then streaked onto CGM medium and incubated for 2-3 days to allow cells to undergo a second recombination event.

To isolate a double crossover mutant, growth from the non-selective plate was resuspened in PBS. Appropriate dilutions of the cell suspension were spread on CGM agar plates supplemented with 100μg/ml 5-FC which were incubated for 48hrs. All of those cells that retain the *codA* gene (either those still containing the autonomously replicating plasmid, or those in which the plasmid had integrated) are sensitive to FC and are, therefore, unable to grow. In contrast, those cells in which the plasmid has excised, through a second recombination event, and been lost due to segregational instability, are able to grow in the presence of FC. The loss of the KO vector was confirmed by patch plating on CGM with or without Tm supplementation.

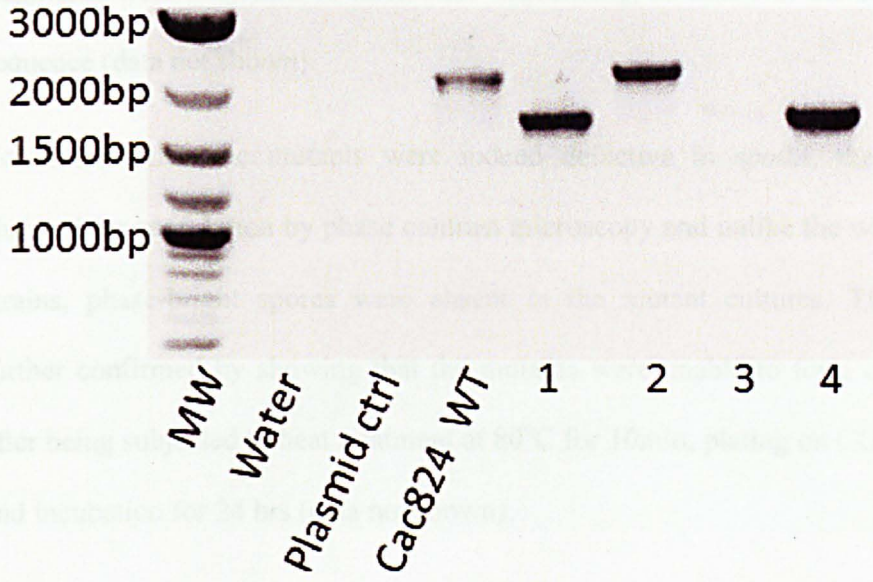


Figure 4.14 PCR screening of the in-frame deletion of *spo0A* using *codA*

PCR screening of four 5-Fluorocytosine resistant (FC^R) colonies using flanking primers Cac-spo0A-sF2 and Cac-spo0A-sR2, PCR produced 1679bp and 2114bp fragments for double crossover deletion mutant and wild-type respectively. PCR products were confirmed by nucleotide sequencing (Data not shown). MW, 2-log DNA marker (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL-SC7515::Cacspo0A-KO-1.5 plasmid DNA control; Cac824 WT, *C. acetobutylicum* 824 genomic DNA control and 1-4 clones.

Those FC^R colonies that had lost the KO plasmid can either be wild-type or the desired in-frame deletion mutants. To confirm which, genomic DNA was prepared from overnight cultures and used as a template in a PCR using primers Cac-spo0A-sF2 and Cac-spo0A-sR2 that flanked the *spo0A* gene. In the case of the wild-type, the predicted fragment that would be generated using these primers was 2114bp, whereas the in-frame deletion variant if present would generate a fragment of 1679bp. As can be seen in Figure 4.14 , of the 4 Tm^S clones screened, one generated a fragment of a size indistinguishable from the wild-type control, whereas three (one of which gave only weak band) of the clones generated a 1677bp fragment indicative of an in-frame deletion mutant. To confirm the identity of the latter as in-frame deletions, the amplified

fragments were subjected to nucleotide sequencing, resulting in the expected sequence (data not shown).

To confirm that the mutants were indeed defective in *spo0A*, they were checked for sporulation by phase contrast microscopy and unlike the wild type strains, phase-bright spores were absent in the mutant cultures. This was further confirmed by showing that the mutants were unable to form colonies after being subjected to heat treatment at 80°C for 10min, plating on CGM agar and incubation for 24 hrs (data not shown).

4.3.2.2 In-frame deletion of Cac1502 (type II restriction system)

Rebase http://tools.neb.com/~vincze/genomes/view.php?enzname=Cac824I and genome sequence analysis revealed that *C. acetobutylicum* ATCC 824 has a very effective restriction endonuclease system *Cac*824*I* encoded by Cac1502. Its recognition sequence is 5′-GCNGC-3′, a sequence that is very frequently found in *E. coli* plasmid vectors. Several other restriction modification enzymes/genes are also predicted but are as yet not characterized (Nolling et al., 2001). Restriction modification systems are the host defence systems against any foreign DNA entering the cells. To protect the foreign DNA against type II restriction endonuclease, *Clostridium-E. coli* shuttle vectors are methylated *in vivo* before electroporation into *C. acetobutylicum* (Heap et al., 2007; Mermelstein & Papoutsakis, 1993). Insertional inactivation or deletion of Cac1502 enables the cells to be transformed with unmethylated DNA as efficiently as with methylated plasmid DNA (Dong et al., 2010; Soucaille et al., 2006).

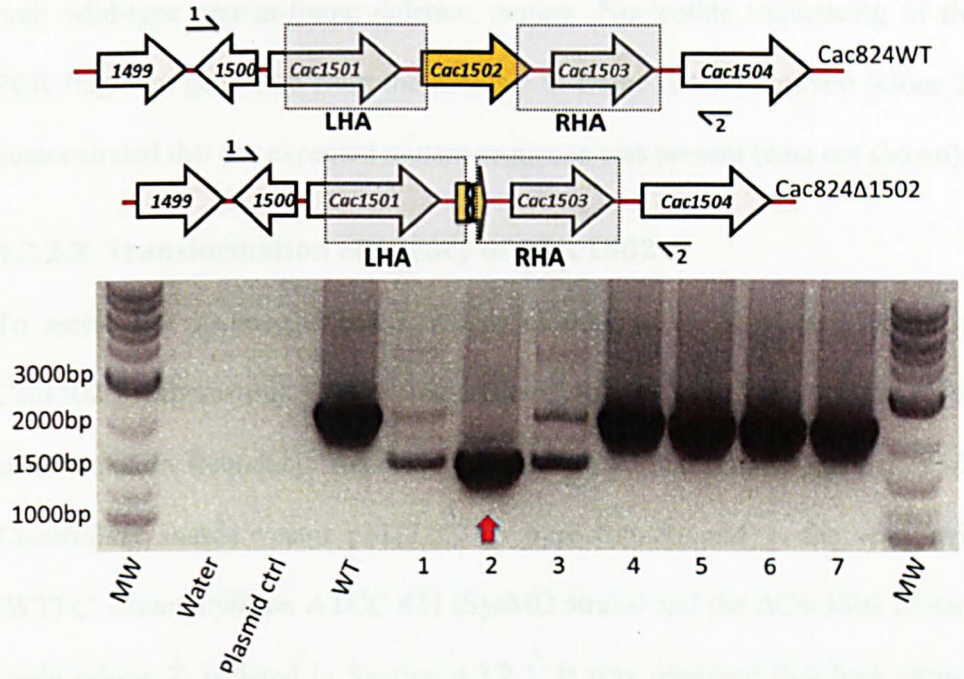


Figure 4.15 PCR screening of the in-frame deletion of Cac1502

PCR screening of clones using flanking primers Cac-1501-sF2 (1) and Cac-1504-sR1(2), PCR produced 1611bp and 2319bp fragments for *C. acetobutylicum* 824 ΔCac1502 and *C. acetobutylicum* 824 wild-type respectively. MW, 2-log DNA marker (NEB) molecular weight marker; Water, PCR-negative control; Plasmid, pMTL-SC7515::Cac1502-KO-1.5 plasmid DNA control; WT, *C. acetobutylicum* 824 genomic DNA control and 1-7 clones.

After successful creation of in-frame deletion mutant in *spo0A* of *C. acetobutylicum,* it was decided to make a deletion mutant of Cac1502. Plasmid pMTL-SC7515::Cac1502-KO-1.5 (Figure 2.3) containing the KO cassette for Cac1502, was transformed into *C. acetobutylicum* ATCC 824. The selection of single and double crossover was performed as described for *spo0A* in Section 4.3.2.1, using *codA* as negative selection marker and PCR screening of the expected double crossover mutants. A total of seven FC^R colonies were screened and only one of them (clone 2) was confirmed as an in-frame deletion mutant (Figure 4.15). Four of the clones (clones 4-7) appeared to have reverted back to wild-type whereas 1 and 3 appeared to be composed of a mixture of

both wild-type and in-frame deletion mutant. Nucleotide sequencing of the PCR fragment generated from the putative in-frame deletion mutant (clone 2) demonstrated that the expected mutant sequence was present (data not shown).

4.3.2.3 Transformation efficiency of ΔCac1502

To assess the phenotypic consequences of making an in-frame deletion of Cac1502, experiments were undertaken to measure any effects on transformation frequency. Accordingly, methylated and unmethylated *E. coli-Clostridium* shuttle vector pMTL85141 were transformed in the wild type (WT) *C. acetobutylicum* ATCC 824 (SysMO strain) and the ΔCac1502 mutant strain (clone 2) isolated in Section 4.3.2.2. It was observed that both strains could be transformed with methylated plasmids but theΔCac1502 mutant produce slightly more transformants (1.3×10^4 +/-39.19)/μg DNA than WT (2.6×10^3 +/-6.51)/μg DNA which were in contradiction to what was achieved for another strain of *C. acetobutylicum* DSM1731 and ClosTron mutant SMB009 of restriction gene Cac1502 (Dong et al., 2010). 1.5×10^4 +/-49.9 transformants per μg DNA were obtained when ΔCac1502 was transformed with unmethylated plasmid in contrast to wild type where no colonies were obtained. 2.6×10^3+/-6.51 transformants /μg DNA were obtained when methylated plasmid prepared from *E. coli* strain carrying pAN2 plasmid was used. DNA isolated from the *E. coli* strain carrying pAN2 contains a mixture of pAN2 and methylated test plasmid. In this case the exact amount of methylated plasmid in the transformation mixture is therefore likely to be lower than that used in experiments using unmethylated plasmid. All assays were repeated three times.

4.4 Discussion

Construction of mutants in the genus *Clostridium* has been hampered for many years due to a lack of effective gene tools. Whilst the ClosTron (Heap et al., 2007) has gone some way towards rectifying this deficiency, it is an insertional mutagen and, therefore, can cause polar effects. An improvement would be a gene tool that could be used where genes can be deleted without affecting the up- and downstream genes. Such stable gene deletions or replacements require an allelic exchange method, and preferably the availability of a negative/ counterselection marker which have not been reported for any *Clostridium* species until recently when *pyrF* was used as a negative selection marker to engineer the metabolic pathway of solventogenic *C. thermocellum* (Tripathi et al., 2010). Later on Nariya et al published the use of *galK* to make toxin gene mutants in *C. perfringens* using a suicide vector. *C. perfringens* has one of the highest transformation efficiency among clostridial strains (Nariya et al., 2011). During the course of this study, Cartman et al published the use of *codA* for making precise changes in the *tcdC* genotype in *C. difficile* (Cartman et al., 2012). Most recently, toxic gene *mazF* was used as an inducible counterselection marker for making stable double crossover gene deletions and integration in *C. acetobutylicum* ATCC 824 (Al-Hinai et al., 2012). Selection of double crossover integrants depended on the use of thiamphenicol marker flanked by FRT sites which were removed later on by expressing the FLP recombinase leaving behind a scar (Al-Hinai et al., 2012) which may affect the expression of genes up- and downstream of the target site. In addition, the FLP recombinase system is not yet established in several other members of the genus *Clostridium* including pathogens such *C. difficile* and *C. botulinum*

(Heap et al., 2010a; Kuehne et al., 2011). Therefore a universal system is required which can be used in a wide range of organisms for making gene deletions and integrations and including precise changes in the genotype.

The opportunity to develop an allelic exchange system based on the use of the counterselection marker *pyrE* was provided by ACE technology and the creation of a specially configured *pyrE* mutant strain of *C. acetobutylicum*. This strain was used as a host in which the *C. sporogenes* 15579 *pyrE* was used as a counterselection marker to make in-frame deletions in *spo0A* and Cap0168 (amylase). In the case of *spo0A*, sporulation minus phenotype was observed in mutants compared to wild-type by phase contrast microscopy and culture of heat shocked cells. In the case of Cap0168 deletion mutation, the phenotypic ability to form halos on starch iodine plates was significantly affected in the in-frame deletion mutant strain compared to the parental *C. acetobutylicum* ATCC 824 *pyrE* mutant strain (CRG CC# 1545 Table 2.1).

Furthermore, the creation of *pyrE* mutant by ACE method provided the opportunity to use this locus for complementation and over-expression of the deleted genes at *pyrE* locus at an appropriate gene dosage. Traditionally, complementation of a deleted or inactivated gene is achieved by introducing a native copy of the gene on a multicopy number plasmid. Frequently, the elevated copy number fails to restore the phenotype to wild-type. In one such study, a sigma factor σ^E, encoded by *spoIIG* gene required for the sporulation cascade in *B. subtilis*, was deleted and then complemented using *spoIIG* genes from *C. acetobutylicum* and *B. subtilis* on plasmids but the sporulation phenotype was only partially restored. This was attributed to the presence of several copies of the complementing genes present in the cells that disrupted

the optimum level of gene expression required for the normal sporulation phenotype. To probe how multiple gene copies influence chromosomal complementation, these authors integrated a functional copy of *B. subtilis spoIIG* into the chromosome and showed that the sporulation phenotype was exactly restored to the wild-type level (Arcuri et al., 2000). In a similar study in *C. acetobutylicum,* the sigma factor σ^F (Cac2306) was disrupted by single crossover integration of plasmid into the chromosome. Complementation of this mutant using a plasmid born copy of the *spoIIA* operon failed to restore the mutant to the wild-type sporulation phenotype. This operon was then integrated into the chromosome in the mutant to see if it could restore the mutant phenotype to wild-type level but the expression was still too high compared to wild-type due to the presence of the two copies of the operon in the *C. acetobutylicum* chromosome (Jones et al., 2011). The effect of multiple copies of genes is also evident from the studies undertaken with the *C. acetobutylicum* degenerated stains M5 and DG1, which have lost the pSOL1 megaplasmid, and therefore do not carry the genes required for solvent production such as acetone and butanol. When the missing genes were introduced back into the cell on multi-copy plasmids, the levels of acetone and butanol were not restored to levels equivalent to that of the wild-type (Cornillot et al., 1997; Sillers et al., 2008). Several other studies demonstrated that plasmid borne complementation, over-expresses the complemented genes *eg.*, the CwpV was over produced in a plasmid complemented *cwpV* mutant of *C. difficile* (Emerson et al., 2009) and approximately 3-fold higher expression of the perfringolysin O titre of a *virR* mutant of the *C. perfringens* strain 56 was

observed when complemented with a multicopy plasmid carrying the *virRS* genes compared to wild type strain (Cheung et al., 2010).

Auxiliary effects could also be due to the supplementation of antibiotics in the media to maintain the plasmid in the cells. One such case was observed in the case of complementation of the *sleC* mutation in *C. difficile* using a plasmid borne copy which did not restore the sporulation phenotype but did restore germination to a reduced level compared to wild-type (Burns et al., 2010). The presence of plasmid alone, or added antibiotic to maintain the plasmid adversely, affected the sporulation-germination process. Moreover plasmid borne complementation systems require the inclusion of a vector only control measurements in any experiment (Burns et al., 2010), increasing the amount of analysis that has to be performed. This is not needed when using ACE-mediated chromosomal complementation at the *pyrE* locus. Moreover the genes are stably integrated into chromosome with single dosage level which makes it the method of choice in those situations where plasmid stability achieved through the use of antibiotics supplementation is not a viable method, such as when complementation in an *in vivo* model is required.

The plasmids pMTL-ME6C and pMTL-ME7C position the complementing gene in the genome immediately downstream of the *pyrE* gene (Cac0027) in the *C. acetobutylicum* ATCC 824 chromosome. This allows read through from the upstream promoter responsible for transcription of *pyrE* in the case of pMTL-ME6C, the identity of which is unknown (but may reside immediately upstream of Cac0025) while in the case of pMTL-ME7C an additional terminator from *Lactococcus lactis* IL1403 strain between the *ldh* and *yngG*

genes was introduced downstream of *pyrE* (de Hoon et al., 2005) to stop transcription through from the *pyrE* promoter.

In certain cases it may be beneficial to bring about 'overexpression' of the complementing gene. For this purpose, two further vectors were constructed pMTL-ME6X and pMTL-ME7X following the same principles used for construction of pMTL-ME6C and pMTL-ME7C but carrying the promoter of the *ferredoxin* gene (*fdx*) of the *C. sporogenes*. As a consequence of its central role in anaerobic electron transfer, the *ferredoxin* gene is highly expressed in *Clostridium*, with the ferredoxin protein representing, for instance, up to 2% of the soluble protein in *C. pasteurianum* (Rabinowitz, 1972).

The sporulation phenotype was restored by complementing the *C. acetobutylicum* ATCC 824 *spo0A* mutant at the *pyrE* locus using pMTL-ME6C-PCacspo0A and also the over-expressed version pMTL-ME6X-Cacspo0A. The complementation and over-expression of *spo0A* showed high number of spore formation at 24 hrs time point compared to wild type (Figure 4.12). It was also observed that over-expression of spo0A has an effect on the colony phenotype (bigger colony size) compared to wild type or complemented version (Figure 4.13). The complementation of the Cap0168 amylase mutant and over-expression of the amylase gene Cap0168 is under progress and will be completed after the submission of this thesis.

During the course of this study, Cartman et al used *codA* as a counterselection marker in *C. difficile* to make precise changes and deletions in *tcdC* (Cartman et al., 2012). It was appropriate to explore the use of *codA*, in parallel to the *pyrE* based system, to make in-frame deletions in *C. acetobutylicum* ATCC

824. Again *spo0A* was successfully deleted using the *codA* system. The second target site was the restriction endonuclease system Cac1502 which was deleted in *C. acetobutylicum* ATCC 824 wild type and the mutant strain was successfully transformed with unmethylated plasmid DNA.

A number of Gram-positive replicons were explored in the past for use in various clostridial species (Heap et al., 2009). During the exemplification of the ACE method in *C. difficile,* a pCB102-based plasmid was initially used to make a *C. difficile* 630 Δ*erm, pyrE* mutant but the isolation of a pure single crossover integrant proved elusive. Success was achieved by inserting 6.5Kb lambda DNA into the ACE plasmid in order to reduce the segregational stability of the plasmid. The resultant increase in instability facilitated the isolation of pure single crossover integrants (Heap et al., 2012). Due to the problems associated with the cloning of allelic exchange cassettes into such a large plasmid (ca. 11kb), another approach was explored. It was suggested that a frame-shift mutation in the pCB102 replicon at the NsiI site causes the plasmid to be more unstable which would make the isolation of pure single crossover integrants easier (Nigel Minton personal communication). The restriction site NsiI (ATGCAT) in the pCB102 replicon in vector pMTL83151 was modified by deleting TGCA which resulted in a frame-shift mutation in the coding sequence, replacing the COOH-terminal region of RepH (CIKYYARSFKKAHVKKSKKKK) with LNIMGALKKLM. The resultant pMTL83*151 was found to be significantly unstable in *C. difficile* compared to pMTL83151 (Ng et al., 2013). The Gram-positive pCB102 replicon in pMTL-ME2 was, therefore, replaced with the frame-shift version from pMTL83*151 and the resultant vector called pMTL-YN3. This plasmid was then used to

make precise changes and in-frame deletions in *C. difficile* 630Δ*erm cwp84*, *mtlD* and *spo0A* genes using *pyrE* as a negative selection marker in a *pyrE* mutant host of *C. difficile*. The generated mutants were analysed in a wild type background after ACE-mediated repair of the *pyrE* gene back to wild-type using pMTL-YN1C. The mutant phenotypes were also complemented by insertion of the unmodified structural gene at the *pyrE* locus using ACE (Ng et al., 2013).

4.5 Key outcomes

The key findings of this chapter were

- The *C. acetobutylicum pyrE* mutant created using ACE technology was successfully used as a host to make in-frame deletions in *spo0A* and *amyB* gene using a heterologous *pyrE* as a counterselection marker.
- The feasibility of using ACE vectors to complement the mutants that had been made through insertion of the complementing gene at the *pyrE* locus was demonstrated.
- The method developed for *C. acetobutylicum* was applied to *C. difficile* (Ng et al., 2013) and several genes deleted, using *pyrE* as a counterselection marker, and complemented through insertion of the wild-type gene at the *pyrE* locus using ACE.
- The use of *codA* as another heterologous negative selection marker was also exploited by making in-frame deletions in *spo0A* and restriction endonuclease system Cac1502.

- The mutant of Cac1502 was successfully transformed with unmethylated DNA with higher transformation efficiency compared to wild-type.

Chapter Five: Isolation of new solventogenic Strains

5 Isolation of new solventogenic strains

5.1 Introduction

A variety of environmental sources (including potatoes and leguminous plant roots, manure, garden waste, agricultural soils, cereal crops, gooseberries, river bottom mud, decayed wood and a putrefying clam) have been used to isolate new *Clostridium* species which share some common characteristics, for example Gram-positive, spore formation, anaerobic and bacilli (Berezina et al., 2009; Calam, 1980; Jones & Woods, 1986; Montoya et al., 2000; Weyer & Rettger, 1927) but their habitats ranged from moderate to a very high temperature (Jones & Keis, 1995). They were first isolated in the early 20th century for the production of acetone for munitions manufacture to support their efforts in World Wars I and II (Jones & Woods, 1986). The type strain ATCC 824 of *C. acetobutylicum* was isolated from Connecticut garden soil in 1924 (Weyer & Rettger, 1927) and was most extensively used in the ABE fermentation process using starch and molasses as substrates for solvent production. Shortage of molasses and its high price were the main causes of loss of competitiveness of the ABE fermentation process for the production of solvents and other valuable chemicals which could also be obtained from petrochemicals. This situation triggered efforts towards the isolation of different strains able to grow on other substrates and produce cheaper biofuels and other chemicals for industry using ABE fermentation. Moreover the high energy prices and the political instability in the oil-rich countries have stimulated the search for new energy resources. This could enable them to be self-reliant in the energy sector and to protect their economies from market

volatility. Global warming, high carbon dioxide emissions from fossil fuels and petrochemical industries have also contributed to the interest on generating energy from environmentally friendly renewable resources. The alternative renewable sources should be less harmful to the environment and include solar energy, wind energy, bio-energy, geothermal energy, hydropower, ocean energy and biomass (Carere et al., 2008; Shunmugam, 2009). Plants biomasses could be used to reduce greenhouse gas emissions significantly (Carere et al., 2008; Crutzen et al., 2008). As this energy source is based on plant material and does not emit any toxic chemicals such as sulphur or carbon dioxide to the environment, these are considered to be environmentally friendly fuels (Shunmugam, 2009). An ideal strain for the ABE fermentation process would be able to utilise a wide range of substrates including crystalline avicel and produce a high butanol titre with the ability to survive in the presence of high butanol concentrations. To look for such a new species, samples from a range of sources including soil, plant wastes, mud were screened for new strains with the required characteristics.

5.2 Aim of this study

The aim of this study were

1. To isolate new solventogenic *Clostridium* species producing a high titre of butanol.
2. To apply the technologies developed during this project to enhance the commercial potential of strains, including extending their ability to utilise different substrates for solvent production.

5.3 Results

5.3.1 Isolation of solvent producing clostridial strains

A total of 24 soil/compost samples were collected from different agricultural sites. For example, some were collected from a garden in Nottinghamshire, UK where potato, kidney beans, French beans and beetroot were cultivated. Samples were screened for solvent producing bacteria which can utilize glucose and xylose as substrates using RCM and CBMx media. Acetone production was detected from purified culture by Rothera's test (Rothera, 1908). Some samples were also collected from different parts of China and Rome, and were collected and screened for solvent producing clostridia that can metabolize glucose and xylose and crystalline cellulose, (followed by purification of the culture).

5.3.2 Phenotypic characterization of new isolates

Microscopically all the new isolated strains/species were Gram-positive, rod-shaped and spore forming microorganisms (Figure 5.1 and Table 5.1The new isolates from soils where kidney bean and French beans were cultivated and soil samples C2 and C5 produced opaque and regular shaped colonies on CBMx agar plate similar to that of *C. acetobutylicum* ATCC 824 grown on CGM agar for 24hrs. Microorganisms isolated from environmental samples where beetroot and potato were cultivated and also soil samples C1, C4 produced translucent crenulated shaped colonies when grown on CBM_{CMC} for 48hrs. Colonies from sample C6 were tiny in size (ca.1mm in diameter) as compared to others which were ca. 3-4mm in diameter. All of these soil

samples produced copious gas after overnight growth in RCM or CBMx medium.

Table 5.1 Colony morphology of some of the isolates on agar media

Soil Sample	Media	Colony morphology
kidney bean	CBMx	3-4 mm, opaque, round
French beans	CBMx	opaque round
C5	CBMx	opaque round
C2	CBMx	3-4 mm opaque round
Beetroot	CBMcmc	translucent crenulated
Potato	CBMcmc	translucent crenulated
C1	CBMcmc	translucent crenulated
C4	CBMcmc	translucent crenulated
C6	CBMcmc	translucent crenulated
Rome	CBMcmc	translucent crenulated
Kidney beans	CBMcmc	3-4 mm opaque round

A total of 24 isolates were subjected to 16S rRNA sequence analysis. 12 of these isolates were isolated after growth on xylose and 12 on crystalline celluslose. 16S rRNA gene was amplified using primers 27F and 1492r and a band of ~1500bp was obtained Figure 5.2 . After gel purification, the region was sequenced using the Geneservice sequencing facility at Nottingham. The nucleotide sequences were subjected to BLAST analysis using the search engine available at http://blast.ncbi.nlm.nih.gov. The results of some of the isolates which showed similarities to the available strains in the NCBI database are shown in Table 5.2 .

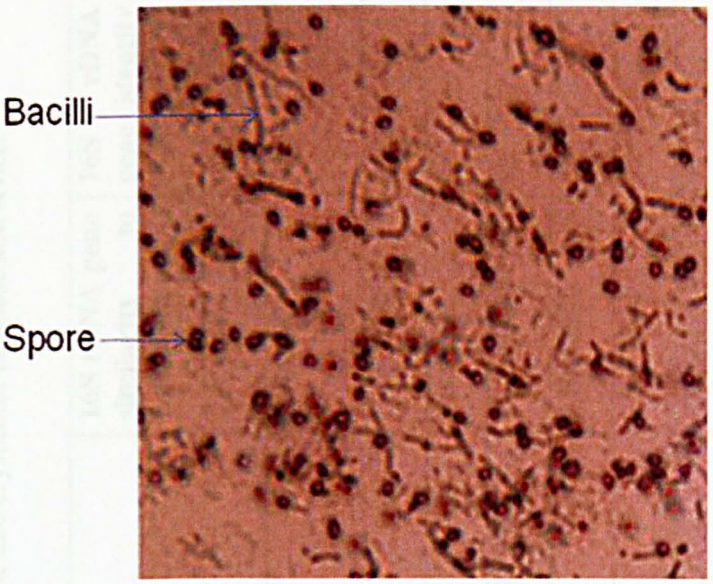


Figure 5.1 Microscopic Observation of soil isolates

An example of a soil isolate viwed by phase contrast microscopy showing spores and rod shaped cells.

Table 5.2 16S rRNA sequence analysis of the isolates obtained from soil samples after growth on xylose and crystalline cellulose and Avicel

Sample and source	Substrate	Organism	16S rDNA gene similarity to Accession no.	16S rDNA gene identity (%)
Sample 3 (C1 Small colonies)	Xylose	*Clostridium* sp. Zx5	EF052864	98
SAMPLE 4 (Kidney beans Big and opaque colonies)	Xylose	*Clostridium beijerinckii*	CP000721	99
SAMPLE 5 (C2 Small and opaque colonies)	Xylose	*Clostridium beijerinckii*	DQ839378	99
SAMPLE 7 (Kidney beans big colonies)	Xylose	*Clostridium* sp. MT10-4G	FJ009573	99
SAMPLE 8 (Potato)	Xylose	*Clostridium aciditolerans* strain JW/YJL-B3	DQ114945	99
SAMPLE 9 (C2 colonies)	Xylose	*Clostridium sordellii*	AB448946	99
SAMPLE 11 (C3)	Xylose	*Clostridium* sp. zx5	EF052864	98
SAMPLE 12 (C6)	Xylose	*Clostridium bifermentans*	FJ424479	97
SAMPLE 13 (C2 big opaque colonies)	Xylose	*Clostridium beijerinckii*	CP000721	99
SAMPLE 15 (C3)	CMC and Avicel	*Clostridium aciditolerans* strain JW/YJL-B3	DQ114945	99

SAMPLE 20 (small colonies)	CMC and Avicel	*Clostridium bifermentans*	FJ424483	99
SAMPLE 21 (C6)	CMC and Avicel	*Clostridium bifermentans*	AB538434	99
SAMPLE 23 (W.g III, 21/12/09)	CMC and Avicel	*Clostridium* sp. zx5	EF052864	98

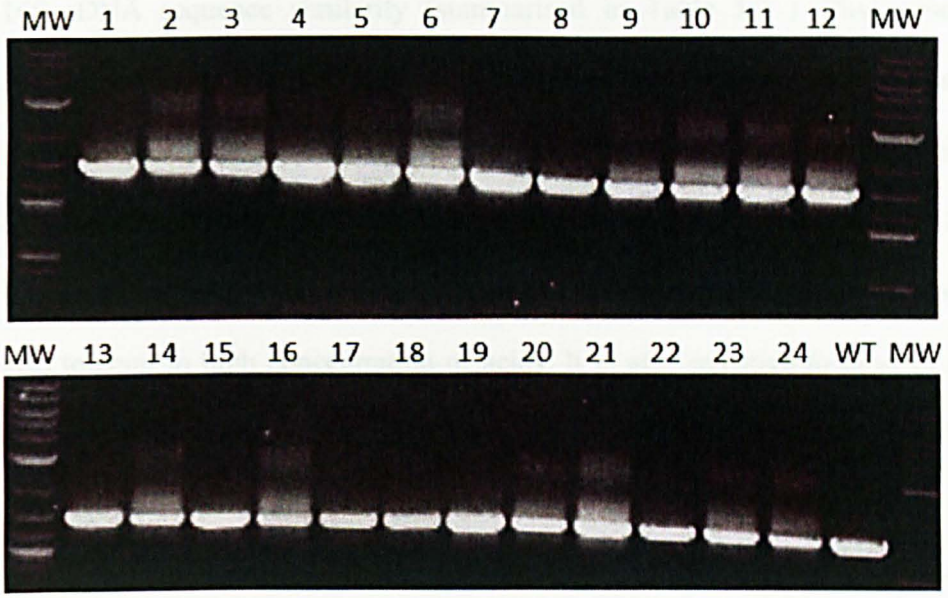


Figure 5.2 16S rDNA PCR screening

16S rDNA sequence was PCR amplified with primers 27F and 1492r and a band of ~1500bp was obtained using genomic DNAs from soil isolates as templates MW, 2-log DNA ladder (NEB) molecular weight marker; WT, *C. acetobutylicum* ATCC 824 and 1-24 soil isolates

5.4 Discussion

C. acetobutylicum ATCC 824 and other saccharolytic clostridia are unable to degrade lignocellulose which is the most abundant form of renewable energy resource in nature. Therefore we sought to isolate new saccharolytic solvent producing Clostridial strains which could degrade crystalline cellulose to produce high titre of butanol. Environmental samples (soil and compost) were collected from a wide range of cultivated lands. Xylose, carboxymethyl cellulose and crystalline cellulose Avicel^{R} were used as carbon sources in CBM medium to assess growth properties of newly isolated strains. After single-colony purification, genomic DNA was isolated and 16S rDNA sequence was determined using universal primers 27F and 1492R. A total of 13 different strains were identified amongst these twenty-four isolates based upon

16S rDNA sequence similarity (summarised in Table 5.2). Phylogenetic analysis based on the 16S rDNA sequence indicated that two of the isolated strains from samples 8 and 15 showed 99% similarity to *Clostridium aciditolerans* type strain JW/YJL-B3^T. According to Lee et al, strain JW/YJL-B3^T produces acetate, butyrate and ethanol by fermenting 20mM glucose and is also tolerant to high concentration of acids. It is also sensitive to most of the antibiotics currently used in Clostridial research (Lee et al., 2007). Consequently, this strain could have potential for metabolic engineering for the production of butanol. In *C. acetobutylicum* ATCC 824 metabolic pathway, butyrate is converted into butyryl-CoA by *ctfAB*-CoA-transferase which is then converted into butanol by the action of *adhE* and *bdhE* genes (Figure 1.1). Utilizing the ACE technology developed in this thesis in Section 3, these genes could be integrated into *C. aciditolerans* strain JW/YJL-B3^T and optimized for the production of butanol.

Two other purified strains from samples 4 and 13, showed similarity to *C. beijerinckii* NCIMB 8052. Several alcohol dehydrogenase genes have been identified in the type strain NCIMB 8052 suggesting great opportunity for production of biofuels (Wang et al., 2012). *C. beijerinckii* strains produce isopropanol encoded by alcohol dehydrogenase (*adh*) which is used as a valuable chemical in many forms in chemical industries and this pathway could be expressed in *C. acetobutylicum* 824 to produce isopropanol in place of acetone (Lee et al., 2012). Three additional isolated strains from soil samples 12, 20 and 21 showed 97-99% similarity to various *C. bifermentans* strains which can be used in bioremediation. Chemical industries producing nitroorganic chemicals generate various hazardous chemicals which can

contaminate the environment. Some bacteria have evolved to use these contaminants as substrates and are therefore able to rid the environment of these toxic products. *C. bifermentans* is one such organism and has been used for the biotransformation of highly toxic 2, 4, 6-Trinitrotoluene (TNT) into toluene which is a non-toxic substance (Lewis et al., 1996).

Due to insufficient time, I was unable to thoroughly assess these isolated strains for ability of genetic modification and sensitiveness to antibiotic selection markers which needs to be done.

5.5 Key outcomes

The key outcomes of this chapter are

- *Clostridium* species are widely distributed in environmental samples such as mud, plant decomposed material, soil,
- New *Clostridium* can be isolated and classified on the basis of 16S rDNA sequence.

Chapter Six: General Discussion

6 General discussion

6.1 Discussion

Clostridium is the only genus which naturally produces the solvents Acetone, Butanol and Ethanol by ABE fermentation process using a wide range of pentose and hexose sugars. This process was made uneconomical by the development of the petrochemical industry (Jones & Woods, 1986). To restore ABE fermentation as a viable commercial process will require the development of more effective bacterial strains. Until now this goal has been hampered by a dearth of effective gene tools. The developments made in this thesis go a long way towards rectifying the situation.

The development of Targetron and ClosTron technologies played a major role in knocking out bacterial genes and could also be used for integration of synthetic genes of upto 1kb size. More effective methods which can deliver large segments of DNA are clearly required, particularly for the successful application of synthetic biology approaches to enhancing the properties of commercially useful strains for biofuel production. DNA can be integrated into the chromosome of organisms in one step by homologous recombination where the transformation frequency is very high such as in yeast and *B. subtilis* (Itaya & Tanaka, 1990). More complex genetic material could be integrated into the chromosome in multiple steps using iterative selection markers flanked by the sequences of interests as in the "domino" method for *Bacillus* (Itaya et al., 2008).

Here a generally applicable method is described to deliver DNA to organisms lacking the required sophisticated gene tools particularly industrially important strains of solventogenic clostridia. Double crossover clones are isolated at certain genomic loci without the use of negative/counter selection markers. Crucially, expression of selectable marker is achieved by the coupling of the two non-functional fragments of the marker gene in the desired double crossover recombinant chromosome. The process was named "Allele-Coupled Exchange (ACE)". This approach has two specialised characteristic: 1) cells which have undergone a particular recombination event can be selected if counter-selectable marker is interrupted or two parts of a positively selectable marker are fused together to give a phenotype; and 2) regions of homology of very different length strongly influence the order of homologous recombination events to achieve a selectable phenotype in the second recombination event in the double crossover integrants.

Genes such as *pyrE* and *pyrF* which encode the pyrimidine biosynthesis enzymes *orotate phosphoribosyl transferase* and *orotidine 5-phosphate decarboxylase* respectively were chosen which could be selected positively and negatively. Cells which contain *pyrE* and *pyrF* can be selected on growth medium lacking uracil, as they are required for uracil biosynthesis. Conversely, cells that lack either *pyrE* or *pyrF* can be selected on growth medium supplemented with 5-fluoroorotic acid (FOA), as this compound is highly toxic only to cells which contain this pathway. FOA has been used for counter selection for some time (Boeke et al., 1984) and in various organisms (Deng et al., 2009; Husson et al., 1990; Knipfer et al., 1997) including very recently in *C. thermocellum* (Tripathi et al., 2010). Asymmetric homology arms are used

(in contrast to general allelic exchange where equal homology arms are used and both have equal chances of integration into chromosome) to control the recombination events. A long region of homology of 1200bp leads the integration of plasmid into the chromosome producing primarily single crossover integrants at this homology arm. A second crossover at the short region of homology provided by a 300bp internal fragment from *pyrF* or *pyrE* would excise the plasmid resulting in FOA-resistant double crossover clones in which *pyrE* or *pyF* is inactivated and the DNA inserted is delivered stably to the chromosome (Figure 3.1and Figure 3.2). This type of allelic exchange in which a negatively selectable marker is replaced could be achieved using a conventional allelic exchange cassette (Knipfer et al., 1997) but the *pyrE* allele in the vector pMTL-JH12 was specially configured to facilitate subsequent genetic modification (Figure 3.8). The *pyrE* in the recombinant strain was not deleted completely but only truncated at the 3`end. In a second step of homologous recombination process, this truncated *pyrE* and its partial counterpart were coupled together to achieve a full-length functional gene as described for vector pMTL-JH14 (Figure 3.8). The coupling of the two non-functional parts of *pyrE* into one functional *pyrE* allowed the cells to be selected on minimal medium *ie.,* they are now prototrophic for pyrimidine biosynthesis (page 93, 96).

As stated above two non-functional *pyrE* fragments were effectively fused to form one fully functional gene via homologous recombination events. In the second approach promoter-less selectable markers were fused to a chromosomally located strong promoter to drive their expression and select the double crossover integrants phenotypically. A macrolide–lincosamide–

streptogramin (MLS) antibiotic resistance marker *ermB*, and a heterologous *pyrE* from *C. sporogenes* ATCC 15579 were selected to integrate downstream of the constitutively expressed *thiolase* promoter in *C. acetobutylicum* wild-type and *pyrE* mutant using pMTL-JH16 and pMTL-JH31 vectors, respectively.

The expression of heterologous markers from chromosomally located primers allows the gene to be cloned without a promoter to avoid toxic effects in the host strain resulting from expression. Positive selection of the double crossover clones using antibiotic markers are widely applicable to a broad range of microorganisms compared to negative selection markers. These markers can be re-cycled and have the potential to introduce large amounts of complex DNA into the chromosome in multi-step process. To assess the proficiency of the method to deliver cargo DNA, initially various HindIII lambda DNA fragments ranged from 2.0-6.5 kb were integrated at the *pyrF* locus. In the multistep ACE procedure, almost the whole lambda genomic DNA was integrated into the *C. acetobutylicum* ATCC 824 *pyrE* mutant chromosome. This is the largest arbitrary sequence integrated in any *Clostridium* by an order of magnitude (Chen et al., 2007; Heap et al., 2010b). Transformation efficiency of the *C. acetobutylicum* ATCC 824 *pyrE* mutant (CRG CC# 1545) was severely affected by large sized plasmid for example pMTL-JH30::L12 and pMTL-JH16::L28 (18Kb-34Kb) and several attempts were made to construct the recombinant strains L28 and L28-L12 in section 3.3.8. It was suggested that the size of inserts used in multistep strategies, at least in *C. acetobutylicum*, should strike a balance between the number of steps and the ease of each step and should be approximately 10Kb.

Allelic exchange is a common method of gene deletion in many microorganisms where a suicide vector is used to deliver an inactivated allele to the chromosome. However, for some microorganisms such as solventogenic *Clostridium* species, this approach has not previously been possible. A number of counterselection markers have been used in other organisms such as *sacB* (Donnenberg & Kaper, 1991; Marx, 2008; Ried & Collmer, 1987) *pyrF* and *pyrE* (Bitan-Banin et al., 2003; Knipfer et al., 1997), *galK* (Ueki et al., 1996), *codA* (van der Geize et al., 2008), *upp* (Fabret et al., 2002) and *mazF* (Zhang et al., 2006). Before the start of this project in August 2009, there were no such counterselection markers available for the genus *Clostridium* except the *upp* based system which was not publically available due to patent protection. Recently, *C. acetobutylicum* ATCC 824 *galK* and *C. thermocellum pyrF* genes were used as negative selection markers in *C. perfringens* (Nariya et al., 2011) and in *C. thermocellum* (Tripathi et al., 2010).

The construction of *C. acetobutylicum pyrE* mutant (CRG CC#1545) by ACE method provided the opportunity to use *C. sporogenes* ATCC 15579 *pyrE* as a heterologous negative selection markers and in-frame deletions were made in *C. acetobutylicum* ATCC 824 genes such as *spo0A,* the master regulator of spore formation and *amyB* involved in starch degradation as described in Sections 4.3.1.2 and 4.3.1.6.

The specially configured *pyrE* locus in *C. acetobutylicum* 824 *pyrE* mutant (CRG CC# 1545) was used to complement and over-express the deleted genes at gene dosage level at the *pyrE* locus in the chromosome compared to the plasmid born complementation which has a number of issues and not ideal for *in vivo* work. Although significant difference was not observed in the case of

the over-expression of *spo0A* which could be due to the weakness of the fdx promoter used in this study compared to *spo0A* native promoter or it's over-expression may require some regulatory factors. The robustness and reliability of the *pyrE*-based system has been demonstrated in *C. difficile* and *C. acetobutylicum* and is under progress in other *Clostridium* and non-*Clostridium* species in our laboratory.

During the course of this study, our Clostridia Research Group (CRG) at Nottingham also developed a method to make precise changes in the *C. difficile tcdC* gene using a heterologous negative selection marker, the *codA* gene from *E. coli* (Cartman et al., 2012). The application of *codA* was also exploited in saccharolytic clostridia and in-frame deletions have been created in *spo0A* and restriction endonuclease system Cac1502. Both of these systems are equally applicable but the *pyrE* system has greater advantages in terms of isolation of the double crossover mutants as selection with FOA is generally easier compared to use of FC in where the majority of the FC^R colonies are random mutant of *codA* and still carry the plasmid. Another big advantage of the *pyrE* system is that it can be used to complement and over-express a deleted gene at the *pyrE* locus concomitant to revert the *pyrE* mutant host strain back to wild type.

Although attempts were made to overcome the degeneration problem associated with *C. acetobutylicum* and integrate the pSOL1 plasmid using a combination of ClosTron and ACE technologies, this approach was not successful. It proved difficult to detect an intron integration event at the Cap0168 (*amylase*) locus without the use of RAM marker. The genetic methods developed in this thesis could be used to do metabolic engineering in

saccharolytic clostridia to enhance its substrate utilisation capability and make the best possible recombinant strains for industrial ABE fermentation to produce high value chemicals such as isopropanol and butanol.

C. acetobutylicum ATCC 824 and other saccharolytic clostridia are unable to degrade lignocellulose which is the most abundant form of renewable energy resource in nature. Screening and isolation of new clostridial strain was not the main aim of this project but during the time available, attempts were made to isolate a new saccharolytic solvent producing Clostridial strain which could degrade crystalline cellulose to produce a high titre of butanol. Environmental samples were collected from a wide range of cultivated lands consists of soils and compost. Phylogenetic analysis based on the 16S rDNA sequence indicated that two of the isolated strains from samples 8 and 15 in (Table 5.2) showed 99% similarity to *C. aciditolerans* type strain JW/YJL-B3^T. According to Lee et al this strain produces acetate, butyrate and ethanol by fermenting 20mM glucose and is also tolerant to high concentration of acids and is sensitive to most of the antibiotics currently used in Clostridia research (Lee et al., 2007). This can be a good organism for metabolic engineering for the production of butanol from butyrate by integrating few genes in *C. aciditolerans* type strain JW/YJL-B3^T from *C. acetobutylicum* ATCC 824. Two other purified strains from samples (4 and 13) showed similarity to *C. beijerinckii* NCIMB 8052. Several alcohol dehydrogenase genes have been identified in the type strain *C. beijerinckii* NCIMB 8052 suggesting great opportunity for production of biofuels (Wang et al., 2012). *C. beijerinckii* strains produces isopropanol encoded by alcohol dehydrogenase (*adh*) which is used as a valuable chemical in many form in chemical industries and this

pathway could be expressed into *C. acetobutylicum* 824 to produce isopropanol rather than acetone (Lee et al., 2012).

6.2 Proposed future work

Having showed the utility of using *pyrE* mutant strains as the host for in-frame deletions, the final part of the study, needed to allow eventual publication, is the demonstration of over-expression of the complementing gene. As the early attempt to use pectinase genes was unsuccessful, attention was turned to the amylase gene CaP0168 (*amyB*). These studies need to be completed by comparing expression using both chromosomal and plasmid-borne complementation with phenotypic evaluation of the complemented strains on CBM medium containing starch and glucose to analyse amylase activity.

Although, the in-frame deletion systems using *pyrE* or *codA* alleles as counter selection markers could be used to make the precise changes needed to generate strains with improved productivity and substrate utilisation, method efficiency issues still remain. The isolation of pure single crossover integrants is crucial during the implementation of the method. Currently, the Gram-positive replicon pIM13 is used in *C. acetobutylicum* ATCC 824 to isolate a pure single crossover integrant in the presence of the *catP* marker. It is the ease with which the single crossover integrants are isolated that currently represents the main bottleneck in the process. The method could be improved if it were possible to engineer a greater degree of 'defectiveness' into the plasmid replicon, or even complete replication deficiency. A completely replication deficient plasmid, in effect a suicide plasmid, would need to be delivered at a much higher transformation frequency than is currently possible (2.6×10^3 +/-

6.51 transformants per μg DNA). Experiments aimed at increasing electroporation frequencies would, therefore, be a worthwhile investment of effort. Such studies could be aided by making further in-frame deletions in the host strain of other restriction genes which may be limiting transfer frequencies, such as Cac3535 and Cac2824.

The alternative approach would be to make the pseudo-suicide vector used less effective at replication, thereby increasing the distinction between cells carrying the autonomous plasmid and those cells in which it is integrated. This could be achieved either by inserting an extra piece of DNA into the plasmid, increasing the size of the plasmid and making it less effective at replication, or, perhaps more logically, by modifying the replication region of the plasmid, thereby reducing its efficiency. The pCB102 replicon-derivative generated in this study is an example of such a plasmid, but unfortunately although it reduced replication efficiency in *C. difficile*, its effectiveness in *C. acetobutylicum* was unaffected. As an alternative, modifications to the pIM13 replicon of pMTL-ME3 could be made. These could include modifications to the Rep gene, or changes to the origin of replication (truncations or DNA substitutions).

Degeneration of solventogenic strains is a big issue in the commercial sector and to try to overcome this, ClosTron technology was tried to deliver the inverse of SHA:LHA of *pyrE/hydA* locus into pSOL1 to convert it into an ACE-like vector which could then integrate into *C. acetobutylicum* chromosome by rACE method. However this approach was not successful. Currently further work is in progress to deliver the inverse of SHA:LHA of *pyrE/hydA* locus into pSOL1 at the Cap0168 locus using the ACE method. The

successful integration of the inverse sequence at Cap0168 will result into an amylase-negative strain and the pSOL1 will be converted into an ACE-like vector with two asymmetric homology arms flanking the whole pSOL1 in reverse orientation targeting the *pyrE/hydA* locus. The resultant double crossover integrant strain with pSOL1 will be *pyrE*-positive and will be stable in terms of pSOL1.

C. acetobutylicum ATCC 824 is a model solventogenic strain used for the production of valuable solvents from simple sugars but due to low titre and its inability to degrade crystalline cellulolytic material, this process has not yet been made commercially viable. Here, an initial attempt was made to isolate a cellulosic butanol producing strain. However, due to time constraints those strains that were isolated were not thoroughly characterised. Thus, they need to be more thoroughly characterised in terms of solvent yield and ability to utilise cellulosic feedstocks. Crucially, the ability to obtain gene transfer in any selected strain would need to be established. On the other hand, *C. cellulolyticum* and other cellulose degrading organism can degrade avicel^{R} (a crystalline cellulose material) but are unable to produce valuable products such as biobutanol. The Allelic exchange method developed in this thesis *ie.*, Allele-Coupled Exchange (ACE) could be used to integrate various cellulolytic genes for example CipA, Cel48S, Cel9K from other organisms into *C. acetobutylicum* ATCC 824 to improve its substrate utilisation ability and make the ABE process a commercially feasible process.

6.3 Concluding remarks

Gene knockout and knock-in technologies are pivotal both to better understand the biology and exploit the commercial properties of *Clostridium* species. Before this study, there were very few methods available to make insertional mutation or gene deletions and all had drawbacks. In this study, a gene integration system has been developed based on homologous recombination called Allele Coupled-Exchange (ACE) which was used to integrate almost the whole phage lambda genome into the chromosome of *C. acetobutylicum* ATCC 824 *pyrE* strain. The creation of the *C. acetobutylicum pyrE* mutant by ACE technology provided the opportunity to exploit the use of the heterologous *pyrE* gene as counterselection marker for making in-frame deletion or precise changes at other loci within the chromosome. Moreover, it provides the facility whereby inactivated genes may be subsequently complemented at an appropriate gene dosage through the rapid insertion of the wild-type gene at the *pyrE* locus. This facility offers considerable advantage over the use of multicopy, autonomous complementation vectors. Another counter selection marker *codA* was also exploited in saccharolytic clostridia to make in-frame deletion mutants. Furthermore, ACE methodology could be used to solve the degeneration problem of *C. acetobutylicum* by integrating the pSOL1 into chromosome.

The technologies developed in this study have now been demonstrated in other clostridial species within the Clostridia Research Group at Nottingham. These could be used to engineer the metabolic pathways of various *Clostridium* species to produce commercially valuable chemicals and enhance their substrate utilisation ability.

Chapter Seven: Bibliography

7 Bibliography

AL-HINAI, M. A., FAST, A. G. & PAPOUTSAKIS, E. T. (2012). Novel system for efficient isolation of *Clostridium* double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and DNA integration. *Applied and Environmental Microbiology* **78,** 8112-21.

ALLCOCK, E. R. & WOODS, D. R. (1981). Carboxymethyl Cellulase and Cellobiase Production by *Clostridium-Acetobutylicum* in an Industrial Fermentation Medium. *Applied and Environmental Microbiology* **41,** 539-541.

ALSAKER, K. V., SPITZER, T. R. & PAPOUTSAKIS, E. T. (2004). Transcriptional analysis of spo0A overexpression in *Clostridium acetobutylicum* and its effect on the cell's response to butanol stress. *Journal of Bacteriology* **186,** 1959-71.

ANDERSCH, W., BAHL, H. & GOTTSCHALK, G. (1983). Level of Enzymes Involved in Acetate, Butyrate, Acetone and Butanol Formation by *Clostridium-acetobutylicum*. *European Journal of Applied Microbiology and Biotechnology* **18,** 327-332.

ARCURI, E. F., WIEDMANN, M. & BOOR, K. J. (2000). Phylogeny and functional conservation of sigma(E) in endospore-forming bacteria. *Microbiology* **146 (Pt 7),** 1593-603.

BAER, S. H., BLASCHEK, H. P. & SMITH, T. L. (1987). Effect of Butanol Challenge and Temperature on Lipid Composition and Membrane Fluidity of Butanol-Tolerant *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* **53,** 2854-61.

BAHL, H., ANDERSCH, W., BRAUN, K. & GOTTSCHALK, G. (1982). Effect of Ph and Butyrate Concentration on the Production of Acetone and Butanol by *Clostridium-acetobutylicum* Grown in Continuous Culture. *European Journal of Applied Microbiology and Biotechnology* **14,** 17-20.

BARTLETT, J. G., MOON, N., CHANG, T. W., TAYLOR, N. & ONDERDONK, A. B. (1978). Role of *Clostridium difficile* in antibiotic-associated pseudomembranous colitis. *Gastroenterology* **75,** 778-82.

BENOIT, L., CAILLIEZ, C., PETITDEMANGE, E. & GITTON, J. (1992). Isolation of cellulolytic mesophilic clostridia from a municipal solid waste digestor. *Microbial Ecology* **23,** 117-125.

BEREZINA, O. V., BRANDT, A., YAROTSKY, S., SCHWARZ, W. H. & ZVERLOV, V. V. (2009). Isolation of a new butanol-producing *Clostridium* strain: high level of hemicellulosic activity and structure of solventogenesis genes of a new *Clostridium saccharobutylicum* isolate. *Systematic and Applied Microbiology* **32,** 449-59.

BHATNAGAR, J., DELEON-CARNES, M., KELLAR, K. L., BANDYOPADHYAY, K., ANTONIADOU, Z. A., SHIEH, W. J., PADDOCK, C. D. & ZAKI, S. R. (2012). Rapid, simultaneous detection of *Clostridium sordellii* and *Clostridium perfringens* in archived tissues by a novel PCR-based microsphere assay: diagnostic implications for pregnancy-associated toxic shock syndrome cases. *Infectious Diseases in Obstetrics and Gynecology* **2012,** 972845.

BIELY, P. (1985). Microbial Xylanolytic Systems. *Trends in Biotechnology* **3,** 286-290.

BITAN-BANIN, G., ORTENBERG, R. & MEVARECH, M. (2003). Development of a gene knockout system for the halophilic archaeon Haloferax volcanii by use of the *pyrE* gene. *Journal of Bacteriology* **185,** 772-8.

BOEKE, J. D., LACROUTE, F. & FINK, G. R. (1984). A Positive Selection for Mutants Lacking Orotidine-5'-Phosphate Decarboxylase Activity in Yeast - 5-Fluoro-Orotic Acid Resistance. *Molecular & General Genetics* **197,** 345-346.

BORDEN, J. R. & PAPOUTSAKIS, E. T. (2007). Dynamics of genomic-library enrichment and identification of solvent tolerance genes for *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* **73,** 3061-8.

BROSIUS, J., DULL, T. J., SLEETER, D. D. & NOLLER, H. F. (1981). Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. *Journal of Molecular Biology* **148,** 107-27.

BROSIUS, J., PALMER, M. L., KENNEDY, P. J. & NOLLER, H. F. (1978). Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. *Proceedings of the National Academy of Sciences* **75,** 4801-5.

BURNS, D. A., HEAP, J. T. & MINTON, N. P. (2010). SleC is essential for germination of *Clostridium difficile* spores in nutrient-rich medium supplemented with the bile salt taurocholate. *Journal of Bacteriology* **192,** 657-64.

CAI, X. & BENNETT, G. N. (2011). Improving the *Clostridium acetobutylicum* butanol fermentation by engineering the strain for co-production of riboflavin. *Journal of Industrial Microbiology and Biotechnology* **38,** 1013-25.

CALAM, C. T. (1980). Isolation of *Clostridium-Acetobutylicum* Strains Producing Butanol and Acetone. *Biotechnology Letters* **2,** 111-116.

CARERE, C. R., SPARLING, R., CICEK, N. & LEVIN, D. B. (2008). Third Generation Biofuels via Direct Cellulose Fermentation. *International Journal of Molecular Sciences* **9,** 1342-60.

CARTMAN, S. T., KELLY, M. L., HEEG, D., HEAP, J. T. & MINTON, N. P. (2012). Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the tcdC genotype and toxin production. *Applied and Environmental Microbiology* **78,** 4683-90.

CHEN, C. K. & BLASCHEK, H. P. (1999). Acetate enhances solvent production and prevents degeneration in *Clostridium beijerinckii* BA101. *Applied Microbiology and Biotechnology* **52,** 170-3.

CHEN, Y., CARUSO, L., MCCLANE, B., FISHER, D. & GUPTA, P. (2007). Disruption of a toxin gene by introduction of a foreign gene into the chromosome of *Clostridium perfringens* using targetron-induced mutagenesis. *Plasmid* **58,** 182-189.

CHEUNG, J. K., KEYBURN, A. L., CARTER, G. P., LANCKRIET, A. L., VAN IMMERSEEL, F., MOORE, R. J. & ROOD, J. I. (2010). The VirSR two-component signal transduction system regulates NetB toxin production in *Clostridium perfringens*. *Infection and Immunity* **78,** 3064-72.

CLEWELL, D. B., YAGI, Y., DUNNY, G. M. & SCHULTZ, S. K. (1974). Characterization of three plasmid deoxyribonucleic acid molecules in a strain of Streptococcus faecalis: identification of a plasmid determining erythromycin resistance. *Journal of Bacteriology* **117,** 283-9.

COOKSLEY, C. M., ZHANG, Y., WANG, H., REDL, S., WINZER, K. & MINTON, N. P. (2012). Targeted mutagenesis of the *Clostridium acetobutylicum* acetone-butanol-ethanol fermentation pathway. *Metabolic Engineering* **14,** 630-41.

CORNILLOT, E., NAIR, R. V., PAPOUTSAKIS, E. T. & SOUCAILLE, P. (1997). The genes for butanol and acetone formation in *Clostridium acetobutylicum* ATCC 824 reside on a large plasmid whose loss leads to degeneration of the strain. *Journal of Bacteriology* **179,** 5442-5447.

CORNILLOT, E. & SOUCAILLE, P. (1996). Solvent-forming genes in clostridia. *Nature* **380,** 489-489.

COUSINEAU, B., SMITH, D., LAWRENCE-CAVANAGH, S., MUELLER, J. E., YANG, J., MILLS, D., MANIAS, D., DUNNY, G., LAMBOWITZ, A. M. & BELFORT, M. (1998). Retrohoming of a bacterial group II intron: Mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* **94,** 451-462.

CRUTZEN, P. J., MOSIER, A. R., SMITH, K. A. & WINIWARTER, W. (2008). N2O release from agro-biofuel production negates global warming reduction by replacing fossil fuels. *Atmospheric Chemistry and Physics* **8,** 389-395.

DATSENKO, K. A. & WANNER, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences U S A* **97,** 6640-5.

DE HOON, M. J., MAKITA, Y., NAKAI, K. & MIYANO, S. (2005). Prediction of transcriptional terminators in *Bacillus subtilis* and related species. *PLoS Computational Biology* **1,** e25.

DE MENORVAL, M. A., MIR, L. M., FERNANDEZ, M. L. & REIGADA, R. (2012). Effects of dimethyl sulfoxide in cholesterol-containing lipid membranes: a comparative study of experiments in silico and with cells. *PLoS One* **7,** e41733.

DENG, L., ZHU, H., CHEN, Z., LIANG, Y. X. & SHE, Q. (2009). Unmarked gene deletion and host-vector system for the hyperthermophilic crenarchaeon Sulfolobus islandicus. *Extremophiles* **13,** 735-46.

DONG, H., ZHANG, Y., DAI, Z. & LI, Y. (2010). Engineering *Clostridium* strain to accept unmethylated DNA. *PLoS One* **5,** e9038.

DONNENBERG, M. S. & KAPER, J. B. (1991). Construction of an eae deletion mutant of enteropathogenic *Escherichia coli* by using a positive-selection suicide vector. *Infection and Immunity* **59,** 4310-7.

DURRE, P. (2007). Biobutanol: an attractive biofuel. *Journal of Biotechnology* **2,** 1525-34.

DURRE, P. (2008). Fermentative butanol production: bulk chemical and biofuel. *Annals of the New York Academy of Sciences* **1125,** 353-62.

DURRE, P., KUHN, A., GOTTWALD, M. & GOTTSCHALK, G. (1987). Enzymatic Investigations on Butanol Dehydrogenase and Butyraldehyde Dehydrogenase in Extracts of *Clostridium-Acetobutylicum*. *Applied Microbiology and Biotechnology* **26,** 268-272.

EMERSON, J. E., REYNOLDS, C. B., FAGAN, R. P., SHAW, H. A., GOULDING, D. & FAIRWEATHER, N. F. (2009). A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein. *Molecular Microbiology* **74,** 541-56.

EZEJI, T. & BLASCHEK, H. P. (2008). Fermentation of dried distillers' grains and solubles (DDGS) hydrolysates to solvents and value-added products by solventogenic clostridia. *Bioresource Technology* **99,** 5232-42.

EZEJI, T., MILNE, C., PRICE, N. D. & BLASCHEK, H. P. (2010). Achievements and perspectives to overcome the poor solvent resistance in acetone and butanol-producing microorganisms. *Applied Microbiology and Biotechnology* **85,** 1697-712.

EZEJI, T. C., QURESHI, N. & BLASCHEK, H. P. (2003). Production of acetone, butanol and ethanol by *Clostridium beijerinckii* BA101 and in situ recovery by gas stripping. *World Journal of Microbiology & Biotechnology* **19,** 595-603.

EZEJI, T. C., QURESHI, N. & BLASCHEK, H. P. (2007). Bioproduction of butanol from biomass: from genes to bioreactors. *Current Opinion in Biotechnology* **18,** 220-7.

FABRET, C., DUSKO EHRLICH, S. & NOIROT, P. (2002). A new mutation delivery system for genome-scale approaches in *Bacillus subtilis*. *Molecular microbiology* **46,** 25-36.

FOND, O., ENGASSER, J. M., MATTA-EL-AMOURI, G. & PETITDEMANGE, H. (1986a). The acetone butanol fermentation on glucose and xylose. I. Regulation and kinetics in batch cultures. *Biotechnology and Bioengineering* **28,** 160-6.

FOND, O., ENGASSER, J. M., MATTA-EL-AMOURI, G. & PETITDEMANGE, H. (1986b). The acetone butanol fermentation on glucose and xylose. II. Regulation and kinetics in fed-batch cultures. *Biotechnology and Bioengineering* **28,** 167-75.

FOND, O., PETITDEMANGE, E., PETITDEMANGE, H. & ENGASSER, J. M. (1983). Cellulose Fermentation by a Coculture of a Mesophilic Cellulolytic *Clostridium* and *Clostridium Acetobutylicum*. *Biotechnology and Bioengineering*, 217-224.

FU, A. S., LIU, R., ZHU, J. & LIU, T. G. (2011). [Genetic engineering of microbial metabolic pathway for production of advanced biodiesel]. *Yi Chuan* **33,** 1121-33.

GABRIEL, C. L. (1928). Butanol fermentation process. *Industrial and Engineering Chemistry* **20,** 1063-1067.

GABRIEL, C. L. & CRAWFORD, F. M. (1930). Development of the butyl-acetonic fermentation industry. *Industrial and Engineering Chemistry* **22,** 1163-1165.

GARNIER, T. & COLE, S. T. (1986). Characterization of a bacteriocinogenic plasmid from *Clostridium perfringens* and molecular genetic analysis of the bacteriocin-encoding gene. *Journal of Bacteriology* **168,** 1189-96.

GARRITY, G. M. A. & HOLT, J. G. (2001). The road map to the manual. In: D.R. Boone and R.W. Castenholz, Editors, Manual of Systematic Bacteriology, the Archaea and the Deeply Branching and Phototrophic Bacteria (Second ed), Springer, New York, USA. 119–166.

GHESHLAGHI, R., SCHARER, J. M., MOO-YOUNG, M. & CHOU, C. P. (2009). Metabolic pathways of clostridia for producing butanol. *Biotechnology Advances* **27,** 764-781.

GIBSON, D. G., YOUNG, L., CHUANG, R. Y., VENTER, J. C., HUTCHISON, C. A., 3RD & SMITH, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6,** 343-5.

GREEN, E. M. & BENNETT, G. N. (1996). Inactivation of an aldehyde/alcohol dehydrogenase gene from *Clostridium acetobutylicum* ATCC 824. *Applied Biochemistry and Biotechnology* **57-58,** 213-21.

GREEN, E. M., BOYNTON, Z. L., HARRIS, L. M., RUDOLPH, F. B., PAPOUTSAKIS, E. T. & BENNETT, G. N. (1996). Genetic manipulation of acid formation pathways by gene inactivation in *Clostridium acetobutylicum* ATCC 824. *Microbiology* **142 (Pt 8),** 2079-86.

GU, Y., HU, S., CHEN, J., SHAO, L., HE, H., YANG, Y., YANG, S. & JIANG, W. (2009). Ammonium acetate enhances solvent production by *Clostridium acetobutylicum* EA 2018 using cassava as a fermentation medium. *Journal of Industrial Microbiology & Biotechnology* **36,** 1225-32.

HAAS, L. O., CREGG, J. M. & GLEESON, M. A. (1990). Development of an integrative DNA transformation system for the yeast Candida tropicalis. *Journal of Bacteriology* **172,** 4571-7.

HARRIS, L. M., DESAI, R. P., WELKER, N. E. & PAPOUTSAKIS, E. T. (2000). Characterization of recombinant strains of the *Clostridium acetobutylicum* butyrate kinase inactivation mutant: need for new phenomenological models for solventogenesis and butanol inhibition? *Biotechnology and Bioengineering* **67,** 1-11.

HARRIS, L. M., WELKER, N. E. & PAPOUTSAKIS, E. T. (2002). Northern, morphological, and fermentation analysis of spo0A inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. *Journal of Bacteriology* **184,** 3586-97.

HARTMANIS, M. G. N., AHLMAN, H. & GATENBECK, S. (1986). Stability of Solvent Formation in *Clostridium-Acetobutylicum* during Repeated Subculturing. *Applied Microbiology and Biotechnology* **23,** 369-371.

HARTMANIS, M. G. N. & GATENBECK, S. (1984). Intermediary Metabolism in *Clostridium-Acetobutylicum* - Levels of Enzymes Involved in the Formation of Acetate and Butyrate. *Applied and Environmental Microbiology* **47,** 1277-1283.

HATHEWAY, C. L. (1995). Botulism: the present status of the disease. *Current Topics in Microbiology and Immunology* **195,** 55-75.

HEAP, J. T., CARTMAN, S. T., KUEHNE, S. A., COOKSLEY, C. & MINTON, N. P. (2010a). ClosTron-targeted mutagenesis. *Methods in Molecular Biology* **646,** 165-82.

HEAP, J. T., EHSAAN, M., COOKSLEY, C. M., NG, Y. K., CARTMAN, S. T., WINZER, K. & MINTON, N. P. (2012). Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Research* **40,** e59.

HEAP, J. T., KUEHNE, S. A., EHSAAN, M., CARTMAN, S. T., COOKSLEY, C. M., SCOTT, J. C. & MINTON, N. P. (2010b). The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. *Journal of Microbiological Methods* **80,** 49-55.

HEAP, J. T., PENNINGTON, O. J., CARTMAN, S. T., CARTER, G. P. & MINTON, N. P. (2007). The ClosTron: a universal gene knock-out system for the genus *Clostridium*. *Journal of Microbiological Methods* **70,** 452-64.

HEAP, J. T., PENNINGTON, O. J., CARTMAN, S. T. & MINTON, N. P. (2009). A modular system for *Clostridium* shuttle plasmids. *Journal of Microbiological Methods* **78,** 79-85.

HEIDELBERGER, C., DANENBERG, P. V. & MORAN, R. G. (1983). Fluorinated Pyrimidines and Their Nucleosides. *Advances in Enzymology and Related Areas of Molecular Biology* **54,** 57-119.

HORTON, R. M., CAI, Z. L., HO, S. N. & PEASE, L. R. (1990). Gene-Splicing by Overlap Extension - Tailor-Made Genes Using the Polymerase Chain-Reaction. *Biotechniques* **8,** 528-&.

HORTON, R. M., HUNT, H. D., HO, S. N., PULLEN, J. K. & PEASE, L. R. (1989). Engineering Hybrid Genes without the Use of Restriction Enzymes - Gene-Splicing by Overlap Extension. *Gene* **77,** 61-68.

HUANG, W. C., RAMEY, D. E. & YANG, S. T. (2004). Continuous production of butanol by *Clostridium acetobutylicum* immobilized in a fibrous bed bioreactor. *Applied Biochemistry and Biotechnology* **113-116,** 887-98.

HUSSON, R. N., JAMES, B. E. & YOUNG, R. A. (1990). Gene replacement and expression of foreign DNA in mycobacteria. *Journal of Bacteriology* **172,** 519-24.

IBRAHIM, A. S. S. & EL-DIWANY, A. I. (2007). Isolation and identification of new cellulases producing thermophilic bacteria from an egyptian hot spring and

some properties of the crude enzyme. In *Australian Journal and Basic Applied Science.*

ITAYA, M., FUJITA, K., KUROKI, A. & TSUGE, K. (2008). Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nature Methods* **5,** 41-3.

ITAYA, M. & TANAKA, T. (1990). Gene-directed mutagenesis on the chromosome of *Bacillus subtilis* 168. *Molecular and General Genetics* **223,** 268-72.

JOHNSON, J. L. & CHEN, J. S. (1995). Taxonomic Relationships among Strains of *Clostridium-Acetobutylicum* and Other Phenotypically Similar Organisms. *Fems Microbiology Reviews* **17,** 233-240.

JOHNSON, J. L., TOTH, J., SANTIWATANAKUL, S. & CHEN, J. S. (1997). Cultures of "*Clostridium acetobutylicum*" from various collections comprise *Clostridium acetobutylicum, Clostridium beijerinckii,* and two other distinct types based on DNA-DNA reassociation. *International Journal of Systematic Bacteriology* **47,** 420-424.

JONES, D. T. & KEIS, S. (1995). Origins and Relationships of Industrial Solvent-Producing Clostridial Strains. *Fems Microbiology Reviews* **17,** 223-232.

JONES, D. T., SHIRLEY, M., WU, X. & KEIS, S. (2000). Bacteriophage infections in the industrial acetone butanol (AB) fermentation process. *Journal of Molecular Microbiology and Biotechnology* **2,** 21-6.

JONES, D. T. & WOODS, D. R. (1986). Acetone-butanol fermentation revisited. *Microbiological Reviews* **50,** 484-524.

JONES, E. W. & FINK, G. R. (1982). "Regulation of amino acid and nucleotide biosynthesis in yeast." Pp.181-299 in The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression, edited by Strathern JN, Jones EW and Broach JR.

JONES, S. W., PAREDES, C. J., TRACY, B., CHENG, N., SILLERS, R., SENGER, R. S. & PAPOUTSAKIS, E. T. (2008). The transcriptional program underlying the physiology of clostridial sporulation. *Genome Biology* **9,** R114.

JONES, S. W., TRACY, B. P., GAIDA, S. M. & PAPOUTSAKIS, E. T. (2011). Inactivation of sigmaF in *Clostridium acetobutylicum* ATCC 824 blocks sporulation prior to asymmetric division and abolishes sigmaE and sigmaG protein expression but does not block solvent formation. *Journal of Bacteriology* **193,** 2429-40.

KARBERG, M., GUO, H. T., ZHONG, J., COON, R., PERUTKA, J. & LAMBOWITZ, A. M. (2001). Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nature Biotechnology* **19,** 1162-1167.

KASHKET, E. R. & CAO, Z. Y. (1993). Isolation of a Degeneration-Resistant Mutant of *Clostridium-Acetobutylicum* Ncimb-8052. *Applied and Environmental Microbiology* **59,** 4198-4202.

KASHKET, E. R. & CAO, Z. Y. (1995). Clostridial Strain Degeneration. *Fems Microbiology Reviews* **17,** 307-315.

KEIS, S., BENNETT, C. F., WARD, V. K. & JONES, D. T. (1995). Taxonomy and Phylogeny of Industrial Solvent-Producing Clostridia. *International Journal of Systematic Bacteriology* **45,** 693-705.

KELLY, J. R., RUBIN, A. J., DAVIS, J. H., AJO-FRANKLIN, C. M., CUMBERS, J., CZAR, M. J., DE MORA, K., GLIEBERMAN, A. L., MONIE, D. D. & ENDY, D. (2009). Measuring the activity of BioBrick promoters using an in vivo reference standard. *Journal of Biological Engineering* **3,** 4.

KILLEFFER, D. H. (1927). Butanol and acetone from corn-A description of the fermentation process. *Industrial and Engineering Chemistry* **19,** 46-50.

KNIGHT, T. (2003). Idempotent vector design for standard assembly of biobricks. DTIC Document.

KNIPFER, N., SETH, A. & SHRADER, T. E. (1997). Unmarked gene integration into the chromosome of Mycobacterium smegmatis via precise replacement of the pyrF gene. *Plasmid* **37,** 129-40.

KUEHNE, S. A., HEAP, J. T., COOKSLEY, C. M., CARTMAN, S. T. & MINTON, N. P. (2011). ClosTron-mediated engineering of *Clostridium*. *Methods in Molecular Biology* **765,** 389-407.

KUTZENOK, A. & ASCHNER, M. (1952). Degenerative processes in a strain of *Clostridium* butylicum. *Journal of Bacteriology* **64,** 829-36.

KWOK, R. (2010). Five hard truths for synthetic biology. *Nature* **463,** 288-90.

LAMBOWITZ, A. M. & ZIRNMERLY, S. (2004). Mobile group II introns. *Annual Review of Genetics* **38,** 1-35.

LAMED, R., SETTER, E. & BAYER, E. A. (1983). Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *Journal of Bacteriology* **156,** 828-36.

LANZA, A. M., CROOK, N. C. & ALPER, H. S. (2012). Innovation at the intersection of synthetic and systems biology. *Current Opinion in Biotechnology* **23,** 712-717.

LEE, J., JANG, Y.-S., CHOI, S. J., IM, J. A., SONG, H., CHO, J. H., PAPOUTSAKIS, E. T., BENNETT, G. N. & LEE, S. Y. (2012). Metabolic engineering of *Clostridium acetobutylicum* ATCC 824 for isopropanol-butanol-ethanol fermentation. *Applied and environmental microbiology* **78,** 1416-1423.

LEE, S. F., FORSBERG, C. W. & GIBBINS, L. N. (1985). Cellulolytic Activity of *Clostridium acetobutylicum*. *Applied and environmental microbiology* **50,** 220-8.

LEE, S. K., CHOU, H., HAM, T. S., LEE, T. S. & KEASLING, J. D. (2008a). Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current Opinion in Biotechnology* **19,** 556-63.

LEE, S. Y., MERMELSTEIN, L. D., BENNETT, G. N. & PAPOUTSAKIS, E. T. (1992). Vector construction, transformation, and gene amplification in *Clostridium acetobutylicum* ATCC 824. *Annals of the New York Academy of Sciences* **665,** 39-51.

LEE, S. Y., PARK, J. H., JANG, S. H., NIELSEN, L. K., KIM, J. & JUNG, K. S. (2008b). Fermentative butanol production by Clostridia. *Biotechnology and Bioengineering* **101,** 209-28.

LEE, Y. J., ROMANEK, C. S. & WIEGEL, J. (2007). *Clostridium* aciditolerans sp nov., an acid-tolerant spore-forming anaerobic bacterium from constructed wetland sediment. *International Journal of Systematic and Evolutionary Microbiology* **57,** 311-315.

LEWIS, T. A., GOSZCZYNSKI, S., CRAWFORD, R. L., KORUS, R. A. & ADMASSU, W. (1996). Products of anaerobic 2,4,6-trinitrotoluene (TNT) transformation by *Clostridium* bifermentans. *Applied and Environmental Microbiology* **62,** 4669-4674.

LI, J., ZHAO, J. B., ZHAO, M., YANG, Y. L., JIANG, W. H. & YANG, S. (2010). Screening and characterization of butanol-tolerant micro-organisms. *Letters in Applied Microbiology* **50,** 373-9.

LIN, Y. L. & BLASCHEK, H. P. (1983). Butanol Production by a Butanol-Tolerant Strain of *Clostridium acetobutylicum* in Extruded Corn Broth. *Applied and Environmental Microbiology* **45,** 966-73.

LINK, A. J., PHILLIPS, D. & CHURCH, G. M. (1997). Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. *Journal of Bacteriology* **179,** 6228-37.

LIYANAGE, H., KASHKET, S., YOUNG, M. & KASHKET, E. R. (2001). *Clostridium beijerinckii* and *Clostridium difficile* detoxify methylglyoxal by a novel mechanism involving glycerol dehydrogenase. *Applied and environmental microbiology* **67,** 2004-10.

LONGLEY, D. B., HARKIN, D. P. & JOHNSTON, P. G. (2003). 5-Fluorouracil: Mechanisms of action and clinical strategies. *Nature Reviews Cancer* **3,** 330-338.

LOPEZ-CONTRERAS, A. M., CLAASSEN, P. A., MOOIBROEK, H. & DE VOS, W. M. (2000). Utilisation of saccharides in extruded domestic organic waste by *Clostridium acetobutylicum* ATCC 824 for production of acetone, butanol and ethanol. *Applied Microbiology and Biotechnology* **54,** 162-7.

LUTKE-EVERSLOH, T. & BAHL, H. (2011). Metabolic engineering of *Clostridium acetobutylicum*: recent advances to improve butanol production. *Current Opinion in Biotechnology* **22,** 634-47.

MADARRO, A., PENA, J. L., LEQUERICA, J. L., VALLES, S., GAY, R. & FLORS, A. (1991). Partial-Purification and Characterization of the Cellulases from *Clostridium-Cellulolyticum* H10. *Journal of Chemical Technology and Biotechnology* **52,** 393-406.

MARX, C. J. (2008). Development of a broad-host-range sacB-based vector for unmarked allelic exchange. *BioMed Central Research Notes* **1,** 1.

MCCOY, E., FRED, E. B., PETERSON, W. H. & HASTINGS, E. G. (1926). A cultural study of the acetone butyl alcohol organism-One plate. *Journal of Infectious Diseases* **39,** 457-483.

MCCOY, E., FRED, E. B., PETERSON, W. H. & HASTINGS, E. G. (1930). A cultural study of certain anaerobic, butyric-acid-forming bacteria. *Journal of Infectious Diseases* **46,** 118-137.

MERMELSTEIN, L. D. & PAPOUTSAKIS, E. T. (1993). Invivo 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. *Applied and Environmental Microbiology* **59,** 1077-1081.

MERMELSTEIN, L. D., WELKER, N. E., BENNETT, G. N. & PAPOUTSAKIS, E. T. (1992). Expression of cloned homologous fermentative genes in *Clostridium acetobutylicum* ATCC 824. *Biotechnology (N Y)* **10,** 190-5.

MERRITT, J., TSANG, P., ZHENG, L., SHI, W. & QI, F. (2007). Construction of a counterselection-based in-frame deletion system for genetic studies of Streptococcus mutans. *Oral Microbiology and Immunology* **22,** 95-102.

MERRYMAN, C. & GIBSON, D. G. (2012). Methods and applications for assembling large DNA constructs. *Metabolic Engineering* **14,** 196-204.

MIHELC, V. A., DUNCAN, C. L., AND & H., C. G. (1978). Characterization of a Bacteriocinogenic Plasmid in *Clostridium perfringens* CW55. *Antimicrobial Agents Chemotherapy* **14,** 771–779.

MINTON, N., BREHM JK, SWINFIELD T-J, WHELAN SM, MAUCHLINE ML, BODSWORTH N & JD, O. (1993). Clostridial cloning vectors. In: Woods DR, editor. The clostridia and biotechnology. Stoneham: Butterworth-Heinemann., 119-150.

MINTON, N. & MORRIS, J. G. (1981). Isolation and Partial Characterization of Three Cryptic Plasmids from Strains of *Clostridium butyricum*. *Journal of General Microbiology* **127,** 325-331.

MINTON, N. P. (2003). Clostridia in cancer therapy. *Nature Reviews Microbiology* **1,** 237-42.

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,** 559-73.

MONOD, M., DENOYA, C. & DUBNAU, D. (1986). Sequence and properties of pIM13, a macrolide-lincosamide-streptogramin B resistance plasmid from *Bacillus subtilis*. *Journal of Bacteriology* **167,** 138-47.

MONTOYA, D., AREVALO, C., GONZALES, S., ARISTIZABAL, F. & SCHWARZ, W. H. (2001). New solvent-producing *Clostridium* sp. strains, hydrolyzing a wide range of polysaccharides, are closely related to *Clostridium butyricum*. *Journal of Industrial Microbiology and Biotechnology* **27,** 329-35.

MONTOYA, D., SPITIA, S., SILVA, E. & SCHWARZ, W. H. (2000). Isolation of mesophilic solvent-producing clostridia from Colombian sources: physiological characterization, solvent production and polysaccharide hydrolysis. *Journal of Biotechnology* **79,** 117-126.

NAIR, R. V., GREEN, E. M., WATSON, D. E., BENNETT, G. N. & PAPOUTSAKIS, E. T. (1999). Regulation of the sol locus genes for butanol and acetone formation in *Clostridium acetobutylicum* ATCC 824 by a putative transcriptional repressor. *Journal of Bacteriology* **181,** 319-30.

NAKAMURA, C. E. & WHITED, G. M. (2003). Metabolic engineering for the microbial production of 1,3-propanediol. *Current Opinion in Biotechnology* **14,** 454-9.

NAKAYAMA, S., IRIE, R., KOSAKA, T., MATSUURA, K., YOSHINO, S. & FURUKAWA, K. (2007). New host-vector system in solvent-producing *Clostridium saccharoperbutylacetonicum* strain N1-4. *Journal of General and Applied Microbiology* **53,** 53-56.

NAKAYAMA, S., KIYOSHI, K., KADOKURA, T. & NAKAZATO, A. (2011). Butanol production from crystalline cellulose by cocultured *Clostridium thermocellum* and *Clostridium saccharoperbutylacetonicum* N1-4. *Applied and Environmental Microbiology* **77,** 6470-5.

NAPOLI, F., OLIVIERI, G., RUSSO, M. E., MARZOCCHELLA, A. & SALATINO, P. (2010). Butanol production by *Clostridium acetobutylicum* in a continuous packed bed reactor. *Journal* of *Industrial Microbiology* and *Biotechnology* **37,** 603-8.

NARIYA, H., MIYATA, S., SUZUKI, M., TAMAI, E. & OKABE, A. (2011). Development and application of a method for counterselectable in-frame deletion in *Clostridium perfringens*. *Applied and Environmental Microbiology* **77,** 1375-82.

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,** e56051.

NIKAWA, J. & KAWABATA, M. (1998). PCR- and ligation-mediated synthesis of marker cassettes with long flanking homology regions for gene disruption in Saccharomyces cerevisiae. *Nucleic Acids Research* **26,** 860-1.

NOLLING, J., BRETON, G., OMELCHENKO, M. V., MAKAROVA, K. S., ZENG, Q. D., GIBSON, R., LEE, H. M., DUBOIS, J., QIU, D. Y., HITTI, J., WOLF, Y. I., TATUSOV, R. L., SABATHE, F., DOUCETTE-STAMM, L., SOUCAILLE, P., DALY,

M. J., BENNETT, G. N., KOONIN, E. V., SMITH, D. R. & FINISHING, G. S. C. P. (2001). Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *Journal of Bacteriology* **183,** 4823-4838.

OBRIEN, R. W. & MORRIS, J. G. (1971). Oxygen and Growth and Metabolism of *Clostridium-acetobutylicum*. *Journal of General Microbiology* **68,** 307-&.

OUNINE, K., PETITDEMANGE, H., RAVAL, G. & GAY, R. (1983). Acetone-Butanol Production from Pentoses by *Clostridium acetobutylicum*. *Biotechnology Letters* **5,** 605-610.

OZKAN, M., DESAI, S. G., ZHANG, Y., STEVENSON, D. M., BEANE, J., WHITE, E. A., GUERINOT, M. L. & LYND, L. R. (2001). Characterization of 13 newly isolated strains of anaerobic, cellulolytic, thermophilic bacteria. *Journal of Industrial Microbiology and Biotechnology* **27,** 275-80.

PAPOUTSAKIS, E. T. (2008). Engineering solventogenic clostridia. *Current Opinion in Biotechnology* **19,** 420-9.

PAREDES, C. J., ALSAKER, K. V. & PAPOUTSAKIS, E. T. (2005). A comparative genomic view of clostridial sporulation and physiology. *Nature Reviews Microbiology* **3,** 969-78.

PAREKH, S., VINCI, V. A. & STROBEL, R. J. (2000). Improvement of microbial strains and fermentation processes. *Applied Microbiology and Biotechnology* **54,** 287-301.

PERUTKA, J., WANG, W. J., GOERLITZ, D. & LAMBOWITZ, A. M. (2004). Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *Journal of Molecular Biology* **336,** 421-439.

PLANTE, I. & COUSINEAU, B. (2006). Restriction for gene insertion within the Lactococcus lactis Ll.LtrB group II intron. *RNA-A Publication of the RNA Society* **12,** 1980-1992.

POPOFF, M. R. (1984). Selective medium for isolation of *Clostridium butyricum* from human feces. *Journal of Clinical Microbiology* **20,** 417-20.

PURDY, D., O'KEEFFE, T. A. T., ELMORE, M., HERBERT, M., MCLEOD, A., BOKORI-BROWN, M., OSTROWSKI, A. & MINTON, N. P. (2002). Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Molecular microbiology* **46,** 439-52.

QURESHI, N., LI, X. L., HUGHES, S., SAHA, B. C. & COTTA, M. A. (2006). Butanol production from corn fiber xylan using *Clostridium acetobutylicum*. *Biotechnology Progress* **22,** 673-680.

RABINOWITZ, J. (1972). Preparation and properties of clostridial ferredoxins. *Methods in Enzymology* **24,** 431-46.

RAVAGNANI, A., JENNERT, K. C., STEINER, E., GRUNBERG, R., JEFFERIES, J. R., WILKINSON, S. R., YOUNG, D. I., TIDSWELL, E. C., BROWN, D. P., YOUNGMAN, P., MORRIS, J. G. & YOUNG, M. (2000). Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. *Molecular Microbiology* **37,** 1172-85.

RIED, J. L. & COLLMER, A. (1987). An nptI-sacB-sacR cartridge for constructing directed, unmarked mutations in Gram-negative bacteria by marker exchange-eviction mutagenesis. *Gene* **57,** 239-246.

ROGERS, P. (1986). Genetics and Biochemistry of *Clostridium* Relevant to Development of Fermentation Processes. *Advances in Applied Microbiology* **31,** 1-60.

ROSGAARD, L., PEDERSEN, S., CHERRY, J. R., HARRIS, P. & MEYER, A. S. (2006). Efficiency of new fungal cellulase systems in boosting enzymatic degradation of barley straw lignocellulose. *Biotechnology Progress* **22,** 493-8.

ROTHERA, A. C. H. (1908). Note on the sodium nitro-prusside reaction for acetone. *Journal of Physiology-London* **37,** 491-494.

RUBNITZ, J. & SUBRAMANI, S. (1984). The minimum amount of homology required for homologous recombination in mammalian cells. *Molecular and Cellular Biology* **4,** 2253-8.

SABATHE, F., CROUX, C., CORNILLOT, E. & SOUCAILLE, P. (2002). amyP, a reporter gene to study strain degeneration in *Clostridium acetobutylicum* ATCC 824. *Fems Microbiology Letters* **210,** 93-98.

SAMBROOK, J., FRITSCH, E. F., AND & MANIATIS, T. (1989). Molecular Cloning: a Laboratory Manual, second edition. *Cold Spring Harbour, New York, Cold Spring Harbour Laboratory.*

SARKER, M. R., CARMAN, R. J. & MCCLANE, B. A. (1999). Inactivation of the gene (cpe) encoding *Clostridium perfringens* enterotoxin eliminates the ability of two cpe-positive C. *perfringens* type A human gastrointestinal disease isolates to affect rabbit ileal loops. *Molecular Microbiology* **33,** 946-58.

SATO, T., FUKUI, T., ATOMI, H. & IMANAKA, T. (2003). Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1. *Journal of Bacteriology* **185,** 210-20.

SEEDORF, H., FRICKE, W. F., VEITH, B., BRUGGEMANN, H., LIESEGANG, H., STRITTIMATTER, A., MIETHKE, M., BUCKEL, W., HINDERBERGER, J., LI, F. L., HAGEMEIER, C., THAUER, R. K. & GOTTSCHALK, G. (2008). The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. *Proceedings of the National Academy of Sciences of the USA* **105,** 2128-2133.

SHEEHAN, J. & HIMMEL, M. (1999). Enzymes, Energy, and the Environment: A Strategic Perspective on the U.S. Department of Energy's Research and Development Activities for Bioethanol. *Biotechnology Progress* **15,** 817-827.

SHETTY, R. P., ENDY, D. & KNIGHT, T. F., JR. (2008). Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering* **2,** 5.

SHUNMUGAM, V. (2009). Biofuels--Breaking the Myth of 'Indestructible Energy'? *Margin: The Journal of Applied Economic Research* **3,** 173-189.

SILLERS, R., CHOW, A., TRACY, B. & PAPOUTSAKIS, E. T. (2008). Metabolic engineering of the non-sporulating, non-solventogenic *Clostridium acetobutylicum* strain M5 to produce butanol without acetone demonstrate the robustness of the acid-formation pathways and the importance of the electron balance. *Metabolic Engineering* **10,** 321-32.

SMIT, S., WIDMANN, J. & KNIGHT, R. (2007). Evolutionary rates vary among rRNA structural elements. *Nucleic Acids Research* **35,** 3339-3354.

SOUCAILLE, P., FIGGE, R. & CROUX, C. (2006). Process for chromosomal integration and DNA sequence replacement in Clostridia. *Depot PCT n° PCT/EP2006/066997.*

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,** 641-54.

SWINFIELD, T. J., OULTRAM, J. D., THOMPSON, D. E., BREHM, J. K. & MINTON, N. P. (1990). Physical characterisation of the replication region of the Streptococcus faecalis plasmid pAM beta 1. *Gene* **87,** 79-90.

THEYS, J., PENNINGTON, O., DUBOIS, L., ANLEZARK, G., VAUGHAN, T., MENGESHA, A., LANDUYT, W., ANNE, J., BURKE, P. J., DURRE, P., WOUTERS, B. G., MINTON, N. P. & LAMBIN, P. (2006). Repeated cycles of *Clostridium*-directed enzyme prodrug therapy result in sustained antitumour effects in vivo. *British Journal of Cancer* **95,** 1212-9.

TOMAS, C. A., WELKER, N. E. & PAPOUTSAKIS, E. T. (2003). Overexpression of groESL in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. *Applied and Environmental Microbiology* **69,** 4951-65.

TRIPATHI, S. A., OLSON, D. G., ARGYROS, D. A., MILLER, B. B., BARRETT, T. F., MURPHY, D. M., MCCOOL, J. D., WARNER, A. K., RAJGARHIA, V. B., LYND, L. R., HOGSETT, D. A. & CAIAZZA, N. C. (2010). Development of pyrF-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a pta mutant. *Applied and Environmental Microbiology* **76,** 6591-9.

TUMMALA, S. B., WELKER, N. E. & PAPOUTSAKIS, E. T. (1999). Development and characterization of a gene expression reporter system for *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology* **65,** 3793-3799.

TUMMALA, S. B., WELKER, N. E. & PAPOUTSAKIS, E. T. (2003). Design of antisense RNA constructs for downregulation of the acetone formation pathway of *Clostridium acetobutylicum*. *Journal of Bacteriology* **185,** 1923-34.

UEKI, T., INOUYE, S. & INOUYE, M. (1996). Positive-negative KG cassettes for construction of multi-gene deletions using a single drug marker. *Gene* **183,** 153-157.

VAN DER GEIZE, R., DE JONG, W., HESSELS, G. I., GROMMEN, A. W., JACOBS, A. A. & DIJKHUIZEN, L. (2008). A novel method to generate unmarked gene deletions in the intracellular pathogen Rhodococcus equi using 5-fluorocytosine conditional lethality. *Nucleic Acids Research* **36,** e151.

VAREL, V. H., TANNER, R. S. & WOESE, C. R. (1995). *Clostridium* herbivorans sp. nov., a cellulolytic anaerobe from the pig intestine. *International journal of systematic bacteriology* **45,** 490-4.

VROOM, J. A. & WANG, G. L. (2008). Modular construction of plasmids through ligation-free assembly of vector components with oligonucleotide linkers. *Biotechniques* **44,** 924-926.

WACH, A. (1996). PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in S. cerevisiae. *Yeast* **12,** 259-265.

WANG, S., ZHAO, Y., LEIBY, M. & ZHU, J. (2009). A new positive/negative selection scheme for precise BAC recombineering. *Molecular Biotechnology* **42,** 110-6.

WANG, Y., LI, X., MAO, Y. & BLASCHEK, H. P. (2012). Genome-wide dynamic transcriptional profiling in *Clostridium beijerinckii* NCIMB 8052 using single-nucleotide resolution RNA-Seq. *BioMed Central Genomics* **13,** 102.

WARNICK, T. A., METHE, B. A. & LESCHINE, S. B. (2002). *Clostridium phytofermentans* sp nov., a cellulolytic mesophile from forest soil. *International Journal of Systematic and Evolutionary Microbiology* **52,** 1155-1160.

WEEKS, A. M. & CHANG, M. C. (2011). Constructing de novo biosynthetic pathways for chemical synthesis inside living cells. *Biochemistry* **50,** 5404-18.

WEISGAL, M. W. & CARMICHAEL, J. (1963). Chaim Weizmann; a biography by several hands, 1st American edition. Atheneum, New York,.

WEYER, E. R. & RETTGER, L. F. (1927). A comparative study of six different strains of the organism commonly concerned in large-scale production of butyl alcohol and acetone by the biological process. *Journal of Bacteriology* **14,** 399-424.

WILKINSON, S. R. & YOUNG, M. (1994). Targeted Integration of Genes into the *Clostridium-Acetobutylicum* Chromosome. *Microbiology-UK* **140,** 89-95.

WOESE, C. R., KANDLER, O. & WHEELIS, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* **87,** 4576-9.

YAMAGISHI, A., TANIMOTO, T., SUZUKI, T. & OSHIMA, T. (1996). Pyrimidine biosynthesis genes (*pyrE* and *pyrF*) of an extreme thermophile, Thermus thermophilus. *Applied and Environmental Microbiology* **62,** 2191-2194.

YANO, T., SANDERS, C., CATALANO, J. & DALDAL, F. (2005). *sacB*-5-fluoroorotic acid-*pyrE*-based bidirectional selection for integration of unmarked Alleles into the chromosome of Rhodobacter capsulatus. *Applied and Environmental Microbiology* **71,** 3014-3024.

ZALDIVAR, J., NIELSEN, J. & OLSSON, L. (2001). Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Applied Microbiology and Biotechnology* **56,** 17-34.

ZHANG, X. Z., YAN, X., CUI, Z. L., HONG, Q. & LI, S. P. (2006). mazF, a novel counter-selectable marker for unmarked chromosomal manipulation in *Bacillus subtilis*. *Nucleic Acids Research* **34,** e71.

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,** 1656-1664.

ZVERLOV, V. V., BEREZINA, O., VELIKODVORSKAYA, G. A. & SCHWARZ, W. H. (2006). Bacterial acetone and butanol production by industrial fermentation in the Soviet Union: use of hydrolyzed agricultural waste for biorefinery. *Applied Microbiology and Biotechnology* **71,** 587-97.

Appendix

Genbank accession numbers of ACE vectors

pMTL-JH12 HQ875753 *C. acetobutylicum pyrE*

pMTL-JH14 HQ875755 *C. acetobutylicum pyrE*

pMTL-JH15 HQ875756 *C. acetobutylicum thl*

pMTL-JH16 HQ875757 *C. acetobutylicum thl*

pMTL-JH30 HQ875766 *C. acetobutylicum thl*

pMTL-JH31 HQ875767 *C. acetobutylicum thl*

Genbank accession number of pyrE KO vector for C. difficile

pMTL-YN3 JX465735

pyrE KO plasmid sequence for C. acetobutylicum

pMTL-ME3

ccccctgcaggataaaaaaattgtagataaatttttataaaatagttttatctacaatttttttatcaggaaacagctatgac
cgcggccgctgtatccatatgaccatgattacgaattcgagctcggtacccggggatcctctagagtcgacgtca
cgcgtccatggagatctcgaggcctgcagacatgcaagcttggcactggccgtcgttttacaacgtcgtgactgg
gaaaaccctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagag
gcccgcaccgatcgcccttcccaacagttgcgcagcctgaatggcgaatggcgctagcataaaaataagaagc
ctgcatttgcaggcttcttatttttatggcgcgccgcattcacttcttttctatataaatatgagcgaagcgaataagcg
tcggaaaagcagcaaaaagtttccttttgctgttggagcatgggggttcaggggtgcagtatctgacgtcaatg
ccgagcgaaagcgagccgaagggtagcatttacgttagataaccccctgatatgctccgacgctttatatagaaa
agaagattcaactaggtaaaatcttaatataggttgagatgataaggtttataaggaatttgtttgttctaatttttcact
cattttgttctaatttcttttaacaaatgttcttttttttttagaacagttatgatatagttagaatagtttaaaataaggagtg
agaaaaagatgaaagaaagatatggaacagtctataaaggctctcagaggctcatagacgaagaaagtggaga
agtcatagaggtagacaagttataccgtaaacaaacgtctggtaacttcgtaaaggcatatatagtgcaattaataa
gtatgttagatatgattggcggaaaaaaacttaaaatcgttaactatatcctagataatgtccacttaagtaacaatac
aatgatagctacaacaagagaaatagcaaaagctacaggaacaagtctacaaacagtaataacaacacttaaaat
cttagaagaaggaaatattataaaaagaaaaactggagtattaatgttaaaccctgaactactaatgagaggcgac
gaccaaaaacaaaaatacctcttactcgaatttgggaactttgagcaagaggcaaatgaaatagattgacctccca
ataacaccacgtagttattgggaggtcaatctatgaaatgcgattaagggccggccagtgggcaagttgaaaaat
tcacaaaaatgtggtataatatctttgttcattagagcgataaacttgaatttgagagggaacttagatggtatttgaa
aaaattgataaaaatagttggaacagaaaagagtattttgaccactactttgcaagtgtaccttgtacctacagcatg
accgttaaagtggatatcacacaaataaaggaaaagggaatgaaactatatcctgcaatgctttattatattgcaat
gattgtaaaccgccattcagagtttaggacggcaatcaatcaagatggtgaattggggatatatgatgagatgata
ccaagctatacaatatttcacaatgatactgaaacattttccagcctttggactgagtgtaagtctgactttaaatcatt

tttagcagattatgaaagtgatacgcaacggtatggaaacaatcatagaatggaaggaaagccaaatgctccgga
aaacatttttaatgtatctatgataccgtggtcaaccttcgatggctttaatctgaatttgcagaaaggatatgattattt
gattcctatttttactatggggaaatattataaagaagataacaaaattatacttcctttggcaattcaagttcatcacg
cagtatgtgacggatttcacatttgccgttttgtaaacgaattgcaggaattgataaatagttaactaaggagaagat
ataaatgagtaatataaatgttatagatatattaaaagaatcaaatgcattattagaagggcatttcttactatcatctg
gaagacatagtaatagatattgtcaatgtgcgaaactacttcaatatccacaaaaagcagaaaaagttattagtgta
gtagcagaaaagcttaaagatgcagactttaatataatagtaggaccagctatgggtggtgtaatagtatcttatga
attagcaaggcaaactaatagaccaggcatatttgcagaaagaaaagatggaattatgtgtatacgaagaggattt
gaaataaaaaaaggagataaagttataatatcagaggatgtaataactacaggaaaatcttctttagaagtagctaa
agttatagaagaaatgggtggagaagtagtaggtatagcatgcatagttgatagaagagcggaaggtataaaaa
caaactacccaatatatagtgcttgtaaattagaaatagaaacttatgaaaaagataattgtgagttatgtaaaaaaa
atataccttttgtaaaaccaggtagtagagaacaaaaataagttaacttcaggtttgtctgtaactaaaaacaagtatt
taagcaaaaacatcgtagaaatacggtgtttttttgttaccctaagtttaaactccttttgataatctcatgaccaaaatc
ccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttct
gcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctacc
aactcttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttagg
ccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtg
gcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacg
gggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctat
gagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacagg
agagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgactt
gagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacg
gttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgc
ctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcgg
aagagcgcccaatacgcagggc

Published online 18 January 2012

Nucleic Acids Research, 2012, Vol. 40, No. 8 e59
doi:10.1093/nar/gkr1321

Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker

John T. Heap, Muhammad Ehsaan, Clare M. Cooksley, Yen-Kuan Ng, Stephen T. Cartman, Klaus Winzer and Nigel P. Minton*

Clostridia Research Group, BBSRC Sustainable BioEnergy Centre, School of Molecular Medical Sciences, Centre for Biomolecular Sciences, The University of Nottingham, University Park, Nottingham NG7 2RD, UK

Received January 18, 2011; Revised November 24, 2011; Accepted December 23, 2011

ABSTRACT

Most bacteria can only be transformed with circular plasmids, so robust DNA integration methods for these rely upon selection of single-crossover clones followed by counter-selection of double-crossover clones. To overcome the limited availability of heterologous counter-selection markers, here we explore novel DNA integration strategies that do not employ them, and instead exploit (i) activation or inactivation of genes leading to a selectable phenotype, and (ii) asymmetrical regions of homology to control the order of recombination events. We focus here on the industrial biofuel-producing bacterium *Clostridium acetobutylicum*, which previously lacked robust integration tools, but the approach we have developed is broadly applicable. Large sequences can be delivered in a series of steps, as we demonstrate by inserting the chromosome of phage lambda (minus a region apparently unstable in *Escherichia coli* in our cloning context) into the chromosome of *C. acetobutylicum* in three steps. This work should open the way to reliable integration of DNA including large synthetic constructs in diverse microorganisms.

INTRODUCTION

The addition of DNA conferring new or altered properties to microorganisms has underpinned biotechnology for decades. Recently, the potential scope and scale of this approach has grown with the acceleration of genome sequencing, development of commercial *de novo* DNA synthesis and advent of Synthetic Biology, the new discipline which brings engineering principles to the design and construction of biological systems (1).

DNA can be added to microorganisms using replicative plasmids, but these are inherently unstable, limiting their applied utility. To stabilize exogenous DNA, it must be irreversibly incorporated into a stable DNA molecule inside the cell, usually a chromosome. This can be accomplished in a one-step homologous recombination procedure (often called 'allele exchange' or 'gene replacement') for those organisms that are efficiently transformed with linear DNA, such as yeast and naturally competent bacteria like *Bacillus subtilis* (2). A selectable marker gene positioned alongside the DNA sequence of interest within an allele exchange cassette is only retained by the desired recombinant cells, allowing these cells to be specifically selected and easily isolated, typically using an antibiotic. Large or multiple sequences can be inserted at a single locus simply by alternating between two selectable markers in a series of integration steps, as in the 'domino' method of Itaya et al. (3).

Most bacteria cannot be transformed with linear DNA, so an integrative plasmid bearing the homologous recombination construct is used instead. As plasmids are circular, a single homologous recombination event can reversibly integrate the entire plasmid into the chromosome, resulting in unstable single-crossover cells with the potential to revert to wild-type. The desired stable double-crossover cells are much rarer, as they result from two homologous recombination events in a single cell or lineage. Double-crossover cells are not easily isolated from single-crossover cells, because both contain the selectable marker in the cassette. This issue can be overcome using a counter-selection marker located on the plasmid 'backbone', but identifying a suitable counter-selection marker and appropriate conditions for its use can be one of the most challenging aspects of developing genetic tools for a particular organism (4). We are interested in several bacterial species for which genetic tools are very limited, and the use of counter-selection markers has not been described.

*To whom correspondence should be addressed. Tel: +44 115 84 67458; Fax: +44 115 82 32120; Email: nigel.minton@nottingham.ac.uk
Correspondence may also be addressed to John T. Heap. Email: j.heap@imperial.ac.uk
Present address:
John T. Heap, Centre for Synthetic Biology and Innovation, Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, UK.

Here we develop generally applicable strategies for the selection of double-crossover clones at certain types of genomic loci without using a plasmid-borne counter-selection marker. Crucially, the expression of a selectable marker gene is coupled to the formation of the desired double-crossover recombinant chromosome, so that double-crossover clones can be isolated using the associated selectable phenotype. We call this approach allele-coupled exchange (ACE). The principles described here bring much of the power and practical simplicity of linear DNA procedures to integration of plasmid-borne DNA, and open the way to the step-by-step insertion of large DNA sequences into diverse microbial chromosomes, as we demonstrate using *Clostridium acetobutylicum*.

MATERIALS AND METHODS

Bacterial strains

Escherichia coli strains were grown in Luria-Bertani (LB) broth at 37°C with rotary shaking at 200 rpm, or on LB plates. All solid media were prepared by adding 1.5 % w/v agar to the corresponding broth. *E. coli* strains containing plasmids (Table 1) were cultured in LB broth supplemented with 12.5 μg/ml chloramphenicol, or on LB plates supplemented with 25 μg/ml chloramphenicol. *E. coli* strain TOP10 (Invitrogen) was used for plasmid cloning and storage, strain CA434 (5) was used as a conjugation donor, and TOP10 containing pAN2 (6) was used for *in vivo* methylation of plasmid DNA prior to transformation of *C. acetobutylicum*. *Clostridium* spp. were grown in static culture at 37°C under an anaerobic atmosphere of N_2:H_2:CO_2 (80:10:10, vol:vol:vol) in an anaerobic workstation (Don Whitley, UK) using media prereduced overnight under the same conditions. Plasmids were transferred into *Clostridium* spp. by electroporation or conjugation as described previously (7).

Plasmid construction

Integration vectors for use in *C. acetobutylicum* ATCC 824, *Clostridium sporogenes* NCIMB 10696 and *Clostridium difficile* 630 are based upon *Clostridium–E. coli* shuttle plasmids pMTL85141, pMTL85151 and pMTL83151, respectively (7), and differ to those parental plasmids only between the SbfI and AscI sites. The integration vectors and their key functional components are listed in Table 1, and annotated sequences are available from GenBank/EMBL/DDJB. Chromosomal DNA of phage lambda *c*I857*ind* 1 *Sam* 7 was obtained from NEB and fragments were cloned as shown in Figure 3. L12 was derived from L18 by a spontaneous deletion of 5996 bp corresponding to nucleotides 21738–27733 of the lambda *c*I857*ind* 1 *Sam* 7 chromosome.

Growth media for integration procedures in *Clostridium acetobutylicum*

After electroporation, transformants of *C. acetobutylicum* were selected on CGM (8) agar plates supplemented with 15 μg/ml thiamphenicol to select for the plasmid-borne resistance marker *catP* and 20 μg/ml uracil to allow *pyrE* mutants to grow (except for those experiments which did not involve *pyrE* mutants, where uracil was not required). CGM agar plates supplemented with 400 μg/ml 5-fluoroorotic acid (FOA) and 1 μg/ml uracil were used to select *pyrE*-minus clones. These conditions were validated using a previously constructed *pyrF* mutant (6) which should have the same phenotype as a *pyrE* mutant. CBM (9) agar plates (which do not contain uracil) were used to select clones able to synthesise their own uracil, which requires a functional *pyrE* gene. Erythromycin-resistant clones were selected on CGM agar plates supplemented with 40 μg/ml erythromycin.

PCR analysis of chromosomal insertions

Genomic DNA from *Clostridium* spp. was prepared using the QIAGEN DNeasy Blood and Tissue Kit in accordance with the manufacturer's instructions and recommended pretreatment for Gram-positive bacteria. PCRs were performed using Taq DNA polymerase (NEB) or KOD DNA Polymerase (Merck) or Phusion polymerase (NEB) in accordance with the manufacturer's instructions. Appropriate combinations of oligonucleotide primers are specified in the legends of relevant figures and in the Supplementary Data, and oligonucleotide sequences are listed in Supplementary Table S1.

Analysis of fermentation products

The fermentation products of wild-type *C. acetobutylicum* ATCC 824 were compared to the *adh*-expressing recombinant strain in batch culture using CBM broth containing 50 g/l glucose and 5 g/l $CaCO_3$. Starter cultures in the same medium were inoculated using fresh colonies, then when these reached an OD_{600} of ~0.5, they were used at a 1% inoculum to start the main cultures. Samples of 1 ml were removed, placed on ice, then centrifuged at 16 000*g* for 1.5 min. Supernatants were removed and stored at −20°C before analysis by gas chromatography. Ethanol, acetone, butanol, acetic acid and butyric acid were quantified using a Thermo Focus GC equipped with a 30 m TR-FFAP column (0.25 mm internal diameter) and a flame ionization detector. H_2 was used as the carrier gas at 0.8 ml/min. The flame was maintained by 35 ml/min H_2, 350 ml/min compressed air and 30 ml/min N_2. The injector and detector temperatures were 240°C and 270°C, respectively. Peaks were resolved using a column profile of 40°C for 2 min, followed by a 30°C/min ramp to 140°C, then a 45°C/min ramp to 210°C and finally 1 min at 210°C. Samples were extracted before injection by adding an equal volume of ethyl acetate (500 μl) to the supernatant sample, vortexing for 10 s and centrifuging for 5 min at 16 000*g*. The 300 μl organic phase was removed to a 2 ml sample vial containing a 300 μl deactivated glass insert. Samples of 1 μl were injected.

Table 1. List of integration vectors

Plasmid	Accession number	Organism	Locus	First region of homology	Second region of homology	Element(s) between regions of homology
pMTL-JH1	HQ875748	*Clostridium acetobutylicum*	*pyrF*	None (to be provided with insert)	300 bp internal fragment of *pyrF*	*lacZ* MCS
pMTL-JH2	HQ875749	*Clostridium acetobutylicum*	*pyrF*	1200 bp immediately following *pyrF*	300 bp internal fragment of *pyrF*	*lacZ* MCS
pMTL-JH3	HQ875750	*Clostridium acetobutylicum*	*pyrF*	None (to be provided with insert)	300 bp internal fragment of *pyrF*	3′ part of *pyrF*, *lacZ* MCS
pMTL-JH4	HQ875751	*Clostridium acetobutylicum*	*pyrF*	1200 bp immediately following *pyrF*	300 bp internal fragment of *pyrF*	3′ part of *pyrF*, *lacZ* MCS
pMTL-JH11	HQ875752	*Clostridium acetobutylicum*	*pyrE*	None (to be provided with insert)	300 bp internal fragment of *pyrE*	*lacZ* MCS
pMTL-JH12	HQ875753	*Clostridium acetobutylicum*	*pyrE*	1200 bp immediately following *pyrE*	300 bp internal fragment of *pyrE*	*lacZ* MCS
pMTL-JH13	HQ875754	*Clostridium acetobutylicum*	*pyrE*	None (to be provided with insert)	300 bp internal fragment of *pyrE*	3′ part of *pyrE*, *lacZ* MCS
pMTL-JH14	HQ875755	*Clostridium acetobutylicum*	*pyrE*	1200 bp immediately following *pyrE*	300 bp internal fragment of *pyrE*	3′ part of *pyrE*, *lacZ* MCS
pMTL-JH15	HQ875756	*Clostridium acetobutylicum*	*thl*	None (to be provided with insert)	Last 300 bp of *thl*	*ermB* lacking promoter, *lacZ* MCS
pMTL-JH16	HQ875757	*Clostridium acetobutylicum*	*thl*	1200 bp immediately following *thl*	Last 300 bp of *thl*	*ermB* lacking promoter, *lacZ* MCS
pMTL-JH17	HQ875758	*Clostridium difficile*	*pyrE*	None (to be provided with insert)	300 bp internal fragment of *pyrE*	*lacZ* MCS
pMTL-JH18	HQ875759	*Clostridium difficile*	*pyrE*	1200 bp immediately following *pyrE*	300 bp internal fragment of *pyrE*	*lacZ* MCS
pMTL-JH19	HQ875760	*Clostridium difficile*	*pyrE*	None (to be provided with insert)	300 bp internal fragment of *pyrE*	3′ part of *pyrE*, *lacZ* MCS
pMTL-JH20	HQ875761	*Clostridium difficile*	*pyrE*	1200 bp immediately following *pyrE*	300 bp internal fragment of *pyrE*	3′ part of *pyrE*, *lacZ* MCS
pMTL-JH26	HQ875762	*Clostridium sporogenes*	*pyrE*	None (to be provided with insert)	300 bp internal fragment of *pyrE*	*lacZ* MCS
pMTL-JH27	HQ875763	*Clostridium sporogenes*	*pyrE*	1200 bp immediately following *pyrE*	300 bp internal fragment of *pyrE*	*lacZ* MCS
pMTL-JH28	HQ875764	*Clostridium sporogenes*	*pyrE*	None (to be provided with insert)	300 bp internal fragment of *pyrE*	3′ part of *pyrE*, *lacZ* MCS
pMTL-JH29	HQ875765	*Clostridium sporogenes*	*pyrE*	1200 bp immediately following *pyrE*	300 bp internal fragment of *pyrE*	3′ part of *pyrE*, *lacZ* MCS
pMTL-JH30	HQ875766	*Clostridium acetobutylicum*	*thl*	None (to be provided with insert)	Last 300 bp of *thl*	*pyrE* lacking promoter, *lacZ* MCS
pMTL-JH31	HQ875767	*Clostridium acetobutylicum*	*thl*	1200 bp immediately following *pyrE*	Last 300 bp of *thl*	*pyrE* lacking promoter, *lacZ* MCS

The elements of the integration cassette of each plasmid are shown, and are best understood with reference to the text and Figures 1–3. Plasmids pMTL-JH12, 14, 15, 16, 30 and 31 are described in the main text, and pMTL-JH2, 18 and 27 are described in the Supplementary Data. Suitable combinations of plasmids targeting the same locus could be used in a series of steps to integrate several overlapping fragments of a large sequence, as described in the text and shown in Figure 3. MCS denotes a multiple cloning site.

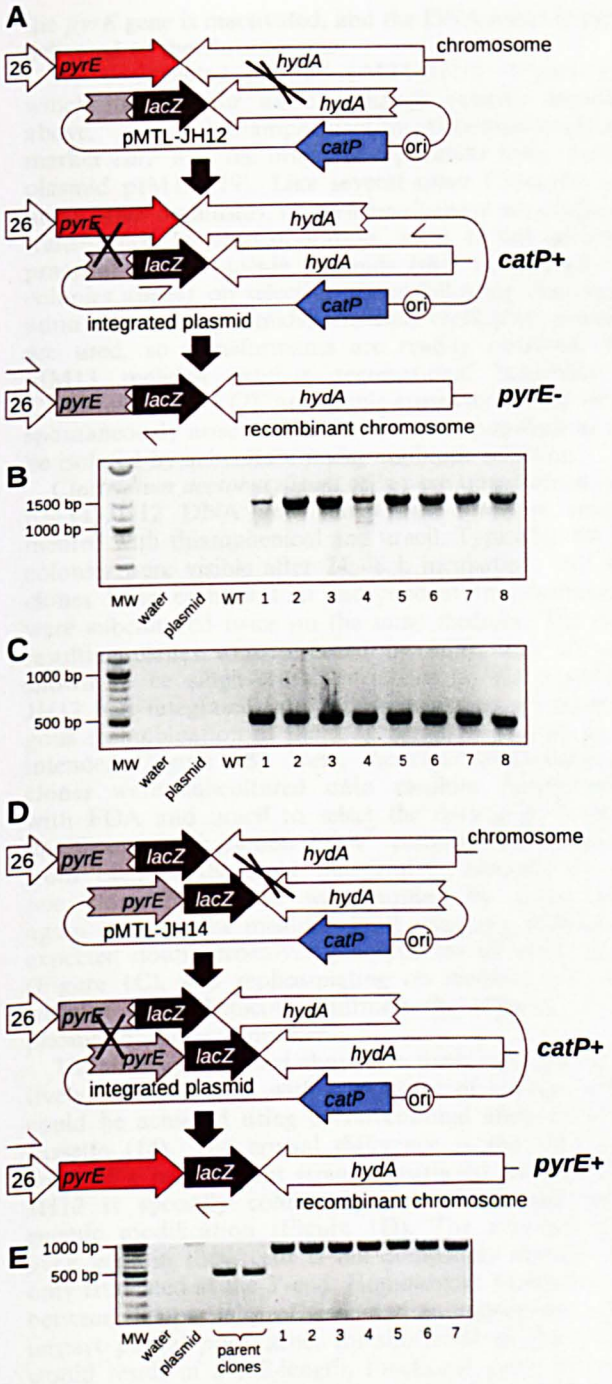


Figure 1. DNA integration at the *pyrE* locus of *C. acetobutylicum*. (**A**) Selection of stable double-crossover clones using pMTL-JH12. The first recombination event (plasmid integration) is mediated by the long region of homology between pMTL-JH12 and *hydA*. Single-crossover clones are obtained on medium containing thiamphenicol. The second recombination event (plasmid excision) is mediated by the short region of homology between pMTL-JH12 and an internal portion of *pyrE*. Double-crossover clones are selected using FOA. (**B**) PCR screening of eight candidate single-crossover clones using primers lacZa-sF2 and Cac-hydA-sR2, which anneal to the single-crossover chromosome where indicated in (A). MW, 2-Log DNA Ladder (NEB) molecular weight marker; plasmid, pMTL-JH12 plasmid DNA control; WT, wild-type *C. acetobutylicum* genomic DNA control; 1–8, candidate clones. All eight candidates show the expected 1428 bp band. (**C**) PCR screening of eight candidate double-crossover clones using primers CAC0026-sF2 and M13F which anneal to the double-crossover chromosome where indicated in (A). MW, 2-Log DNA Ladder (NEB) molecular weight marker; plasmid, pMTL-JH12 plasmid DNA control; WT, wild-type *C. acetobutylicum* genomic DNA control; 1–8, candidate clones. All eight candidates show the expected 558 bp band. (**D**) Selection of stable double-crossover clones using pMTL-JH14. The first recombination event (plasmid integration) is mediated by the long region of homology between pMTL-JH14 and *hydA*/*lacZ*. Single-crossover clones are obtained on medium containing thiamphenicol. The second recombination event (plasmid excision) is mediated by the short region of homology between pMTL-JH14 and the corresponding portion of *pyrE*. Double-crossover clones are selected using growth medium lacking uracil. (**E**) PCR screening of seven candidate double-crossover clones using primers CAC0026-sF2 and M13F which anneal to the double-crossover chromosome where indicated in (C). MW, 2-Log DNA Ladder (NEB) molecular weight marker; plasmid, pMTL-JH14 plasmid DNA control; 1–7, candidate clones. Controls using DNA from two of the *pyrE*-minus clones obtained in (A) are marked as parent clones. All seven candidates show the expected 861 bp band, and the parent clones show the expected 558 bp band. This increase of 303 bp corresponds to the restoration of the truncated *pyrE* gene to full length.

RESULTS

Specific selection of double-crossovers by switching the state of *pyrE*

Clostridium acetobutylicum is an industrial organism that naturally produces the excellent biofuel butanol, but the yield, specificity and feedstock utilization of the fermentation process might each be improved by the addition of DNA encoding heterologous enzymes (10). This motivated us to develop a robust DNA integration method for *C. acetobutylicum*. Our first approach requires a gene which is both positively and negatively selectable, such as *pyrE* or *pyrF*, which encode the pyrimidine biosynthesis enzymes orotate phosphoribosyltransferase and orotidine 5-phosphate decarboxylase, respectively. Cells which contain *pyrE* and *pyrF* can be selected on growth medium lacking uracil, as they are required for uracil biosynthesis. Conversely, cells that lack either *pyrE* or *pyrF* can be selected on growth medium supplemented with FOA, as this compound is highly toxic only to cells which contain this pathway. FOA has been used for counter selection for some time (11) and in various organisms (12–17) including very recently in *Clostridium thermocellum* (18).

To isolate double-crossover clones by exploiting a change in the state of *pyrE*, we designed a special allele exchange cassette (Figure 1A). The two regions of homology are of very different lengths, in an attempt to control the order of the recombination events. A long region of homology corresponding to the 1200 bp immediately downstream of *pyrE* is intended to direct the first recombination event so that a large majority of single-crossover clones are the result of recombination in this region, which would not inactivate *pyrE*. A subsequent recombination event at a second, much smaller region of homology corresponding to a 300 bp internal portion of *pyrE* would excise the plasmid, resulting in FOA-resistant (FOA^R) double-crossover cells in which

the *pyrE* gene is inactivated, and the DNA insert is stably delivered to the chromosome.

We constructed plasmid pMTL-JH12 (Figure 1A), which includes the allele exchange cassette described above, the chloramphenicol/thiamphenicol-resistance marker *catP* and the origin of replication from *Bacillus* plasmid pIM13 (19). Like several other Clostridia and many other organisms, *C. acetobutylicum* is not efficiently transformed by electroporation, so it is not generally practical to use suicide plasmids (that is, typically no colonies appear on selective plates following electroporation of suicide plasmids). Instead, replicative plasmids are used, so transformants are readily obtained. The pIM13 replicon exhibits segregational instability in *C. acetobutylicum* (7), and single-crossover clones which spontaneously arise within transformant populations can be isolated by subculture under antibiotic selection.

Clostridium acetobutylicum cells were transformed with pMTL-JH12 DNA and plated onto medium supplemented with thiamphenicol and uracil. Typically, 10–100 colonies were visible after 24–48 h incubation, and two clones from each of four independent transformations were subcultured twice on the same medium. The eight resulting clones were screened by PCR, and all were shown to be single-crossover clones in which pMTL-JH12 had integrated into the chromosome via homologous recombination at the long region of homology, as intended (Figure 1B). Next, the eight single-crossover clones were subcultured onto medium supplemented with FOA and uracil to select the desired *pyrE*-minus double-crossover clones. FOA^R colonies were obtained from each of the eight independent subcultures, and one clone from each was purified by subculturing again on the same medium. PCR screening showed the expected double-crossover genotype for all eight clones (Figure 1C), and replica-plating on medium with and without thiamphenicol confirmed the absence of the plasmid-borne *catP* marker.

The strategy described above effectively replaces a negatively selectable gene with a sequence of interest, which could be achieved using a conventional allele exchange cassette (14). The crucial difference is that the *pyrE* locus of a recombinant strain constructed using pMTL-JH12 is specially configured to facilitate subsequent genetic modification (Figure 1D). The non-functional *pyrE* gene in these cells is not completely deleted, it is only truncated at the 3′-end. Homologous recombination between this partial *pyrE* gene and an appropriate counterpart partial *pyrE* gene, foreshortened at the 5′-end, would result in a full-length, functional gene. We constructed plasmid pMTL-JH14, which is almost identical to pMTL-JH12, except the 300 bp internal portion of *pyrE* that comprises the small homology region is followed immediately by the remainder of the *pyrE* coding sequence (Figure 1D). As before, the long region of homology is intended to direct plasmid integration, without affecting the *pyrE* phenotype. Subsequent plasmid excision via the short 300-bp homology region will result in the desired double-crossover clones, specifically selectable by their *pyrE*-positive, uracil prototrophic phenotype. We transformed one of the previously constructed *pyrE* mutant clones of *C. acetobutylicum* with pMTL-JH14 DNA, selected transformants on plates supplemented with thiamphenicol and uracil and subcultured them twice on this medium. To select *pyrE*-positive clones, cells were subcultured onto medium lacking uracil. Seven independent clones were purified and shown to be the intended double-crossovers by PCR (Figure 2D) and thiamphenicol sensitivity.

It would be useful to apply the strategy above to *Clostridium* spp. other than *C. acetobutylicum*. We constructed plasmids equivalent to pMTL-JH12 and pMTL-JH14 for *C. sporogenes* NCIMB 10696, which shows potential in novel tumor therapies (20); and strain 630Δ*erm* of the important human pathogen *C. difficile* (21) (Table 1). Double-crossover clones of both these organisms could be selected using media supplemented with FOA, as described in the Supplementary Data and shown in Supplementary Figure S2. Integration at the *pyrF* locus of *C. acetobutylicum*, including insertions of fragments of phage lambda DNA of various sizes, is also described in the Supplementary Data and shown in Supplementary Figure S1.

Coupling expression of heterologous selectable markers to a chromosomal promoter

Using plasmid pMTL-JH14 in the specially configured *pyrE* mutant constructed using pMTL-JH12, we had effectively fused together two partial *pyrE* genes to result in one complete, functional *pyrE* gene via recombination in the middle of the coding sequence. We realized that a recombination event bringing together two partial genes to form one whole gene need not occur in the coding sequence. In an alternative arrangement, the full-length coding sequence of a selectable marker, lacking only a promoter, could be linked by recombination to a promoter situated on the chromosome, completing the gene and leading to its expression. We designed plasmid pMTL-JH31 (Figure 2A) to implement this modified concept by integrating a *pyrE* gene lacking its own promoter into the chromosome of the previously constructed *pyrE* mutant of *C. acetobutylicum*. The insertion was targeted to a site downstream of the thiolase (*thl*) promoter, which is known to exhibit strong expression throughout growth (22,23). To prevent possible homologous recombination at the *pyrE* locus, pMTL-JH31 uses the orthologous *pyrE* gene from *C. sporogenes* ATCC 15579, which is only 47.6% identical to the *C. acetobutylicum* gene. As before, a long (1200 bp) region of homology directs plasmid integration without altering the *pyrE* phenotype, and single-crossover clones are obtained on medium containing thiamphenicol. Subsequently, plasmid excision mediated by a second region of homology places the silent (non-expressed) *pyrE* gene immediately downstream of the *thl* gene on the chromosome, leading to its co-expression with *thl*, and allowing these double-crossover clones to be selected on medium lacking uracil (Figure 2A). To test this scheme, we transformed the *pyrE* mutant of *C. acetobutylicum* with pMTL-JH31 DNA, and performed the integration procedure as described above for pMTL-JH14.

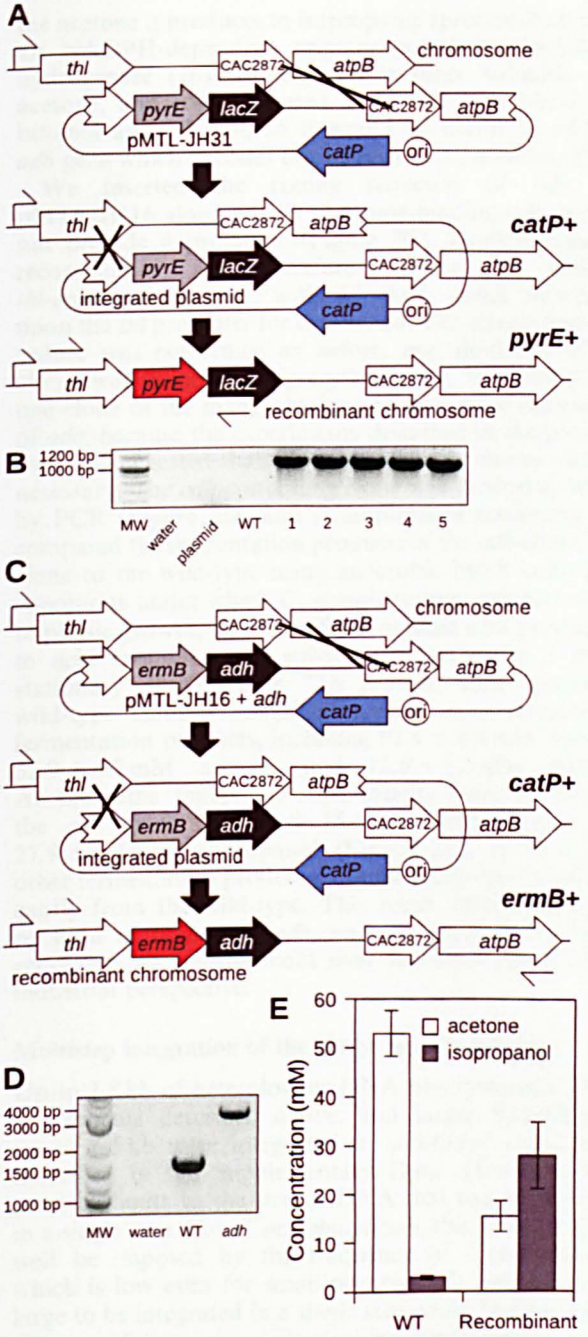


Figure 2. DNA integration at the *thl* locus of *C. acetobutylicum*. (**A**) Selection of stable double-crossover clones using pMTL-JH31. The first recombination event (plasmid integration) is mediated by the long region of homology between pMTL-JH31 and CAC2872/*atpB*. Single-crossover clones are obtained on medium containing thiamphenicol. The second recombination event (plasmid excision) is mediated by the short region of homology between pMTL-JH31 and the 3′-end of *thl*. Double-crossover clones are selected using growth medium lacking uracil. (**B**) PCR screening of five candidate double-crossover clones using primers Cac-thl-sF1 and M13F which anneal where indicated in (A). MW, 2-Log DNA Ladder (NEB) molecular weight marker; plasmid, pMTL-JH31 plasmid DNA control; WT, wild-type *C. acetobutylicum* genomic DNA control; 1–5, candidate clones. All five candidates show the expected 1101 bp band. (**C**) Selection of stable double-crossover clones using pMTL-JH16 containing *adh*. The first recombination event (plasmid integration) is mediated by the long region of homology between pMTL-JH16 and CAC2872/*atpB*. Single-crossover clones are obtained on medium containing thiamphenicol. The second recombination event (plasmid excision) is mediated by the short region of homology between pMTL-JH16 and the 3′-end of *thl*. Double-crossover clones are selected using erythromycin. (**D**) PCR screening of one candidate *adh*-expressing double-crossover clone using primers Cac-thl-sF1 and Cac-atpB-sR1 which anneal where indicated in (C). MW, 2-Log DNA Ladder (NEB) molecular weight marker; WT, wild-type *C. acetobutylicum* genomic DNA control which shows the expected 1660 bp band; *adh*, candidate clone which shows the expected 3523 bp band. (**E**) Concentrations of acetone and isopropanol in the supernatant of cultures of wild-type *C. acetobutylicum* (WT) or the *adh*-expressing recombinant clone after 72 h growth. The concentrations of both fermentation products differ significantly between the two strains ($P < 0.01$).

Five independent clones were screened by PCR (Figure 2B) and thiamphenicol sensitivity, and all were verified, confirming the utility of this modified approach.

The use of recombination to link a chromosomal promoter to a selectable marker within an allele exchange cassette represents a very flexible strategy, not limited to restoration of an auxotrophic mutant to prototrophy. We constructed pMTL-JH16, which is identical to pMTL-JH31, except the ribosome-binding site and coding sequence of *pyrE* are replaced by those of the macrolide–lincosamide–streptogramin (MLS) antibiotic-resistance marker *ermB* (24). As before a long (1200 bp) region of homology directs plasmid integration. Next, the recombination event mediating plasmid excision positions *ermB* immediately downstream of *thl*, leading to *ermB* expression and an MLS-resistant phenotype which can be used to specifically select double-crossover clones (Figure 2C). We tested pMTL-JH16 in wild-type *C. acetobutylicum* using a similar integration procedure as before, culturing transformants initially on thiamphenicol, and were readily able to specifically select double-crossover clones using the macrolide antibiotic erythromycin. Six recombinant clones were screened in the same way as before, and all were confirmed to be the desired recombinants (data not shown).

Integration and expression of an alcohol dehydrogenase

In the previous strategy, the expression of heterologous marker genes was coupled to the formation of a recombinant double-crossover chromosome, providing a means to select these clones. Next we sought to extend this principle to the expression of other heterologous genes of interest, for other purposes. Besides butanol, the other major products of sugar fermentation by *C. acetobutylicum* are acetone and ethanol, and the process is often called the acetone–butanol–ethanol (ABE) fermentation (25). Butanol is the most valuable product, but ethanol also has value both as a commodity chemical and as a biofuel. Acetone is less valuable and is not suitable as a fuel, so it represents an undesirable by-product or waste product of the industrial ABE process. *Clostridium beijerinckii* NRRL B593 has a similar physiology to *C. acetobutylicum*, but reduces

the acetone it produces to isopropanol (propan-2-ol) using an NADPH-dependent primary/secondary alcohol dehydrogenase (26). Isopropanol is more valuable than acetone, and could be used as biofuel in a blend with butanol and ethanol, so it would be useful to add the *adh* gene which encodes this activity to *C. acetobutylicum*.

We inserted the coding sequence of *adh* into pMTL-JH16 along with a ribosome-binding site, but did not provide a promoter (Figure 2C). Double-crossover recombinants will therefore contain an artificial *thl-ermB-adh* operon, with all three genes dependent upon the *thl* promoter for expression. The integration procedure was performed as before, and double-crossover clones were selected using erythromycin. We picked only one clone of the many obtained to screen for integration of *adh*, because the experiments described in the previous sections suggested that screening multiple clones was unnecessary. The *adh*-containing clone was verified as before by PCR (Figure 2D) and thiamphenicol sensitivity. We compared the fermentation products of the *adh*-containing clone to the wild-type using anaerobic batch cultivation conditions under which *C. acetobutylicum* exhibits classic bi-phasic growth, switching from organic acid production to acid re-uptake and solvent production as it enters stationary phase. After 72 h growth, supernatants of wild-type cultures contained typical concentrations of fermentation products, including 91.9 ± 8.4 mM butanol, 52.9 ± 4.7 mM acetone and 12.6 ± 1.5 mM ethanol. At the same timepoint, supernatants from cultures of the *adh* clone contained 15.4 ± 3.1 mM acetone and 27.9 ± 6.7 mM isopropanol (Figure 2E). None of the other fermentation product concentrations differed significantly from the wild-type. This result demonstrates expression of functional *adh*, and the recombinant strain represents an improvement over the wild-type from an industrial perspective.

Multistep integration of the phage lambda genome

Up to 1.8 kb of heterologous DNA was integrated in the experiments described above, and larger fragments of up to 6.5 kb were integrated in additional experiments described in the Supplementary Data. However, there must be limits to the size of DNA that can be delivered in a single step. For *C. acetobutylicum*, this limitation may well be imposed by the frequency of transformation, which is low even for small plasmids (7). Sequences too large to be integrated in a single step could be delivered in a series of steps using overlapping subfragments of the desired sequence, as in the domino method for *Bacillus subtilis* (3). This might be achieved either by alternately switching on and off a positively and negatively selectable gene such as *pyrE* (Figure 1) or by alternately linking expression of two different heterologous selectable markers to a chromosomal promoter (Figure 2). In either scheme, the short region of homology would remain the same in every step, whereas the long region of homology would target the end of the previous insert in the second and subsequent steps.

To test multistep DNA delivery using ACE, we attempted to insert the entire genome of phage lambda into the chromosome of *C. acetobutylicum*. We identified three restriction fragments of circular phage lambda DNA (circularized by ligation of the cohesive ends of the linear chromosome) that covered the entire genome and provided regions of overlap suitable to direct the initial recombination events in the second and third ACE steps (Figure 3). We used restriction ligation cloning to insert the 28 kb fragment (L28), 18 kb fragment (L18) and 6.5 kb fragment (L6.5) into ACE vectors pMTL-JH16, pMTL-JH30 and pMTL-JH15, respectively. Vectors pMTL-JH30 and pMTL-JH15 are similar to pMTL-JH31 and pMTL-JH16, but do not include a long region of homology, so these vectors are appropriate when the long region of homology depends upon a previous step, and must therefore be provided with the insert or as part of the insert (Table 1). When we inserted L18 into pMTL-JH30, we obtained clones with deletions in the lambda sequence. It appears that cloning this region (which includes lysis genes) into pMTL-JH30 is toxic to *E. coli*, so clones with spontaneous deletions are selected. In one of these constructs, 6 kb of DNA is deleted, but the regions required for recombination between fragments are not affected. We used this 12 kb of lambda DNA (L12) instead of L18, and proceeded to the multistep integration.

We integrated each of the three lambda DNA fragments in turn into the chromosome of the *pyrE* mutant of *C. acetobutylicum* (Figure 3). Double-crossover clones were selected using erythromycin (*ermB*+) for the L28 insertion, then medium lacking uracil (*pyrE*+) for the L12 insertion, then erythromycin again for the L6.5 insertion. At each step, the insertion was initially verified by thiamphenicol sensitivity and PCR. After all three insertions were completed, a Southern blot of EcoRI-digested genomic DNA from the three recombinant strains (containing one, two or three lambda DNA insertions) was performed, using lambda DNA as the probe (Figure 3D). EcoRI-digested genomic DNA from the parental *pyrE* mutant strain was also included, as was EcoRI-digested plasmid DNA of the three integration plasmids. All samples showed the expected distinguishing pattern of fragments (none in the case of the *pyrE* strain control) confirming the successful multistep insertion. The insertions were also verified by sequencing.

DISCUSSION

In this work, we addressed the need to reliably and stably deliver DNA to organisms lacking mature genetic tools, which include bacterial species of applied importance. The specialized allele exchange approaches we have demonstrated share two common principles: (i) cells in which a particular homologous recombination event has occurred can be selected if the event interrupts a counter-selectable gene, or appropriately fuses together two parts of a positively-selectable gene; and (ii) regions of homology of very different lengths strongly bias the order of homologous recombination events, and can therefore be exploited to couple acquisition of a selectable phenotype to the second recombination event in a lineage—the formation of a double-crossover clone.

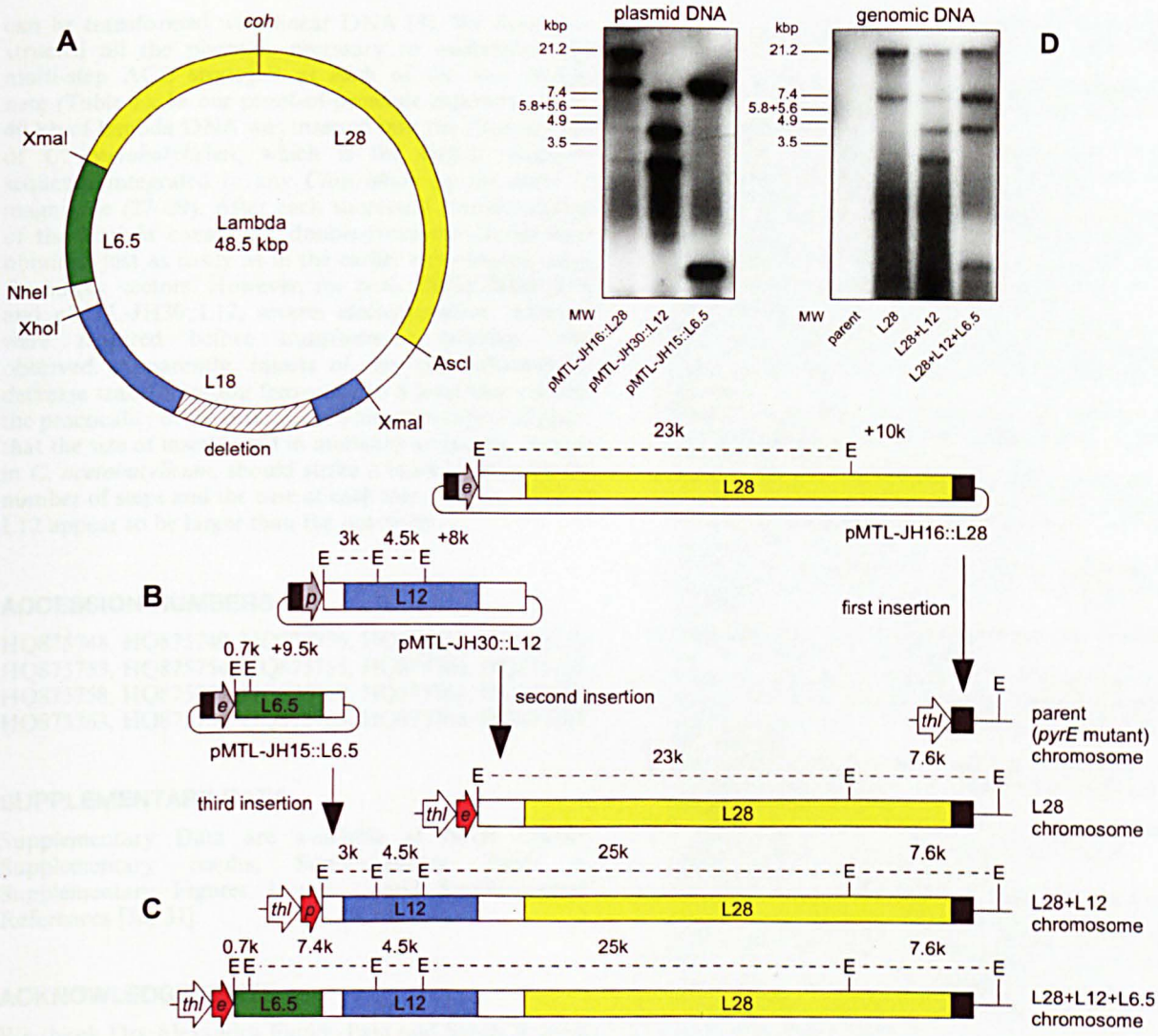


Figure 3. Multistep insertion of the chromosome of phage lambda into the chromosome of *C. acetobutylicum*. (**A**) Chromosome of phage lambda showing restriction sites used to excise the three overlapping fragments. Yellow, 28 kb XmaI-XmaI fragment L28; blue, 18 kb NheI-AscI fragment L18; green, 6.5 kb XmaI-XhoI fragment L6.5; white, regions of overlap between fragments; cross-shaded deletion, 6 kb region of L18 absent from L12; *coh*, ligated cohesive ends. (**B**) Integration plasmids pMTL-JH16::L28, pMTL-JH30::L12 and pMTL-JH15::L6.5. The lambda sequences are colored as in (A). Gray arrows *e* and *p*, inactive (non-expressed) *ermB* and *pyrE*, respectively; black, homology to *thl* locus; E, EcoRI sites; dashed lines, EcoRI fragments (except those spanning the plasmid backbone); numeric labels, EcoRI fragment lengths in kilo base pair. (**C**) The *thl* locus of *C. acetobutylicum* before and after 1, 2 and 3 insertions of lambda DNA. Elements are labeled as in (B). The first recombination event at each step, indicated, is directed by a long region of homology. A short region of homology mediates plasmid excision and simultaneously activates *ermB* or *pyrE* in alternate steps, shown by red arrows *e* and *p*, respectively, by positioning them under the control of the chromosomal *thl* promoter. (**D**) Southern blot of EcoRI digests of the plasmids and chromosomes shown in (B) and (C), using lambda DNA as probe. MW, HindIII-digested lambda DNA molecular weight marker.

The combination of these principles constitutes a novel, generally applicable strategy which we call allele-coupled exchange (ACE). The variant demonstrated using pMTL-JH15, 16, 30 and 31 is particularly interesting and broadly applicable, because providing the entire coding sequence of a heterologous marker on the plasmid means that any genomic locus with a promoter can be a target for integration. Furthermore, using a chromosomal promoter to direct expression of a heterologous gene of interest allows the gene to be cloned without a promoter, precluding toxicity which might otherwise result from over-expression in the cloning host. This variant also allows specific selection of double-crossover clones using only positively selectable markers, like antibiotic-resistance genes, which are more easily applied to diverse organisms than counter-selection markers.

Perhaps the most exciting implication of this work is the potential to iteratively introduce large insertions (beyond the size which could be delivered in a single step) in a similar way to procedures available for organisms which

can be transformed with linear DNA (3). We have constructed all the plasmids necessary to undertake such multi-step ACE strategies at each of the loci studied here (Table 1). In our proof-of-principle experiment over 40 kb of lambda DNA was inserted into the chromosome of *C. acetobutylicum*, which is the largest arbitrary sequence integrated in any *Clostridium* by an order of magnitude (27–29). After each successful transformation of the lambda constructs, double-crossover clones were obtained just as easily as in the earlier experiments using the empty vectors. However, for both pMTL-JH16::L28 and pMTL-JH30::L12, several electroporation attempts were required before transformant colonies were observed. Apparently, inserts of this size substantially decrease transformation frequency to a level that reduces the practicality of the procedure. This observation suggests that the size of inserts used in multistep strategies, at least in *C. acetobutylicum*, should strike a balance between the number of steps and the ease of each step. Inserts L28 and L12 appear to be larger than the optimum.

ACCESSION NUMBERS

HQ875748, HQ875749, HQ875750, HQ875751, HQ875752, HQ875753, HQ875754, HQ875755, HQ875766, HQ875757, HQ875758, HQ875759, HQ875760, HQ875761, HQ875762, HQ875763, HQ875764, HQ875765, HQ875766, HQ875767.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary results, Supplementary Table 1, Supplementary Figures 1 and 2, and Supplementary References [30, 31].

ACKNOWLEDGEMENTS

We thank Drs Alexandra Faulds-Pain and Sarah Kuehne for valuable criticism and suggestions during preparation of the manuscript.

FUNDING

Biotechnology and Biological Sciences Research Council (BB/F003390/1, BB/G016224/1, BB/E021271/1); Medical Research Council (G0601176); European Union (HEALTH-F3-2008-223585); TMO Renewables Ltd. Funding for open access charge: The University of Nottingham.

Conflict of interest statement. The University of Nottingham has filed a patent application encompassing some of the work described in this article. The patent application names J.TH. and N.P.M. as inventors.

REFERENCES

1. Collins,J.J., Endy,D., Hutchison,C.A. and Roberts,R.J. (2010) Editorial-synthetic biology. *Nucleic Acids Res.*, **38**, 2513.
2. Itaya,M. and Tanaka,T. (1990) Gene-directed mutagenesis on the chromosome of *Bacillus subtilis* 168. *Mol. Gen. Genet.*, **223**, 268–272.
3. Itaya,M., Fujita,K., Kuroki,A. and Tsuge,K. (2008) Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nat. Methods*, **5**, 41–43.
4. Reyrat,J.M., Pelicic,V., Gicquel,B. and Rappuoli,R. (1998) Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. *Infect. Immun.*, **66**, 4011–4017.
5. Purdy,D., O'Keeffe,T.A.T., Elmore,M., Herbert,M., McLeod,A., Bokori-Brown,M., Ostrowski,A. and Minton,N.P. (2002) Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol. Microbiol.*, **46**, 439–452.
6. Heap,J.T., Pennington,O.J., Cartman,S.T., Carter,G.P. and Minton,N.P. (2007) The ClosTron: a universal gene knock-out system for the genus *Clostridium*. *J. Microbiol. Methods*, **70**, 452–464.
7. Heap,J.T., Pennington,O.J., Cartman,S.T. and Minton,N.P. (2009) A modular system for *Clostridium* shuttle plasmids. *J. Microbiol. Methods*, **78**, 79–85.
8. Hartmanis,M.G.N. and Gatenbeck,S. (1984) Intermediary metabolism in *Clostridium acetobutylicum*: levels of enzymes involved in the formation of acetate and butyrate. *Appl. Environ. Microbiol.*, **47**, 1277–1283.
9. O'Brien,R.W. and Morris,J.G. (1971) Oxygen and the growth and metabolism of *Clostridium acetobutylicum*. *J. Gen. Microbiol.*, **68**, 307–318.
10. Papoutsakis,E.T. (2008) Engineering solventogenic clostridia. *Curr. Opin. Biotechnol.*, **19**, 420–429.
11. Boeke,J.D., LaCroute,F. and Fink,G.R. (1984) A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.*, **197**, 345–346.
12. Husson,R.N., James,B.E. and Young,R.A. (1990) Gene replacement and expression of foreign DNA in mycobacteria. *J. Bacteriol.*, **172**, 519–524.
13. Sander,P., Meier,A. and Böttger,E.C. (1995) rpsL+: a dominant selectable marker for gene replacement in mycobacteria. *Mol. Microbiol.*, **16**, 991–1000.
14. Knipfer,N., Seth,A. and Shrader,T.E. (1997) Unmarked gene integration into the chromosome of *Mycobacterium smegmatis* via precise replacement of the *pyrF* gene. *Plasmid*, **37**, 129–140.
15. Sato,T., Fukui,T., Atomi,H. and Imanaka,T. (2003) Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.*, **185**, 210–220.
16. Bitan-Banin,G., Ortenberg,R. and Mevarech,M. (2003) Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene. *J. Bacteriol.*, **185**, 772–778.
17. Deng,L., Zhu,H., Chen,Z., Liang,Y.X. and She,Q. (2009) Unmarked gene deletion and host-vector system for the hyperthermophilic crenarchaeon *Sulfolobus islandicus*. *Extremophiles*, **13**, 735–746.
18. Tripathi,S.A., Olson,D.G., Argyros,D.A., Miller,B.B., Barrett,T.F., Murphy,D.M., McCool,J.D., Warner,A.K., Rajgarhia,V.B., Lynd,L.R. *et al.* (2010) Development of pyrF-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a pta mutant. *Appl. Environ. Microbiol.*, **76**, 6591–6599.
19. Monod,M., Denoya,C. and Dubnau,D. (1986) Sequence and properties of pIM13, a macrolide-lincosamide-streptogramin B resistance plasmid from *Bacillus subtilis*. *J. Bacteriol.*, **167**, 138–147.
20. Theys,J., Pennington,O., Dubois,L., Anlezark,G., Vaughan,T., Mengesha,A., Landuyt,W., Anné,J., Burke,P.J., Dürre,P. *et al.* (2006) Repeated cycles of *Clostridium difficile*-directed enzyme prodrug therapy result in sustained antitumour effects *in vivo*. *Br. J. Cancer*, **95**, 1212–1219.
21. Hussain,H.A., Roberts,A.P. and Mullany,P. (2005) Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630Deltaerm) and demonstration that the conjugative

transposon Tn916DeltaE enters the genome of this strain at multiple sites. *J. Med. Microbiol.*, **54**, 137–141.
22. Tummala,S.B., Welker,N.E. and Papoutsakis,E.T. (1999) Development and characterization of a gene expression reporter system for *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.*, **65**, 3793–3799.
23. Harris,L.M., Welker,N.E. and Papoutsakis,E.T. (2002) Northern, morphological, and fermentation analysis of *spo0A* inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.*, **184**, 3586–3597.
24. Clewell,D.B., Yagi,Y., Dunny,G.M. and Schultz,S.K. (1974) Characterization of three plasmid deoxyribonucleic acid molecules in a strain of *Streptococcus faecalis*: identification of a plasmid determining erythromycin resistance. *J. Bacteriol.*, **117**, 283–289.
25. Jones,D.T. and Woods,D.R. (1986) Acetone-butanol fermentation revisited. *Microbiol. Rev.*, **50**, 484–524.
26. Ismaiel,A.A., Zhu,C.X., Colby,G.D. and Chen,J.S. (1993) Purification and characterization of a primary-secondary alcohol dehydrogenase from two strains of *Clostridium beijerinckii*. *J. Bacteriol.*, **175**, 5097–5105.
27. Heap,J.T., Kuehne,S.A., Ehsaan,M., Cartman,S.T., Cooksley,C.M., Scott,J.C. and Minton,N.P. (2010) The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. *J. Microbiol. Methods*, **80**, 49–55.
28. Chen,Y., Caruso,L., McClane,B., Fisher,D. and Gupta,P. (2007) Disruption of a toxin gene by introduction of a foreign gene into the chromosome of *Clostridium perfringens* using targetron-induced mutagenesis. *Plasmid*, **58**, 182–189.
29. Cartman,S.T. and Minton,N.P. (2010) A mariner-based transposon system for *in vivo* random mutagenesis of *Clostridium difficile*. *Appl. Environ. Microbiol.*, **76**, 1103–1109.
30. Brefort,G., Magot,M., Ionesco,H. and Sebald,M. (1977) Characterization and transferability of *Clostridium perfringens* plasmids. *Plasmid*, **1**, 52–66.
31. Karlsson,S., Burman,L.G. and Akerlund,T. (1999) Suppression of toxin production in *Clostridium difficile* VPI 10463 by amino acids. *Microbiology*, **145**, 1683–1693.

Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker

John T Heap*† Muhammad Ehsaan, Clare M Cooksley, Yen-Kuan Ng, Stephen T Cartman, Klaus Winzer and Nigel P Minton*

Clostridia Research Group, BBSRC Sustainable BioEnergy Centre, School of Molecular Medical Sciences, Centre for Biomolecular Sciences, The University of Nottingham, University Park, Nottingham, NG7 2RD, UK

Supplementary Data

Integration of small and medium-sized DNA inserts at the *pyrF* locus of *Clostridium acetobutylicum*

The *pyrF* gene encodes orotidine 5-phosphate decarboxylase, which participates in the same pyrimidine biosynthesis pathway as orotate phosphoribosyltransferase, encoded by *pyrE*. Inactivation of either *pyrF* or *pyrE* will lead to the same FOA^R phenotype, and both genes are required for uracil prototrophy. These selections can be applied to *pyrF* in the same way as *pyrE*. We constructed plasmid pMTL-JH2 according to the same design principles as pMTL-JH12, but targeting the *pyrF* locus (Supplementary Figure 1A). We transformed *C. acetobutylicum* cells with pMTL-JH2 DNA, followed the same integration procedure as for pMTL-JH12, and screened four independent FOA-resistant clones. All four clones showed the desired double-crossover recombinant genotype by PCR (Figure 2B), and were thiamphenicol-sensitive, indicating the loss of plasmid DNA. This experiment used the 'empty' pMTL-JH2 vector, so the only heterologous sequence delivered to the chromosome was the small *lacZ* sequence which provides the vector's multiple cloning site. At time of these experiments, we had not yet demonstrated integration of large DNA inserts. To confirm that pMTL-JH2 could also be used to deliver larger DNA sequences, we constructed four derivatives of pMTL-JH2 by inserting different-

*To whom correspondence may be addressed.
Nigel Minton: nigel.minton@nottingham.ac.uk.
John Heap: j.heap@imperial.ac.uk.
†Present address: Centre for Synthetic Biology and Innovation, Division of Molecular Biosciences, Imperial College London, London, SW7 2AZ, UK.

sized fragments of phage lambda DNA into the allele exchange cassette (Supplementary Figure 1C).

Lambda DNA-HindIII Digest (NEB) is a preparation of phage lambda (*cI857ind1 Sam 7*) DNA digested to completion with HindIII. The mixture of fragments in this preparation was heated to 60 °C for 5 min to separate the lambda cohesive ends of the fragments derived from the ends of the lambda chromosome, treated with T4 polymerase (NEB) to convert the HindIII cohesive ends and lambda cohesive ends to blunt ends, then individual blunt fragments were purified by agarose gel electrophoresis. Four derivatives of pMTL-JH2 were constructed by ligating the 2.0, 2.3, 4.3 or 6.5 kbp blunt fragment of lambda DNA with pMTL-JH2 DNA linearised using StuI and treated with Antarctic Phosphatase (NEB) to prevent re-circularisation. This positioned the lambda DNA in each pMTL-JH2 derivative between the two regions of homology so that the lambda DNA would be delivered to the chromosome of a double-crossover clone by a successful integration experiment. These plasmids were designated pMTL-JH2-lambda2.0, pMTL-JH2-lambda2.3, pMTL-JH2-lambda4.3 and pMTL-JH2-lambda6.5 accordingly. The derivatives of pMTL-JH2 were each sequenced using the M13F primer to determine the orientation of the insert (because blunt cloning is not directional) which would subsequently inform the design of the screening PCR.

C. acetobutylicum cells were transformed with each of the lambda DNA-containing derivatives of pMTL-JH2, and the integration procedure was performed as described in the main text for pMTL-JH12. FOA-resistant clones were screened by PCR and thiamphenicol-sensitivity by replica-plating. Each of these four inserts was successfully delivered to the *C. acetobutylicum* chromosome.

DNA integration at the *pyrE* locus of *Clostridium sporogenes* and *Clostridium difficile*

We constructed plasmids pMTL-JH18 and pMTL-JH27 to target the *pyrE* locus of *C. difficile* 630 and *C. sporogenes* NCIMB 10696 respectively, using a cassette design equivalent to pMTL-JH12 (Figure 1, Table 1). The choice of replicon for pMTL-JH18 and pMTL-JH27 was guided by our recent work [7, 29, and our unpublished observations] which suggests that the pIM13 replicon would limit the growth of *C. sporogenes* NCIMB 10696 colonies when under antibiotic selection for a resistance marker on the plasmid, providing the means to enrich and isolate faster-growing single-crossover clones; while the replicon from pCB102 would be more appropriate to achieve the same in *C. difficile* 630. Both plasmids also include a transfer function to allow conjugation from an *E. coli* donor. We tested pMTL-JH18 in *C. difficile* 630Δ*erm* [21] and pMTL-JH27 in *C. sporogenes* NCIMB 10696, using equivalent integration procedures as for pMTL-JH12 in *C.*

acetobutylicum, but with media and conditions adapted to suit the different hosts.

Plasmids were transferred by conjugation from *E. coli* CA434 as described previously [7]. Transconjugants (clones which successfully received pMTL-JH27) of *C. sporogenes* were selected on TYG [30] agar supplemented with 250 µg/ml D-cycloserine to counter-select *E. coli* and 15 µg/ml thiamphenicol to select for the plasmid-borne resistance marker *catP*. Cells were sub-cultured twice on the same medium, then sub-cultured twice on TYG agar supplemented with 3 mg/ml 5-fluoroorotic acid (FOA) to select and isolate *pyrE*-minus clones. Four FOA-resistant colonies were screened by PCR (Supplementary Figure 1A) and screened for thiamphenicol-sensitivity by replica-plating, and the desired genotype was confirmed for all four.

Transconjugants of *C. difficile* 630Δ*erm* containing pMTL-JH18 were selected on BHI agar supplemented with 250 µg/ml D-cycloserine and 8 µg/ml cefoxitin to counter-select *E. coli*, and 15 µg/ml thiamphenicol to select for the plasmid-borne resistance marker *catP*. Thiamphenicol-resistant *C. difficile* colonies containing pMTL-JH18 all grew quickly, at a rate similar to the wild-type on unsupplemented medium. Apparently the replication defect of pMTL-JH18 does not substantially limit the growth of *C. difficile* colonies on BHI agar supplemented with thiamphenicol, meaning that enrichment for single-crossover clones is not easily achieved. Consequently we were not able to obtain single-crossover clones using thiamphenicol, nor double-crossover clones using FOA. However, we have noted that the extent of the replication defect of some plasmids is dependent upon the size of the plasmid (Heap, Cartman, Ehsaan, Ng and Minton, unpublished observations). To attempt to exacerbate the replication defect of pMTL-JH18, a 6.5 kbp fragment of phage lambda DNA was inserted into the allele exchange cassette (in the same way as described above for pMTL-JH2) increasing the size of the plasmid from 5.9 kbp to 12.5 kbp, and the integration experiment was repeated. The thiamphenicol-resistant *C. difficile* colonies obtained showed a variety of growth rates, and the faster-growing colonies (corresponding to single-crossover clones) were enriched by subculturing twice on the same medium. Cells were then sub-cultured twice on CDM [31] agar supplemented with 2 mg/ml FOA and 5 µg/ml uracil to select and isolate *pyrE*-minus clones. FOA-resistant colonies were easily obtained, and five independent clones were screened as before by PCR (Supplementary Figure 1B) and thiamphenicol-sensitivity, and confirmed to be recombinants.

References

[7] Heap, J. T., Pennington, O. J., Cartman, S. T., and Minton, N. P. (Jul, 2009) A modular system for *Clostridium* shuttle plasmids. *J Microbiol*

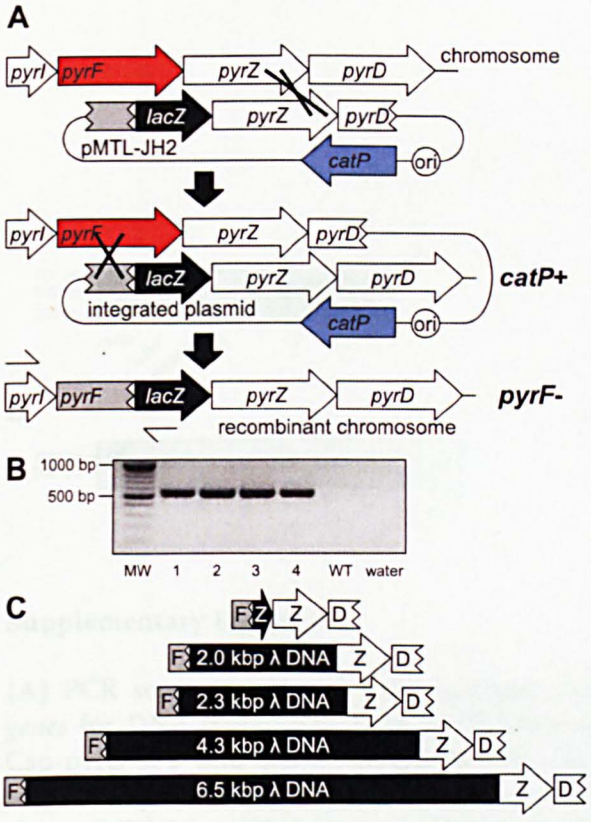
Methods, **78**(1), 79–85.

[29] Cartman, S. T. and Minton, N. P. (Feb, 2010) A mariner-based transposon system for *in vivo* random mutagenesis of *Clostridium difficile*. *Appl Environ Microbiol*, **76**(4), 1103–1109.

[21] Hussain, H. A., Roberts, A. P., and Mullany, P. (Feb, 2005) Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630Deltaerm) and demonstration that the conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites. *J Med Microbiol*, **54**(Pt 2), 137–141.

[30] Brefort, G., Magot, M., Ionesco, H., and Sebald, M. (1977) Characterization and transferability of *Clostridium perfringens* plasmids. *Plasmid*, 1(1), 52–66.

[31] Karlsson, S., Burman, L. G., and Akerlund, T. (Jul, 1999) Suppression of toxin production in *Clostridium difficile* VPI 10463 by amino acids. *Microbiology*, **145**(Pt 7), 1683–1693.

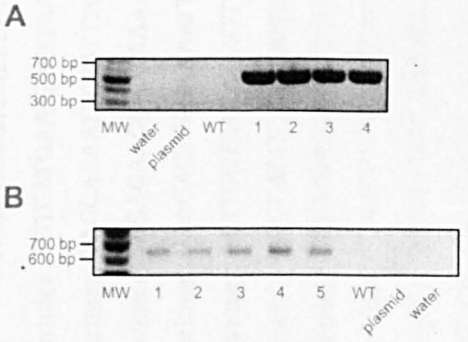


Supplementary Figure 1: DNA integration at the *pyrF* locus of *C. acetobutylicum*.

(**A**) Selection of stable double-crossover clones using pMTL-JH2. The first recombination event (plasmid integration) is mediated by the long region of homology between pMTL-JH2 and *pyrZ*/*pyrD*. Single-crossover clones are obtained on medium containing thiamphenicol. The second recombination event (plasmid excision) is mediated by the short region of homology between pMTL-JH2 and an internal fragment of *pyrF*. Double-crossover clones are selected using FOA.

(**B**) PCR screening of four candidate double-crossover clones using primers Cac-pyrI-sF1 and M13F which anneal where indicated in (A). MW, 2-Log DNA Ladder (NEB) molecular weight marker; 1–4, candidate clones; WT, wild-type *C. acetobutylicum* genomic DNA control. All four candidates show the expected 549 bp band.

(**C**) Homologous recombination cassette of pMTL-JH2 and four derivatives containing different fragments of lambda DNA, shown to scale.



Supplementary Figure 2:

(A) PCR screening of four FOA-resistant candidate clones of *C. sporogenes* for DNA integration at the *pyrE* locus using pMTL-JH27. Primers Csp-pyrD-sF2 and M13F anneal to the chromosome and *lacZ* insert respectively, resulting in a 541 bp PCR product from double-crossover clones. MW, 2-Log DNA Ladder (NEB) molecular weight marker; plasmid, pMTL-JH27 plasmid DNA control; WT, wild-type *C. sporogenes* genomic DNA control; 1–4, candidate clones. All four candidates show the expected 558 bp band.

(B) PCR screening of five FOA-resistant candidate clones of *C. difficile* for DNA integration at the *pyrE* locus using pMTL-JH18::lambda6.5 (pMTL-JH18 with the 6.5 kbp lambda DNA insert). Primers Cdi630-pyrD-sF1 and lambda6.5-sF2 anneal to the chromosome and lambda insert respectively, resulting in a 623 bp PCR product from double-crossover clones. MW, 2-Log DNA Ladder (NEB) molecular weight marker; plasmid, pMTL-JH18::lambda6.5 plasmid DNA control; WT, wild-type *C. sporogenes* genomic DNA control; 1–5, candidate clones. All five candidates show the expected 623 bp band.

Oligonucleotide	Binding site	Purpose	Sequence 5′ to 3′
M13F	*lacZ* in integration vectors	Plasmid-specific reverse primer	TGTAAAACGACGGCCAGT
lacZa-sF2	*lacZ* in integration vectors	Plasmid-specific forward primer	ACTGGCCGTCGTTTTACAACGTCGTG
Cac-pyrI-sF1	*pyrI* of *C. acetobutylicum*	Chromosome-specific forward primer	TGTGATGAAATATATAAGGGAGCAAAGGCGC
CAC0026-sF2	CAC0026 of *C. acetobutylicum*	Chromosome-specific forward primer	TAGCACAATTGTATTTGGACTTCTTTAAATAAAAACATGG
Cdi630-pyrD-sF1	*pyrD* of *C. difficile*	Chromosome-specific forward primer	TAGAGAAGGAATAAAAAGTTTAGACGAAATAAGAGG
Csp-pyrD-sF2	*pyrD* of *C. sporogenes*	Chromosome-specific forward primer	GAAGACTTAGAAAATTATATGAAAGAAGAAGGT
Cac-thl-sF1	*thl* of *C. acetobutylicum*	Chromosome-specific forward primer	ACTTGCTAAGATAGTTTCTTATGGTTCAGCAGG
Cac-pyrD-sR1	*pyrD* of *C. acetobutylicum*	Chromosome-specific reverse primer	AGCCATATCCACTATATCCTCTGCATTAGG
Cac-hydA-sR2	*hydA* of *C. acetobutylicum*	Chromosome-specific reverse primer	TTGATGATGTTTGTCTTGATGACTCAACATGC
Cdi630-CD0189-sR3	CD0189 of *C. difficile*	Chromosome-specific reverse primer	CCAAGCTCTATGACAGACAGCTCATTGTTTAGAAC
Csp-3233-sR4	CBO3233 homolog of *C. sporogenes*	Chromosome-specific reverse primer	AATATGGTTCAAAAGCACTTGAAAA
Cac-atpB-sR1	*atpB* of *C. acetobutylicum*	Chromosome-specific reverse primer	ATGATACTGGTATTGTAACCTTTTCTAAAAGGTTCATAGG

Supplementary Table 1:

List of oligonucleotide primers. PCR screening was performed using appropriate combinations of primers as described in the text and figure legends. Validated chromosome-specific reverse primers for each locus targeted in this study are also listed to provide researchers with further screening options.

OPEN ACCESS Freely available online

PLOS | ONE

Expanding the Repertoire of Gene Tools for Precise Manipulation of the *Clostridium difficile* Genome: Allelic Exchange Using *pyrE* Alleles

Yen Kuan Ng[1☯], Muhammad Ehsaan[1☯], Sheryl Philip[1], Mark M. Collery[1], Clare Janoir[2], Anne Collignon[2], Stephen T. Cartman[1], Nigel P. Minton[1*]

1 Clostridia Research Group, NIHR Biomedical Research Unit in GI Disease, Centre for Biomolecular Sciences, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom, **2** Université Paris-Sud, Faculté de Pharmacie, Département de Microbiologie, Unité EA 40-43, Châtenay-Malabry, France

Abstract

Sophisticated genetic tools to modify essential biological processes at the molecular level are pivotal in elucidating the molecular pathogenesis of *Clostridium difficile*, a major cause of healthcare associated disease. Here we have developed an efficient procedure for making precise alterations to the *C. difficile* genome by *pyrE*-based allelic exchange. The robustness and reliability of the method was demonstrated through the creation of in-frame deletions in three genes (*spo0A*, *cwp84*, and *mtlD*) in the non-epidemic strain 630Δ*erm* and two genes (*spo0A* and *cwp84*) in the epidemic PCR Ribotype 027 strain, R20291. The system is reliant on the initial creation of a *pyrE* deletion mutant, using Allele Coupled Exchange (ACE), that is auxotrophic for uracil and resistant to fluoroorotic acid (FOA). This enables the subsequent modification of target genes by allelic exchange using a heterologous *pyrE* allele from *Clostridium sporogenes* as a counter-/negative-selection marker in the presence of FOA. Following modification of the target gene, the strain created is rapidly returned to uracil prototrophy using ACE, allowing mutant phenotypes to be characterised in a PyrE proficient background. Crucially, wild-type copies of the inactivated gene may be introduced into the genome using ACE concomitant with correction of the *pyrE* allele. This allows complementation studies to be undertaken at an appropriate gene dosage, as opposed to the use of multicopy autonomous plasmids. The rapidity of the 'correction' method (5–7 days) makes *pyrE*$^-$ strains attractive hosts for mutagenesis studies.

Citation: Ng YK, Ehsaan M, Philip S, Collery MM, Janoir C, et al. (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. doi:10.1371/journal.pone.0056051

Editor: Michel R. Popoff, Institute Pasteur, France

Received August 10, 2012; **Accepted** January 8, 2013; **Published** February 6, 2013

Funding: The authors acknowledge the financial support of the MRC (G0601176), the European Community's Seventh Framework Programme 'HYPERDIFF' (HEALTH-F3-2008-223585), the Biotechnology and Biological Sciences Research Council (BB/G016224/1), and TMO686 Renewables Ltd. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: One of the authors, Muhammad Ehsaan is a PhD student sponsored by TMO Renewables Ltd. His stipend and fee is received by the University from TMO and administered by them. Neither he nor Nigel P. Minton have any other association with this company (neither in terms of salary, consultancy or shares). The work has not been patented. The funding of Muhammad Ehsaan's studentship does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: Nigel.Minton@nottingham.ac.uk

☯ These authors contributed equally to this work.

Introduction

Clostridium difficile is the major cause of nosocomial diarrhoea and a major burden to healthcare services worldwide. The organism is resistant to various broad-spectrum antibiotics and capitalises on the disruption of the normal intestinal flora to cause disease symptoms ranging from asymptomatic carriage to a fulminant, relapsing and potentially fatal colitis [1,2]. Sophisticated genetic tools to modify essential biological processes at the molecular level are pivotal in enabling the systematic study of the basis of colonization, virulence and pathogenesis of *C. difficile*. The requisite systems ideally need to be able to both make precise, *in situ* alterations to existing alleles, as well as introduce entirely new alleles to progress hypothesis driven research. The former include the generation of in-frame deletions and the introduction or correction of single or multiple nucleotide substitutions, deletions and insertions. The latter is required both to add entirely new, or altered alleles or provide the facility for the complementation of distal mutant alleles at chromosomal gene dosage.

For many years the required level of sophistication has been unavailable. Thus, directed mutants could only be made using insertional mutagens, reliant either on replication deficient [3] or defective [4,5] plasmids, or on the deployment of the ClosTron and group II intron re-targeting [6,7]. Recently, however, the cytosine deaminase gene (*codA*) of *Escherichia coli* was developed as a negative/counter selection marker for *C. difficile*, which enabled precise manipulation of the *C. difficile* chromosome for the first time [8]. In parallel, a second method (Allele-Coupled Exchange, ACE) has been formulated that allows the rapid insertion of heterologous DNA, of any size or complexity, into the genome [9]. Whilst a number of different genetic loci may be used to insert heterologous DNA via ACE, one exemplification of the method exploits the native *pyrE* gene, bringing about its inactivation by replacement of the wild type allele with a mutant allele lacking the codons from both the 5′ and 3′ end of the structural gene. The

pyrE gene encodes orotate phosphoribosyltransferase [E.C.2.4.2.10]), which is an enzyme involved in *de novo* pyrimidine biosynthesis. It may be used as a positive/negative selection marker as it is essential in the absence of exogenous pyrimidines and it also renders 5-fluoro-orotate (FOA) toxic to cells. Toxicity occurs via a series of steps which result in misincorporation of fluorinated nucleotides into DNA and RNA and hence, cause cell death.

Mutant strains defective in *pyrE* created using ACE become auxotrophs requiring exogenous uracil to grow. They also lends themselves to the use of a functional *pyrE* allele as a negative/counter selection marker in a similar way to *codA* [8].The functionally equivalent gene, *URA5*, and *URA3* (encoding orotidine-phosphate decarboxylase) have been widely used as negative selection markers in *Saccharomyces cerevisiae* [10]. Indeed, this approach has allowed homologous gene replacement in a range of different microbes [11–14], including most recently *Clostridium thermocellum* [15] using the organisms own *pyrF* gene (equivalent to *URA3*) in a specifically created Δ*pyrF* mutant. Not unexpectedly, the Δ*pyrF* strain created required the addition of exogenous uracil to achieve equivalent cell density to the wild type in the rich media employed. With supplementation, whilst growth was slightly delayed, the eventual the growth rate was comparable to wildtype [15].

In this study, we have developed a procedure for the generation of allelic exchange mutants in *pyrE* mutants of two different strains of *C. difficile* (strain 630Δ*erm* and a BI/NAP1/PCR-Ribotype 027 strain, R20291) using a heterologous *pyrE* allele from *Clostridium sporogenes* as a negative/counter-selection marker. Use of a heterologous *pyrE* allele avoids homologous recombination with the native *pyrE* locus of *C. difficile*. The system has been used to make in-frame deletions in three different genes, *spo0A* (the master regulator of sporulation), *cwp84* (which encodes a cysteine protease) and *mtlD* (which encodes mannitol-1-phosphate 5-dehydrogenase). Crucially, having created these mutants, a specific ACE-vector is used to rapidly (within 5–7 days) restore the chromosomal *pyrE* allele to wild-type, allowing the specific in-frame deletion mutant to be characterised in a clean, wild-type background. Moreover, variants of the same vector may be used to deliver the wild-type allele of the deleted gene, either under the control of its native promoter or the strong *fdx* promoter, into the genome. The former allows complementation studies to be performed at an appropriate gene dosage, while the latter potentially allows the assessment of the effect of overexpressing the gene.

Results

Construction of a *pyrE*$^{-}$ mutant of R20291

The *pyrE*$^{-}$ mutant of strain 630Δ*erm* was previously made using Allele-Coupled Exchange (ACE) and the purpose built, replication defective vector pMTL-JH18::lambda6.5 [9]. It is based on the replicon of plasmid pCB102 [16], previously shown to be the least effective (most unstable in terms of segregation into daughter cells) in strain 630Δ*erm* of the four replicons available in the pMTL80000 modular vector series [17]. In strain R20291, however, the plasmid pBP1 replicon is the most defective, *ie.*, plasmids that use this replication region are more rapidly lost in the absence of antibiotic selection [18]. Accordingly, pMTL-YN18 was constructed in order to generate a *pyrE* mutant of *C. difficile* R20291 (Figure 1). pMTL-YN18 is equivalent to the pMTL-JH18::lambda6.5 vector used to construct a Δ*pyrE* mutant in *C. difficile* 630Δ*erm*, except the pCB102 replicon is substituted with that of pBP1 and the 6.5 kb fragment of phage lambda DNA is absent. It carries two asymmetric homology arms that flank a *lacZ*' multiple cloning site (MCS) region. The shorter, 300 bp left-hand homology arm (LHA) is composed of a variant *pyrE* allele that lacks 50 nucleotides from the 5'-end of the gene and is devoid of 235 bp from the 3'-end of the gene. The larger 1200 bp right-hand homology arm (RHA) encompasses the region residing immediately downstream of the *pyrE* gene in the *C. difficile* genome.

Plasmid pMTL-YN18 was used to generate a *pyrE* mutant of *C. difficile* R20291. Five independent fluoroorotic acid (FOA) resistant colonies were chosen and shown to be *pyrE* deletion mutants. Thus, they all required exogenous uracil for growth on minimal media, were no longer thiamphenicol resistant, consistent with plasmid loss, and generated a PCR fragment of the expected size (623 bp) using primers that flanked the *pyrE* locus (Supporting Information File S1). Nucleotide sequence of the 623 bp fragment generated confirmed the presence of the expected deletion.

Restoration of *pyrE*+ phenotype in 630Δ*erm* and R20291

With *pyrE* mutants of both *C. difficile* 630Δ*erm* and R20291, it was desirable to demonstrate that they could be converted back to uracil prototrophy, through restoration of the *pyrE* locus to wild-type. Accordingly, ACE plasmids pMTL-YN1 and pMTL-YN2 were constructed to correct the *pyrE* mutations in *C. difficile* 630Δ*erm* and R20291, respectively (Figure 2). These plasmids were

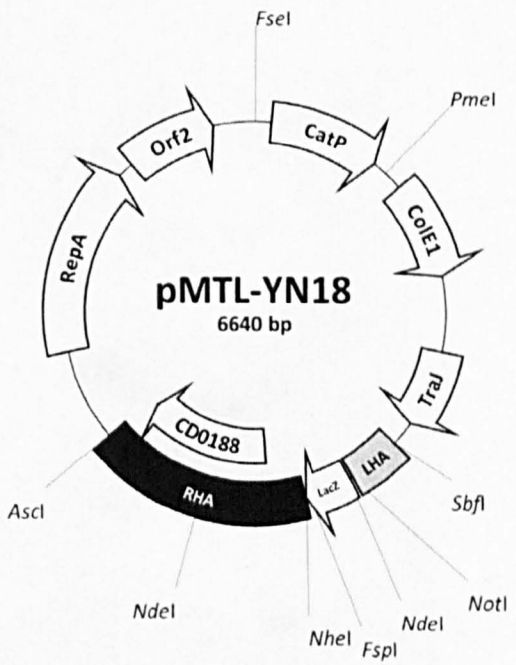


Figure 1. Plasmid pMTL-YN18. The ACE vector pMTL-YN18 is designed to create a deletion mutant specifically in the *C. difficile* strain R20191. Plasmid components are: CatP, the *catP* gene of *Clostridium perfringens* conferring thiamphenicol resistance; ColE1, the replication region of the *E.coli* plasmid ColE1; TraJ, transfer function of the RP4 *oriT* region; RepA and Orf2, the replication region of the *Clostridium botulinum* plasmid pBP1; LHA, left-hand homology arm encompassing a 300 bp internal fragment of the R20291 *pyrE* gene lacking 50 nucleotides from the 5'-end, and 235 bp from the 3'-end; RHA, right-hand homology arm comprising encompassing the 1200 bp region of DNA immediately downstream of *pyrE*, and; *lacZ'*, gene encoding the alpha fragment of the *E.coli* β-galactosidase (and containing a multiple cloning site region derived from plasmid pMTL20 [19]).
doi:10.1371/journal.pone.0056051.g001

broadly equivalent to pMTL-JH14, used to correct a similar *pyrE* mutation in *C. acetobutylicum* [9], with a *pyrE* allele that lacked the first 50 nucleotides of the open-reading-frame forming the short 300 bp LHA. Crucially, they differed from pMTL-JH14 in that they did not carry the *lacZ'* region present in this plasmid between the LHA and RHA. As a consequence, the strains generated using them would be indistinguishable from the wild-type at the *pyrE* locus. Both plasmids were conjugated into their respective *pyrE* mutants from an *E. coli* CA434 donor and, following the procedure outlined in MATERIALS AND METHODS, several dozen colonies were obtained that were able to grow without uracil supplementation. Six random colonies from each strain were screened by PCR using primers which annealed to the chromosomal genes that reside up- and down-stream of the *pyrE* gene. In all cases, the expected 2,058 bp band was generated, equivalent to the wild-type control, whereas no band was generated when the template DNA was derived from the *pyrE* mutant (Supporting Information File S1). Nucleotide sequencing of the fragment generated confirmed that the cells carried a wild type *pyrE* gene.

Formulation of the components of an Allelic Exchange vector

Aside from the gene specific knock-out cassettes, the two most important components of the desired allelic exchange vector are the *pyrE* allele, to be used as the negative/counter-selection marker, and the clostridial plasmid replicon. In the case of the *pyrE* allele, it is desirable that the gene used is heterologous, to avoid recombination at the native chromosomal *pyrE* locus of *C. difficile*. Therefore, we selected the *pyrE* gene of *Clostridium sporogenes*, which shares only 65% identity with its *C. difficile* equivalent. The *pyrE* gene was cloned into pMTL83151 and the resultant plasmid (pMTL-ME2, Figure 4) shown to be able to restore the *C. difficile pyrE* mutants to prototrophy, *ie.*, they were able to grow on minimal media lacking uracil supplementation (data not shown).

The isolation of the initial, single cross-over plasmid integrants required for the creation of allelic exchange mutants is facilitated by the use of 'pseudo-suicide' vectors [8,18]. These are plasmids carrying *catP* (encoding resistance to thiamphenicol) that are sufficiently replication defective for there to be a significant growth disadvantage in the presence of thiamphenciol compared to cells in which the plasmid, together with *catP*, has integrated. Such integrated clones grow faster, and produce larger colonies, on agar media supplemented with thiamphenicol, because all of the progeny carry a copy of *catP*. Cells carrying *catP* on a non-integrated, defective autonomous plasmid, in contrast, grow slower because in the presence of antibiotic they are limited by the rate at which the plasmid is segregated amongst the daughter population. Previously, the pCB102-based plasmid had been made more defective in strain 630Δ*erm* by increasing the size of the final vector through the insertion of 6.5 kb of DNA derived from phage

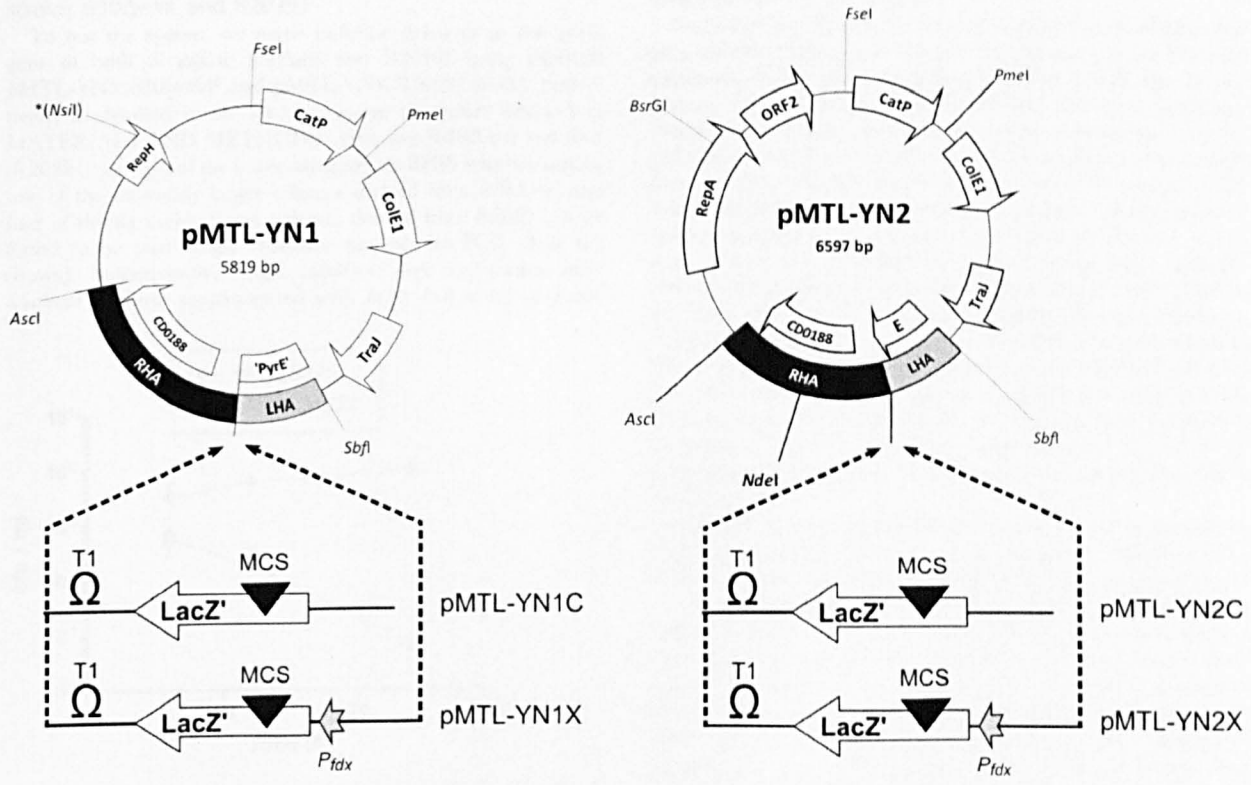


Figure 2. PyrE ACE correction vectors for *C. difficile* 630Δ*erm* (pMTL-YN1) and R20291 (pMTL-YN2). Both vectors carry identical components between their FseI and SbfI restriction sites. These are: CatP, the *catP* gene of *Clostridium perfringens* conferring thiamphenicol resistance; ColE1, the replication region of the *E.coli* plasmid ColE1, and; TraJ, transfer function of the RP4 *oriT* region. Plasmids pMTL-YN1C and pMTL-YN2C have an additional segment of DNA inserted between the left-hand homology arm (LHA) and the right-hand homology arm (RHA) which carries: a transcriptional terminator (T1) of the ferredoxin gene of *Clostridium pasteurianum*; a copy of the *lacZ'* gene encoding the alpha fragment of the *E.coli* β-galactosidase, and; a multiple cloning site (MCS) region derived from plasmid pMTL20 [19]. Plasmids pMTL-YN1X and pMTL-YN2X differ from pMTL-YN1C and pMTL-YN2C, respectively, in that they carrying the promoter region (P_{fdx}) of the *Clostridium sporogenes* ferredoxin gene.
doi:10.1371/journal.pone.0056051.g002

lambda [9]. Here we took a different strategy, and explored the effect of introducing a frame-shift mutation into the 3′-end of putative *repH* gene of the pCB102 replication region. Accordingly, the modular plasmid pMTL83251 was cleaved with NsiI, treated with T4 polymerase and subjected to self-ligation, yielding pMTL83*251. Nucleotide sequencing confirmed that the expected modification had occurred, altering the sequence ATGCAT to AT. The resulting deletion of 4 nucleotides (TGCA) causes a frame-shift in the coding sequence, replacing the COOH-terminal region of RepH (CIKYYARSFKKAHVKKSKKKK) with LNIMGALKKLM. The effect on segregational stability was tested by growing cells of 630Δ*erm* carrying either pMTL83251 or pMTL83*251 in the absence of antibiotic selection and measuring plasmid loss. The results are shown in Figure 3, and demonstrate that the frame-shift has caused a significant reduction in the segregational stability of the plasmid.

Accordingly, the NsiI site of plasmid pMTL-ME2 was similarly frame-shifted, and the resulting plasmid was designated pMTL-YN3 (Figure 4). For manipulations in R20291, the pCB102 replication region of pMTL-YN3 was simply replaced with that of pBP1, by cleaving pMTL-YN3 with AscI and FseI and replacing the excised fragment carrying the pCB102 replicon with the equivalent fragment from the pBP1-based plasmid pMTL82251 [17]. The resulting plasmid was designated pMTL-YN4 (Figure 4).

Exemplification of the System to create *spo0A* mutants of strains 630Δ*erm* and R20291

To test the system, we made in-frame deletions in the *spo0A* gene of both *C. difficile* 630Δ*erm* and R20291 using plasmids pMTL-YN3::630*spo0A** and pMTL-YN4::R20291*spo0A**, respectively, as detailed in the allele exchange procedure outlined in MATERIALS AND METHODS. After two (630Δ*erm*) and four (R20291) passages of the transconjugants on BHIS selective media, one of the six visibly larger colonies derived from 630Δ*erm*, and four of the six visibly larger colonies derived from R20291, were found to be pure single crossover mutants by PCR (data not shown). Subsequently, single colonies were re-streaked onto minimal medium supplemented with FOA and uracil to select for cells in which the integrated plasmid had excised. The isolated FOA resistant colonies (four in the case of 630Δ*erm* and twelve in the case of R20291) were then screened by PCR using primers spo0A-YN-F2 and spo0A-YN-R2 that anneal to the upstream and the downstream sequence of *spo0A*, respectively. Of the four 630Δ*erm* colonies screened, one yielded the expected 1,845 bp DNA fragment, indicative of an in-frame deletion (Supporting Information File S1). The other colonies yielded a 2,331 bp DNA fragment, consistent with the presence of a wild-type copy of the gene. In the case of R20291, two of the twelve colonies screened yielded the 1,845 bp DNA fragment indicative of an in-frame deletion (Supporting Information File S1). Nucleotide sequencing of the 1,845 bp DNA fragments confirmed that the expected allelic exchange event had taken place in all three putative mutants.

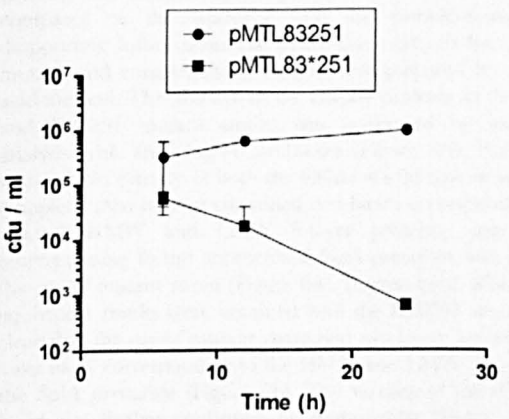


Figure 3. Segregational stability of pMTL83*251 or pMTL83251 in *C. difficile* strain 630Δ*erm*. The two plasmids differ only in that pMTL83*251 has a frame shift in the pCB102 RepH gene, introduced by blunt-end ligation following cleavage with NsiI. Cells carrying the two plasmids were grown in BHIS media in the absence of antibiotic and then CFUs estimated on agar media supplemented with thiamphenicol after 6, 12 and 24 h of growth. The illustrated data was derived from three independent experiments.
doi:10.1371/journal.pone.0056051.g003

The sporulation minus phenotype of the mutants was confirmed by assaying colony forming units (cfu) on BHIS supplemented with 0.1% [w/v] sodium taurocholate before and after heat shock (65°C for 30 min), as previously described [20]. In each case, no colonies were obtained after heat shock, whereas the wild type controls gave counts of 1.28×10^4 and 9.17×10^4 cfu for CRG2548 (R20191Δ*pyrE*) and CRG2547 (630Δ*erm*Δ*pyrE*), respectively. Phase contrast microscopy confirmed that unlike the wild type strains, phase-bright spores were absent in the mutant cultures.

Creation and *in situ* complementation of in-frame deletion mutants of *cwp84*

To further test the system, we constructed mutants of the *cwp84* gene of both 630Δ*erm* and R20291. Cwp84 is a cysteine protease responsible for the post-translational cleavage of SlpA into the two proteins High Molecular Weight (HMW) and Low Molecular Weight (LMW) SlpA, which are the major components of the *C. difficile* surface layer. ClosTron mutants of *cwp84* have been made elsewhere in *C. difficile* 630Δ*erm* [21,22], and in the one study [22] complemented using an autonomous plasmid. *cwp84* in-frame deletion mutants were constructed following an identical procedure to that used to isolate the *spo0A* mutants, using plasmids pMTL-YN3::630*cwp84** and pMTL-YN4::R20291*cwp84**. One of the three 630Δ*erm* FOA resistant colonies screened yielded a 1,218 bp PCR product, indicative of the desired in-frame deletion (Supporting Information File S1). In the case of R20291, three of four FOA resistant colonies screened yielded the 1,218 bp PCR product, indicative of the desired in-frame deletion (Supporting Information File S1). Nucleotide sequencing of the 1,218 bp fragments confirmed that the desired in-frame deletion had been obtained.

A second round of *pyrE*-based allele exchange was carried out in the 630Δ*erm* and R20291 *cwp84* mutants, using pMTL-YN3::630*cwp84*′ and pMTL-YN4::R20291*cwp84*′, respectively, to restore the mutant alleles back to wild-type, *in situ*. A conservative base substitution was made to introduce a ScaI restriction site into the restored *cwp84* sequence (Figure 5A). Three 630Δ*erm* FOA resistant colonies were screened and yielded a 1.1 kb PCR product with primers cwp84-F4 and cwp84-R4, indicating that the *cwp84* allele had been successfully restored (Figure 5B). Subsequent digestion of the PCR products with ScaI confirmed that restoration of the *cwp84* allele had been successful in each of the three clones screened (Figure 5C). Similar analysis was carried out on nine R20291 FOA resistant clones. Successful restoration of *cwp84* was confirmed in each of the nine clones, as evidenced by a 1.1 kb PCR product which was cleaved by Sca I (data not shown).

Before carrying out any phenotypic characterisation, the *cwp84* mutants and their restored derivatives were converted back to a *pyrE*-positive (uracil prototrophy) phenotype using pMTL-YN1

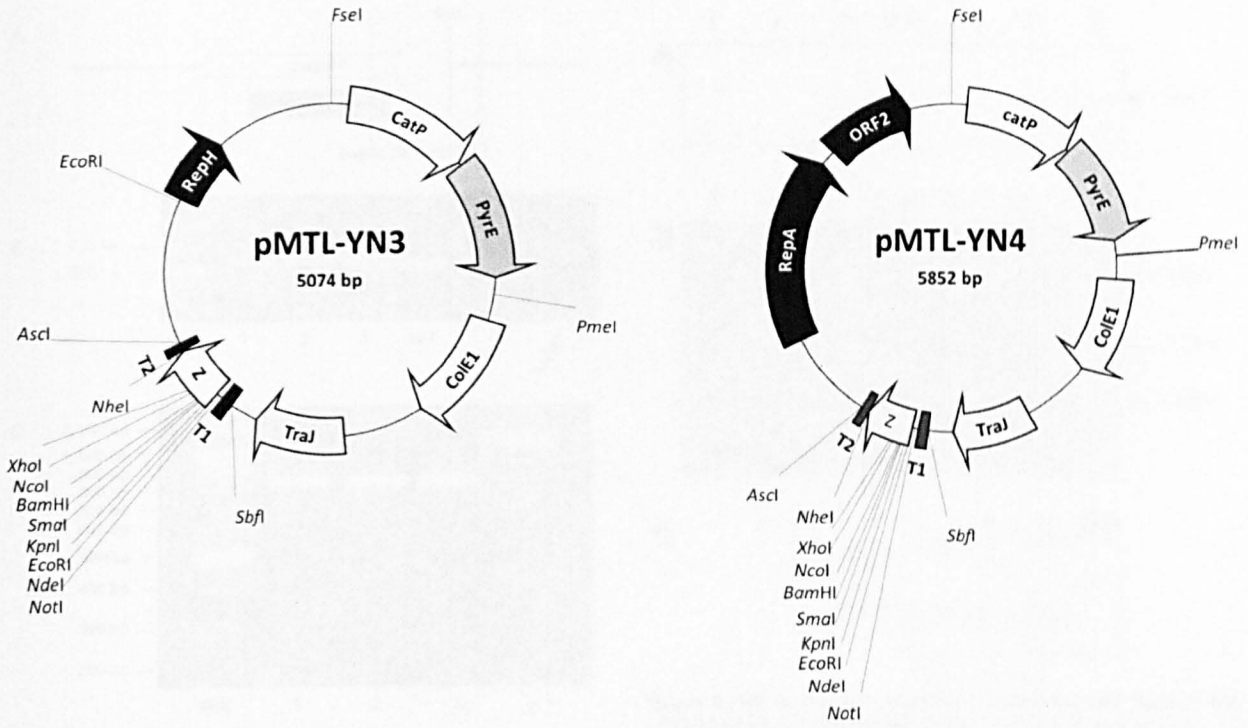


Figure 4. Allelic Exchange vectors for manipulation of *C. difficile* 630Δ*erm* (pMTL-YN3) and R20291 (pMTL-YN4). Common plasmid components are: CatP, the *catP* gene of *Clostridium perfringens* conferring thiamphenicol resistance; PyrE, the *pyrE* gene of *Clostridium sporogenes*; ColE1, the replication region of the *E.coli* plasmid ColE1, and; TraJ, transfer function of the RP4 *oriT* region; Z, the *lacZ'* gene encoding the alpha fragment of the *E.coli* β-galactosidase (and containing a multiple cloning site, MCS, region derived from plasmid pMTL20); T1, a transcriptional terminator isolated from downstream of the *Clostridium difficile* strain 630 CD0164 gene, and; T2, a transcriptional terminator of the ferredoxin gene of *Clostridium pasteurianum*. The position of the frame-shift generated at the NsiI site is indicated by an asterick. Plasmid pMTL-ME2 is identical to plasmid pMTL-YN3, except it carries an NsiI site at the 3'-end of RepH at the position marked by an asterick.
doi:10.1371/journal.pone.0056051.g004

and pMTL-YN2 for the 630Δ*erm*-derived strains and the R20291-derived strains, respectively. In keeping with previous observations [21,22], the mutants grew poorly on anaerobic blood agar compared to the wild-type and the complemented strains (Supporting Information File S1). S-layer extracts from the WT, mutant and complemented strains were prepared by the glycin acid method. The absence of the Cwp84 protease in the 630Δ*erm* and R20291 mutant strains was confirmed by immunoblot analysis with anti-Cwp84 antibodies (Figure 6A). Furthermore, while glycin extracts of both the 630Δ*erm* wild type strain and the complemented mutant contained two bands corresponding to the cleaved HMW and LMW S-layer proteins, only a band corresponding to the unprocessed SlpA precursor was evident in the *cwp84* mutant strain (Figure 6B). Interestingly, whilst broadly equivalent results were obtained with the R20291 strains, it was clear that the *cwp84* mutant contained two lower intensity bands, most likely corresponding to the HMW and LMW, in addition to the SplA precursor (Figure 6B). The identity of the HMW-SLP band was further confirmed by immunoblot (Figure 6C). This suggests that some proteolytic cleavage of SlpA may have occurred in the R20219 mutant, despite the absence of Cwp84.

Creation and rapid *in trans* complementation of in-frame deletion mutants

Whilst the *in situ* complementation of mutant alleles represents the 'gold standard', the procedure required is lengthy (approx. 3–4 weeks), and therefore adds to the time needed to characterise a mutant. An alternative strategy would be, having made a specific in-frame deletion in a *pyrE* strain, to use ACE to introduce a wild type copy of the gene into the chromosome at the *pyrE* locus concomitant with the correction of this allele back to a PyrE-positive phenotype. Accordingly, complementation vectors (pMTL-YN1C and pMTLYN2C, for strains 630 and R20291, respectively) were constructed (MATERIALS AND METHODS) especially for this purpose, together with derivatives (pMTL-YN1X and pMTLYN2X, for strains 630 and R20291, respectively) carrying a strong promoter (the *fdx* promoter) to direct the over expression of the complementing gene. In contrast to *in situ* complementation, and as a consequence of being able to directly select for uracil prototrophy without the need to isolate pure single crossover clones, the successful introduction of genes using these vectors takes between 5 and 7 days.

To test the utility of this approach, we targeted the *mtlD* gene which forms part of the operon responsible for mannitol metabolism in *C. difficile* strain 630Δ*erm*, using plasmid pMTL-YN3::630mtlD. The ability of *C. difficile* to ferment mannitol is a distinguishing feature and forms the basis of CDSA (*Clostridium difficile* Selective Agar) developed by Becton Dickinson [23]. Fermentation of mannitol in this medium causes a pH drop, which causes the indicator present to change in color, from red to yellow. In many bacteria mannitol is transported into the cell via a typical phosphotransferase system (PTS), and therefore couples transport to phosphorylation of the sugar [24]. The *mtl* operon in *E. coli* consists of the *mtlA*, *mtlR*, and *mtlD* genes that respectively encode

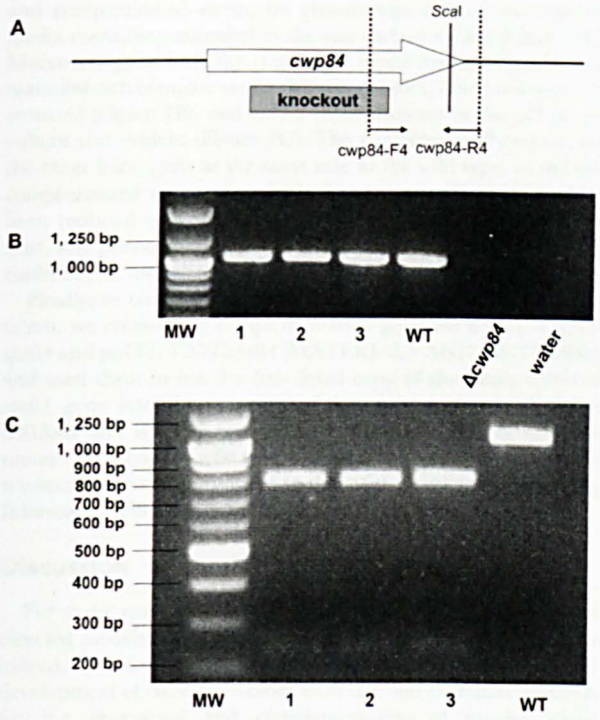


Figure 5. PCR screening of double crossover candidate clones for complementation of the *cwp84* gene in *C. difficile* 630 Δ*erm* Δ*cwp84*. (A) Schematic diagram of the complementation of *cwp84*, with a single nucleotide change to base 2280 of *cwp84* from a T to an A, without changing the corresponding valine amino acid residue and at the same time creating a *ScaI* site. The purpose of this single nucleotide change was to prove the occurrence of the complementation event. (B) PCR screening of candidate clones of the complemented *cwp84* gene. Primers cwp84-F4 and cwp84-R4 anneal to the internal sequence of the knockout cassette and the downstream sequence of *cwp84*, respectively, resulting in a 1, 026 bp PCR product from double-crossover complemented clones and wild-type, while no PCR product is expected from Δ*cwp84* mutants. MW is a 2-Log DNA Ladder (NEB) molecular weight marker, WT is a wild-type *C. difficile* DNA control, and 1–3 are the candidate clones. All candidates 1 to 3 show the expected complemented 1, 026 bp band, thereby confirmed as *cwp84* complemented clones, as seen in the wildtype control. (C) PCR products amplified using primers cwp84-F4 and cwp84-R4 from candidates clones and wildtype were analysed by RE digestion with *ScaI*. PCR products amplified from cwp84 complemented clones were cut into two fragments (786 and 240 bp), whereas PCR products amplified from the wildtype control did not.
doi:10.1371/journal.pone.0056051.g005

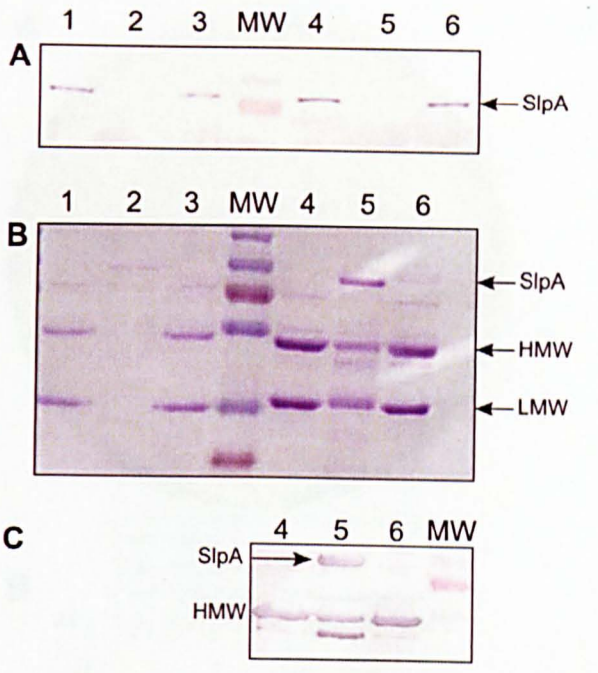


Figure 6. Glycin extracts analysis of 630Δerm and R20291 WT, mutant and complemented strains. (A) Immunoblot analysis with anti-Cwp84 antibodies of glycin extracts, showing complete absence of Cwp84 in the mutants compared to WT and complemented strains. (B) SDS-PAGE of glycin extracts of 630Δerm and R20291 WT, mutant and complemented strains, showing no processing of SlpA precursor in the 630Δ*cwp84* mutant, and in contrast, an incomplete processing of SlpA in the R20291Δ*cwp84* mutant. (C) Identification of the HMW-SLP in the glycin extract of the R20291Δ*cwp84*, showing that a partial processing of SlpA takes place in this mutant even in absence of the Cwp84 protease. Lanes 1, 630Δ*erm*; lanes 2, 630Δ*erm*Δ*cwp84*; lanes 3, 630Δ*erm*Δ*cwp84* complemented; lanes 4, R20291; lanes 5, R20291Δ*erm*Δ*cwp84*; lanes 6, R20291Δ*erm*Δ*cwp84* complemented; MW, molecular weight standard.
doi:10.1371/journal.pone.0056051.g006

the mannitol transporter (enzyme IICBA*mtl*), a transcriptional regulator, and mannitol-1-phosphate dehydrogenase [25]. The arrangement of genes in *C. difficile* 630Δ*erm* is predicted to be equivalent to that of *B. stearothermophilus* and comprises four putative genes: *mtlA* (enzyme IICB*mtl*), *mtlR*, *mtlF* (enzyme IIA*mtl*), and *mtlD* [26,27].

Following transfer of the knock-out plasmid pMTL-YN3::630*mtlD* into *C. difficile* 630Δ*erm*, three FOA resistant clones were isolated. PCR screening using primers mtlD-F3 and mtlD-R3, that anneal to the upstream and the downstream sequence of *mtlD*, demonstrated that two clones were Δ*mtlD* mutants, yielding a PCR product of 1,418 bp, while the third clone was a wild type revertant, yielding a PCR product of 2,582 bp (Supporting Information File S1). Nucleotide sequencing of the 1,418 bp products confirmed that the desired *mtlD* in-frame deletion had been created.

On isolation of the *mtlD* mutant, the *pyrE*$^-$ gene was converted back to *pyrE*$^+$ using plasmid pMTL-YN1 as described above. In addition, two further derivatives of pMTL-YN1 were employed which simultaneously delivered a functional copy of the *mtlD* wild type gene into the genome immediately downstream of the corrected *pyrE* gene. Plasmid pMTL-YN1C*mtlD* introduced the *mtlD* structural gene into the chromosome together with its native promoter. In contrast, pMTL-YN1X*mtlD* placed the inserted gene under the control of the strong *fdx* promoter. In both instances, the inserted *mtlD* gene was followed by the transcriptional terminator sequence [6] derived from the ferredoxin gene of *Clostridium pasteurianum*. The purpose of the terminator was to prevent polar effects being exerted by the inserted gene on the downstream gene, CD0188. In both cases, six independent PyrE+ clones able to grow on minimal media in the absence of uracil were screened for the presence of the wild type allele by PCR using primers Cdi630-pyrD-sF1 and Cdi630-CD0189-SR3. All twelve clones gave the expected 2,058 bp DNA product, equivalent to a wild type control. Nucleotide sequencing confirmed its identity as a wild type *pyrE* allele.

Phenotypic analysis confirmed that the *mtlD* mutant was no longer able to ferment mannitol. Thus, in contrast to the wild type

and complemented strain, no growth was evident on minimal media containing mannitol as the sole carbon source (Figure 7A). Moreover, growth of the mutant in liquid media comprising a mannitol-rich complex media (BD Diagnostics, USA) was severely retarded (Figure 7B), and only a slight decrease in the pH of the culture was evident (Figure 7C). The complemented mutant, on the other hand grew at the same rate as the wild type, as did the complemented mutant in which the native *mtlD* promoter had been replaced with the strong *fdx* promoter. In this instance, at least, any phenotypic benefit of increasing the expression of *mtlD* could not be measured.

Finally, to confirm the utility of this approach to complementation, we created the complementation plasmids pMTL-YN1C-*spo0A* and pMTL-YN2C*spo0A* (MATERIALS AND METHODS) and used them to insert a functional copy of the strain 630Δ*erm* *spo0A* gene into the genomes of the *spo0A* mutants of strains 630Δ*erm* and R20291 concomitant with restoration of the *pyrE* mutant allele back to wild type. The sporulation phenotypes of the resultant strains were restored to that of the wildtype (Supporting Information File S1).

Discussion

For many years the gene tools available to bring about specific, directed modifications of *C. difficile* at the genetic level were scant indeed. In the present study we have capitalised on the development of ACE technology to devise and exemplify methods for the generation and complementation of precise allelic exchange mutants in the *C. difficile* genome based on the use of various *pyrE* alleles. The use of genes involved in uracil metabolism, such as *pyrE* and *pyrF*, as a negative-/counter-selection marker has previously been demonstrated in a number of different bacteria and yeast [10–15], where it is reliant on the initial creation of a mutant defective in the chosen gene to act as the host for mutant construction. The deployment of ACE technology ensures this is a relatively simple and rapid undertaking. More important, as demonstrated here, ACE may be used to rapidly convert the host back to $pyrE^{+}$, allowing the assessment of the mutant phenotype in a 'clean', wild type background. Such an undertaken has not generally been applied in other systems. Thus, for instance, the mutants created in *C. thermocellum* [15] using the organisms own *pyrF* gene as a counter-selection marker were always analysed in the Δ*pyrF* strain in which they were created. Similarly, a mutagenesis system developed in *Bacillus subtilis*, based on the use of the *upp* gene [28], has been adapted for use in *C. acetobutylicum* (P Soucaille, pers comm.) and is always undertaken, and the resultant mutants made analysed in, a *upp*- background. The use of a background defective in nucleoside metabolism can complicate the assessment of the effect of the mutation under analysis. This is particularly the case in the investigation of pathogens and virulence, where mutants in nucleoside metabolism are considered a form of disablement in their own right [29]. This has considerable impact on studies designed to assess the effect of mutation of other loci on virulence, particularly *in vivo*.

We have demonstrated the robustness and reliability of this *pyrE* method of mutagenesis in both *C. difficile* 630 Δ*erm* and *C. difficile* R20291, through the creation of in-frame deletion mutants in *spo0A*, *cwp84* and *mtlD*. We have re-affirmed the value of strategies based on pseudo-suicide vectors, and in the case of 630Δ*erm* improved the options available through the further disablement of the pCB102 replicon used by the introduction of a frameshift in the *repH* gene. This was found to facilitate the ease with which single crossover integrants were isolated, which tend to emerge after just two passages of transconjugants on selective media. This

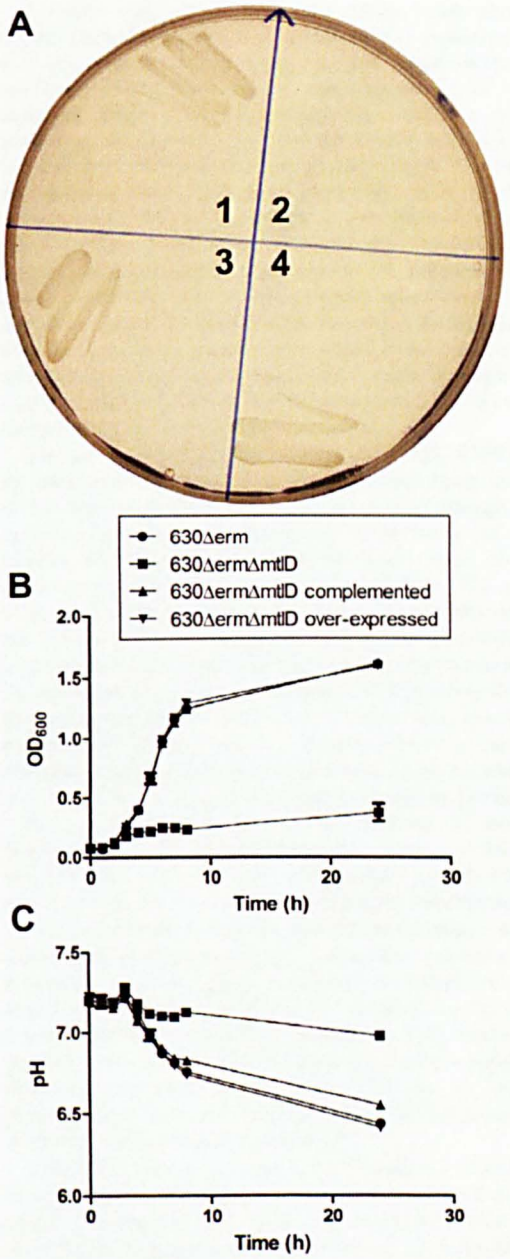


Figure 7. Growth of *C. difficile* 630Δerm strains with mannitol as the sole carbon source. (A) Clock-wise from top-left, *C. difficile* 630Δ*erm* (1) 630Δ*erm*Δ*mtlD* mutant (2), and 630Δ*erm*Δ*mtlD*-complemented (3) and 630Δ*erm*Δ*mtlD*-overexpressed (4) were streaked onto minimal media agar with mannitol as the sole carbon source and incubated for 48 h to observe growth. In contrast to the wild type, complemented and overexpressed strains, no growth was evident for the 630ΔermΔ *mtlD* mutant. (B) The growth of Δ*mtlD* was limited in mannitol broth, while growth of the Δ*mtlD* complemented and *mtlD* overexpressed strains were restored to wildtype levels. (C) The pH of the media broth showed a dip in pH caused by the fermentation of mannitol for the wildtype, Δ*mtlD* complemented and Δ*mtlD* overexpressed strains, which correlate to their growth. The 630 Δ*erm* Δ*mtlD* mutant strain grew very weakly in mannitol broth, which was reflected in the sustained pH levels of the media. All experiments were undertaken in triplicate. The data, complete with error bars is provided in the Supporting Information File S1.
doi:10.1371/journal.pone.0056051.g007

is in contrast to strain R20291 allelic exchange vectors, which make use of the pBP1 vector. During the course of this work we have confirmed that the annotated *mtlD* locus is indeed involved in mannitol utilisation and re-affirmed that Cwp84 plays a role in the processing of SlpA. Intriguingly, we found that despite the introduction of the *cwp84* frame-shift in R20291, unlike 630Δ*erm*, some processing of SlpA was still occurring in this strain. The in-frame deletion of both strains results in an internal deletion of 406 amino acids from the protein (total 804 amino acids). Whilst the truncated protein that results still contains Cys116, an important residue in the active site of the protease [22], it seems unlikely that the protein retains proteolytic activity, particularly as the truncated protein in the 630Δ*erm* mutant is identical. Moreover, no signal was detected by immunoblot analysis in the two mutant strains, suggesting that there is no expression at all of the protease, even in a truncated form, or that this form is quickly degraded in the cell. Perhaps strain R20291 produces a secondary protease also able to process SlpA? At this stage no one explanation can be ruled out. Here our focus has been on the development and description of the mutagenesis method, and our data on individual mutants remains observational.

The method developed here is in many respects equivalent to our recent description of the use of *codA* as a counter selection [8]. Indeed, the allelic exchange cassettes and methodological procedures are the same, except counter-selection is based on differential resistance to FOA (PyrE) rather than FC (CodA). Accordingly, the isolation of mutants using either protocol takes around 3–4 weeks to accomplish once the allelic exchange vectors have been constructed, and both methods may be subsequently employed to undertake *in situ* complementation of the mutated allele (exchange of the mutant allele with the wild type gene). A pivotal difference is that *codA*-based mutagenesis [8] can be undertaken in a wild type background (provided the host does not carry a native *codA* gene, as is the case in *C. botulinum*, *C. sporogenes* and *C. ljungdahlii*), whereas the method described here requires the initial creation of a *pyrE* mutant which needs to be corrected back to wild type once a mutant at a distal locus has been created. However, using ACE, this is rapid (5 days), and provides the opportunity to complement the mutant in parallel, through concomitant insertion of a wild type copy of the gene either under the control of its native promoter, or the strong *fdx* promoter.

Complementation is essential to unequivocally confirm the role of a genetic determinant in a specific phenotype [30]. Complementation is traditionally achieved using multicopy plasmids. The consequent abnormal gene dosage can, however, frequently fail to restore phenotype to wildtype levels. Thus, CwpV was over produced in a plasmid complemented *cwpV* mutant of *C. difficile* [31], the perfringolysin O titre of a *virR* mutant of the *C. perfringens* strain 56 was approximately 3-fold higher than the wild type strain when complemented with a multicopy plasmid carrying the *virRS* genes [32], while in a study of *Clostridium septicum* 'haemolytic activity of the complemented strain was higher than the parent strain ..... presumably a reflection of the presence of the *csa* gene on a multicopy plasmid.' [33]. In the study of Dingle and co-workers [34], 'although the wild-type 630Δ*erm* strain and the *fliC* mutant strain produced FliD, the *fliC*-complemented strain did not, suggesting that there was a reduction in *fliD* gene expression when FliC is expressed in trans off a plasmid rather than in *cis* from the chromosomal gene'. Similarly, plasmid complemented *agr* mutants of *Clostridium perfringens* failed to completely regain wild-type sporulation levels [35], a phenomenon attributed by the authors to 'complementation involving a multicopy plasmid carrying the cloned *agr* locus'.

Ancillary effects can also arise when using plasmids for complementation studies due to the need to maintain plasmids, and the complementing gene, through supplementation with antibiotics. Thus, for instance, complementation of the a *sleC* mutation using a plasmid-located gene did not restore the phenotype to wild type [20], but did restore germination to the reduced level obtained in a vector only control. The presence of the plasmid alone, and most likely the effect of the added thiamphenicol needed to maintain the plasmid, detrimentally effected spore germination. Consequently, the use of plasmid complementation systems may require the inclusion of a vector only control [20]. This is not required when using ACE-based complementation at the *pyrE* locus. Crucially, the intrinsic stability of chromosomal complementation makes it the method of choice in those situations where plasmid maintenance through antibiotic supplementation is not a viable option, such as when complementation in an *in vivo* model is required.

The use of plasmids pMTL-YN1C and pMTL-YN2C positions the complementing gene in the genome immediately downstream of the *pyrE* gene (CD0187). This allows read through from the upstream promoter responsible for transcription of *pyrE*, the identity of which is unknown but may reside immediately upstream of CD0187 or, more likely, CD0184. The availability of transcriptional read through is important in those cases where the identity of the natural promoter of the complementation gene is not defined, and cannot be guaranteed to be cloned along with the structural gene, allowing, for instance, a promoter-less copy of the structural gene to be cloned. In those instances where the promoter is defined, and the operator wishes to exclude read through, a transcriptional terminator may simply be added during the cloning of the complementation gene and its promoter.

In certain instances it may be desirable to bring about 'overexpression' of the complementing gene. For this purpose, two further vectors were constructed (pMTL-YN1X and pMTLYN2X, for strains 630 and R20291, respectively) carrying the promoter of the ferredoxin gene (*fdx*) of *Clostridium sporogenes*. As a consequence of its central role in anaerobic electron transfer, the ferredoxin gene is highly expressed in *Clostridium*, with the ferredoxin protein representing, for instance, up to 2% of the soluble protein in *Clostridium pasteurianum* [36]. Indeed the *fdx* promoter of *C. perfringens* has been used to derive a highly effective clostridial expression system [37], while the *C. sporogenes fdx* promoter used here was employed to maximise expression of a nitroreductase gene in *C. sporogenes* [38].

Whichever vectors are used, their deployment involves only marginal additional effort compared to the use of autonomous complementation vectors. Thus, both methods (plasmid-based and ACE-based complementation) involve the assembly of the requisite vector in *E. coli* and its transformation into the appropriate clostridial mutant and the selection of transconjugants on antibiotic supplemented plates. ACE-mediated complementation merely requires re-streaking of the primary transconjugant onto minimal media lacking uracil. All those colonies that develop represent complemented strains in which the *pyrE* allele has been corrected to wildtype. The additional time required, therefore, equates to merely the 2 days needed for uracil prototrophic colonies to develop on the minimal media.

Conclusion

We have demonstrated the use of a *pyrE*-based allele exchange system to make precise alterations to the genome of *C. difficile*. In using a *pyrE*- strain as the host, by whatever means a mutant is made (e.g., allelic exchange or ClosTron), the facility to insert a wild type copy of the gene at the *pyrE* locus provides a rapid

mechanism for generating a stable complementation clone that is not compromised by inappropriate gene dosage effects. As such, the *pyrE* mutants made here are available to the scientific community for mutagenesis studies, along with the requisite *pyrE* correction vectors. The genomes of both strains have been subjected to genome re-sequencing and found to be devoid of any additional SNPs or Indels compared to the progenitor 630Δ*erm* strain or R20291 parental strains.

Materials and Methods

Bacterial strains and growth conditions

Escherichia coli TOP10 (Invitrogen) and *E. coli* CA434 [39] were cultured in Luria-Bertani (LB) medium, supplemented with chloramphenicol (25 μg/ml), where appropriate. Routine cultures of *C. difficile* 630 Δ*erm* [40] and *C. difficile* R20291 were carried out in BHIS medium (brain heart infusion medium supplemented with 5 mg/ml yeast extract and 0.1% [wt/vol] L-cysteine) [41]. *C. difficile* medium was supplemented with D-cycloserine (250 μg/ml), cefoxitin (8 μg/ml), lincomycin (20 μg/ml), and/or thiamphenicol (15 μg/ml) where appropriate. A defined minimal media [18] was used as uracil-free medium when performing genetic selections. A basic nutritive mannitol broth for growth assays of *C. difficile* strains were prepared as follows : Proteose peptone no. 2 4% [wt/vol] (BD Diagnostics, USA), sodium phosphate dibasic 0.5%[wt/vol], potassium phosphate monobasic 0.1%[wt/vol], sodium chloride, 0.2% [wt/vol], magnesium sulfate, 0.01% [wt/vol], mannitol, 0.6% [wt/vol] with final pH at +/−7.35. For solid medium, agar was added to a final concentration of 1.0% (wt/vol). *Clostridium sporogenes* ATCC 15579 was cultivated in TYG media [7]. All *Clostridium* cultures were incubated in an anaerobic workstation at 37°C (Don Whitley, Yorkshire, United Kingdom). Uracil was added at 5 μg/ml, and 5-Fluoroorotic acid (5-FOA) at 2 mg/ml. All reagents, unless noted, were purchased from Sigma-Aldrich.

Conjugation into *C. difficile*

Plasmids were transformed into *E. coli* donor, CA434 and then conjugated into *C. difficile* [42]. Thiamphenicol (15 μg/ml) was used to select for *catP*-based plasmid. D-cycloserine (250 μg/ml) and cefoxitin (8 μg/ml) were used to counter select against *E. coli* CA434 after conjugation.

DNA manipulations

DNA manipulations were carried out according to standard techniques. *C. difficile* genomic DNA for use in cloning and PCR analysis was prepared as described previously [9]. Polymerase chain reactions (PCR) used Failsafe™ PCR system (Epicentre) and Taq polymerase (Promega) in accordance with the manufacturers' protocols using primers detailed in Table 1. Primer design for amplification of DNA from *C. difficile* 630 Δ*erm* and R20291 strains were based on the available *C. difficile* genomes from the EMBL/GenBank databases with accession numbers AM180356 and FN545816 respectively. The oligonucleotides and the plasmids/strains used in this study are listed in Table 1 and Table 2, respectively. All DNA manipulations and cloning procedures were performed as per Sambrook [43].

Vector Construction

Plasmid pMTL-YN18 (accession number; JX465728), used to create a *pyrE*⁻ mutant in R20291, is an equivalent vector to pMTL-JH18 [9] in which the pCB102 replicon was replaced with that of pBP1 (isolated from pMTL82251), using the flanking AscI and FseI sites specifically created for this purpose [17]. Plasmid pMTL-YN1 (accession number; JX465729) was made by cloning a 1,753 bp SbfI/AscI fragment between the corresponding restriction sites of plasmid pMTL-JH20 [9]. This fragment was generated from the 630Δ*erm* genome by PCR using the primers pyrE-int-SbfI-F and pyrEcomplement-AscI-R, and carries a truncated copy of *pyrE*, lacking the first 50 nucleotides of the structural gene, together with 1200 bp from immediately 3′ to the stop codon of *pyrE*, which includes the entire coding sequence of CD0188 of 630Δ*erm*. To generate the equivalent *pyrE* correction vector (pMTL-YN2, accession number; JX465732) for R20291, the pCB102 replicon of pMTL-YN1 was replaced with that of pBP1 from pMTL82251 (using the AscI and FseI sites) and the 630Δ*erm* DNA between the SbfI and AscI sites replaced with the equivalent region from the R20291 genome. In this case, the DNA was synthesized by DNA2.0 specific to the R20291 chromosome by PCR as two contiguous SbfI/NdeI and NdeI/AscI fragments. The PyrE complementation vector pMTL-ME2 was constructed by inserting the *C. sporogenes* ATCC 15579 *pyrE* gene into pMTL83151. The *pyrE* gene was amplified by PCR from genomic DNA prepared from *Clostridium* sporogenes ATCC 15579 using primers Csp-pyrE-*Hpa*I-sF1 and Csp-pyrE-*Hpa*I-sR1. The PCR product was digested with HpaI and gel purified followed by cloning in same orientation as *catP* in pMTL83151 [17] linearized with the same enzyme.

To construct the ACE overexpression plasmids pMTL-YN1X (accession number; JX465731) and pMTL-YN2X (accession number; JX465734), a 214 bp NotI/NdeI fragment encompassing the *C. sporogenes fdx* promoter was isolated from plasmid pMTL82253 and inserted between the equivalent sites of pMTL-JH14 [9]. The plasmid obtained (pMTL-JH14::Pcspfdx) was cleaved with NheI, and ligated to two, annealed oligonucleotides (Cpa-TT-Oligo-1 and Cpa-TT-Oligo-2) which encompassed the transcriptional terminator of the *Clostridium pasteurianum* ferredoxin gene. A 517 bp NotI/NheI fragment (encompassing the *fdx* promoter, *lacZ*′ and multiple cloning sites, and the transcriptional terminator of the *C. pasteurianum* ferredoxin gene) was then isolated from the resultant plasmid (pMTL-ME6X) and inserted between the equivalent sites in pMTL-YN1 and pMTLYN2, to yield pMTL-YN1X and pMTL-YN2X, respectively. In parallel, pMTL-ME6X was digested with EcoRI and NheI and the released 293 bp fragment inserted between the equivalent sites of pMTL-JH14 [9] to make pMTL-ME6C. This plasmid was then cleaved with NotI and NheI, and the released 323 bp fragment inserted between the equivalent sites of pMTL-YN1 and pMTLYN2, to yield the complementation vectors pMTL-YN1C (accession number; JX465730) and pMTL-YN2C (accession number; JX465733), respectively.

Allelic Exchange Cassettes

Allele exchange cassettes were assembled, specific to the three target loci (*spo0A*, *cwp84* and *mtlD*) of the two strains, composed of a left-hand homology arm (LHA) and a right-hand homology arm (RHA), each of approximately 500 to 800 bp in size. Each cassette was cloned between the *Sbf*I and *Asc*I restriction recognition sites in pMTL-YN3 (accession number; JX465735) to create pMTL-YN3::630spo0A*, pMTL-YN3::630cwp84* and pMTL-YN3::630mtlD*. Similarly, pMTL-YN3::R20291spo0A* and pMTL-YN3::R20291cwp84* were made by cloning the allelic exchange cassettes between the *Asc*I and *Sbf*I sites in pMTL-YN4 (accession number; JX465736).

The *spo0A* knockout cassette was synthesized by DNA2.0 which consists of a 672 bp LHA (bases 1412012 to 1412789 on the forward strand of *C. difficile* 630 Δ*erm* genome) fused to a 800 bp RHA (bases 1413275 to 1414074 on the forward strand of *C. difficile* 630 Δ*erm* genome) which were designed to make an in-

Table 1. List of oligonucleotides used in this study.

Oligonucleotide	Binding site	Sequence (5′ to 3′)
Csp-pyrE-*Hpa*I-sF1	*pyrE* of C.*sporogenes*	AATATTGTTAACTAAGGAGAAGATATAAATGAGTAATATAAAT GTTATAGATATATTAAAAGAATCAAAT
Csp-pyrE-*Hpa*I-sR1	*pyrE* of C.*sporogenes*	AATATGTTAACTTATTTTTGTTCTCTACTACCTGGTTTTACAAAAGGT
Cdi630-pyrD-sF1	*pyrD* of C. *difficile*	AGAGAAGGAATAAAAAGTTTAGACGAAATAAGAGG
λ6.5-sF2	6.5 kbp HindIII fragment of phage lambda, insert-specific reverse primer	TATGAGTACCCTGTTTTTTCTCATGTTCAGG
Cdi630-CD0189-sR3	CD0189 of C. *difficile*,chromosome-specific reverse primer	CCAAGCTCTATGACAGACAGCTCATTGTTTAGAAC
pyrE-int-SbfI-F	*pyrE* of C. *difficile* 630 Δ*erm* and R20291	CTGCAGGGGAGGGACATTTTTTATTATCTTCAGG
pyrEcomplement-Asc1-R	CD0188 of C. *difficile* 630 Δ*erm* and R20291	GGCGCGCCATAGTATATAACATTAATAAAATTTAAAATC
spo0A-YN-F2	upstream of *spo0A* of C. *difficile*	GGCAAGTATAAACTTGGATTATGGGTAAGAGAT
spo0A-YN-R2	downstream of *spo0A* of C. *difficile*	CTATATATCTTTCCATCTACAACTTCTATAG
cwp84-F3	upstream of *cwp84* of C. *difficile*	TTCTATAATTAATATGTACTCATAATCC
Cwp84-F4	knockout region of *cwp84*	CTGGACAAGCTACTTCAGGAG
Cwp84-R4	downstream of *cwp84* of C. *difficile*	CTGGACAAGCTACTTCAGGAG
mtlD-F3	upstream of *mtlD* of C. *difficile*	CTAGAGAATAGAATCGTGCTAGATTCAAATGAAG
mtlD-R3	downstream of *mtlD* of C. *difficile*	CTTTAACTGAATACTCTCTTGCCTTAG
mtlD-NotI-F	upstream of *mtlD*, inclusive of RBS and promoter sequence	GCGGCCGCTTTTTAATCACTCCTTATATTTTTATAC
mtlD-BamHI-R	end of coding sequence of *mtlD*	GGATCCTTATAAATTTTTCATAAATATATAACTTTTTTCGATAT TATTTAAAAGTTCTTCG
mtlD-NdeI-F	start of *mtlD* coding sequence	CATATGAAAAAGGCAATTCAGTTTGGAGCAGG
Cpa-TT-Oligo-1	transcriptional terminator of the *Clostridium pasteurianum* ferredoxin gene	CTAGTATAAAAATAAGAAGCCTGCATTTGCAGGCTTCTTATTTTTATG
Cpa-TT-Oligo-2	transcriptional terminator of the *Clostridium pasteurianum* ferredoxin gene	CTAGCATAAAAATAAGAAGCCTGCAAATGCAGGCTTCTTATTTTTATA
spo0A-SpeI-F	*spo0A* gene of *C. difficile*	ATATACTAGTGGTATTTTTATAGATGAAATGATAAAATTGTAG
spo0A-BamHI-R	*spo0A* gene of *C. difficile*	ATATGGATCCTCAGTTTACAACTTGTAAAGACAC

doi:10.1371/journal.pone.0056051.t001

frame deletion removing codons 65 to 226, inclusive, out of a total of 275 codons). For the equivalent in *C. difficile* R20291, a 778 bp LHA (bases 1319650 to 1320427 on the forward strand of the *C. difficile* R20291 genome) was fused to a 800 bp RHA (bases 1320913 to 1321712 on the forward strand of the *C. difficile* R20291 genome). Each cassette was designed to incorporate a *Sbf*I site at the 5′ end and an *Asc*I site at the 3′ end to facilitate cloning into the pMTL-YN3 plasmid. The *cwp84* knockout cassette consists of a 653 bp LHA (bases 1045907 to 1046559 on the reverse strand of the *C. difficile* 630 Δ*erm* genome) fused to a 558 bp RHA (bases 1047778 to 1048335 on the reverse strand of the *C. difficile* 630 Δ*erm* genome) which were designed to remove codons 214 to 629, inclusive, representing the removal of 417 codons from a total of 814). For the equivalent in *C. difficile* R20291, a 713 bp LHA (bases 1067119 to 1067771 on the reverse strand of the *C. difficile* R20291 genome) was fused to a 558 bp RHA (bases 1068990 to 1069547 on the reverse strand of the *C. difficile* R20291 genome). The *mtlD* knockout cassette consists of a 654 bp LHA (bases 1593564 to 1594217 on the reverse strand of the *C. difficile* 630 Δ*erm* genome) fused to a 654 bp RHA (bases 1595353 to 1596006 on the reverse strand of the *C. difficile* 630 Δ*erm* genome) which were designed to make an in-frame deletion between codon 3 and codon 387 of *mtlD* (total codons 388).

Complementation Vectors

Plasmids for the complementation of the *spo0A* mutants of strains R20291 and 630Δ*erm* were constructed as follows. The *spo0A* of strain *C. difficile* 630Δ*erm* was PCR amplified as a 1216 bp DNA fragment using the two primers spo0A-SpeI-F and spo0A-BamHI-R and cloned in pCR2.1 vector (Invitrogen) and verified by sequencing. The *spo0A* was re-isolated as a 1.2 kb EcoRI and BamHI fragment and inserted between the equivalent sites of pMTL-YN2C to give the R20291 complementation plasmid pMTL-YN2C*spo0A*. To derive an equivalent plasmid for strain 630Δ*erm*, the *spo0A* gene was excised from pMTL-YN2C*spo0A* as a 1.2 kb DNA fragment following cleaved with NotI and BamHI and inserted into the appropriate sites of pMTL-YN1C to give the complementation plasmid pMTL-YN1C*spo0A*.

To generate the two required allelic exchange cassettes need for complementation of the *cwp84* in-frame deletions, the portion of *cwp84* missing in the two plasmids (pMTL-YN4::R20291*cwp84** and pMTL-YN3::630*cwp84**) originally used to make the in-frame deletions were synthesised (DNA2.0) as a 1,082 bp NdeI-KpnI fragment (R20291) and a 1,255 bp NdeI-XbaI fragment (630Δ*erm*) generated by cloned into the appropriate allelic exchange vectors to yield plasmids pMTL-YN4::R20291*cwp84′* and pMTL-YN3::630*cwp84′*. In order to be able to distinguish the complementation allele from wild type, the synthesised region included a single nucleotide change (T to A) to base 2280 of *cwp84* which retains the codon specificity (codon 760, Valine) but creates a ScaI restriction enzyme recognition site.

The complementation plasmid pMTL-YN1C*mtlD* was constructed by the PCR amplification of the *C. difficile* 630Δ*erm mtlD* gene, together with its 5′ non-coding region, as a 1,779 bp

Table 2. List of strains and plasmids used in this study.

Strains/Plasmids	Relevant features	Source
E. coli Top10	*mcrA* Δ(*mrr-hsdRMS-mcrBC*) *recA1*	Invitrogen
E. coli CA434	*hsd20*(r^B-, m^B-, *recA13, rpsL20, leu, proA2*, with IncPβ conjugative plasmid R702	[31]
C. difficile 630Δ*erm*	Erythromycin sensitive strain of strain 630	[34]
CRG1496	*C. difficile* 630Δ*erm* Δ*pyrE*	This study
CRG2547	Derived from CRG1496, in-frame deletion within *spo0A*, *pyrE* repaired to wildtype sequence	This study
CRG2302	Derived from CRG1496, in-frame deletion within *cwp84*, *pyrE* repaired to wildtype sequence	This study
CRG2445	Derived from CRG2302, *in situ* complementation of Δ*cwp84* with a single nucleotide change, *pyrE* repaired to wildtype sequence	This study
CRG2923	Derived from CRG1496, in-frame deletion within *mtlD*, *pyrE* repaired to wildtype sequence	This study
CRG2536	*C. difficile* 630_*erm mtlD*::intron *ermB*	This study
CRG2926	Derived from CRG2923, chromosomal complementation of *mtlD* with its natural RBS and promoter region, downstream of *pyrE*	This study
CRG2929	Derived from CRG2923, chromosomal complementation of *mtlD* preceded by the promoter region (P_{fdx}) of the *Clostridium sporogenes* ferredoxin gene, downstream of *pyrE*	This study
C. difficile R20291 wild-type	BI/NAP1/027 Stoke Mandeville (2004–2005) isolate	Anaerobe Reference Laboratory, Cardiff, Wales, United Kingdom
CRG2359	*C. difficile* R20291 Δ*pyrE*	This study
CRG2548	Derived from CRG2359, in-frame deletion within *spo0A*, *pyrE* repaired to wildtype sequence	This study
CRG2549	Derived from CRG2359, in-frame deletion within *cwp84*, *pyrE* repaired to wildtype sequence	This study
CRG3059	Derived from CRG2549, *in situ* complementation of Δ*cwp84* with a single nucleotide change, *pyrE* repaired to wildtype sequence	This study
pMTL-YN18	same as pMTLJH18::λ6.5, but without the λ6.5 fragment, homology arms are specific to R20291 sequence and pCB102 replicon replaced with pBP1	This study
pMTL-ME2	Derived from pMTL83151 through addition of a *C. sporogenes pyrE* and *C. perfringens catP* gene. Carries an unaltered pCB102 replicon, which includes a NsiI site in the *repH* gene.	This study
pMTL-YN3	Derived from pMTL-ME2 by blunt-end ligation of the NsiI site within *repH*, causing a frame-shift in RepH	This study
pMTL-YN3::630spo0A*	Same as pMTL-YN3, *spo0A* KO cassette cloned into SbfI/AscI sites	This study
pMTL-YN3::630cwp84*	Same as pMTL-YN3, *cwp84* KO cassette cloned into SbfI/AscI sites	This study
pMTL-YN3::630cwp84c	Same as pMTL-YN3, full sequence of cwp84 with single base change cloned into SbfI/AscI sites	This study
pMTL-YN4	Derived from pMTL-YN3 by replacing the pCB102 replicon with that of pBP1	This study
pMTL-YN4::R20291spo0A*	Same as pMTL-YN4, *spo0A* KO cassette cloned into SbfI/AscI sites	This study
pMTL-YN4::R20291cwp84*	Same as pMTL-YN4, *cwp84* KO cassette cloned into SbfI/AscI sites	This study
pMTL-YN4::R20291cwp84c	Same as pMTL-YN4, full sequence of *cwp84* with single base change cloned into SbfI/AscI sites	This study
pMTL-YN3::630mtlD*	Same as pMTL-YN3, *mtlD* KO cassette cloned into SbfI/AscI sites	This study
pMTL-YN1	same backbone as JH-18 but without the 6.5kb fragment of λphage DNA, entire *pyrE* coding sequence minus the first 50 bases	This study
pMTL-YN1C	same as pMTL-YN1, except for a MCS inserted after *pyrE* gene to enable cloning of target gene for purpose of complementation	This study
pMTL-YN1C::mtlD	same as pMTL-YN1C, with the full coding sequence of mtlD and its natural RBS cloned into the *NdeI* and *BamHI* sites	This study
pMTL-YN1X	same as pMTL-YN1C, for a *fdx* promoter inserted after *pyrE* gene and before the MCS to enable overexpression of target gene	This study
pMTL-YN1X::mtlD	same as pMTL-YN1C::mtlD, with the full coding sequence of *mtlD* cloned downstream of the fdx promoter into the *NdeI* and *BamHI* sites	This study
pMTL-YN2	same as pMTL-YN1, but with homology arms specific to R20291 sequence and pCB102 replicon replaced with pBP1	This study
pMTL-YN2C	same as pMTL-YN1C, but with homology arms specific to R20291 sequence and pCB102 replicon replaced with pBP1	This study
pMTL-YN2X	same as pMTL-YN1X, but with homology arms specific to R20291 sequence and pCB102 replicon replaced with pBP2	This study

doi:10.1371/journal.pone.0056051.t002

fragment (using primers mtlD-NotI-F and mtlD-BamHI-R), and its subsequent insertion, following cleaved with NotI and BamHI, between the NotI and BamHI sites of pMTL-YN1C. pMTL-YN1X*mtlD* was derived by the PCR amplification of the *C. difficile* 630Δ*erm mtlD* structural gene alone (ie., lacking its native promoter), as a 1,161 bp fragment (using primers mtlD-NdeI-F and mtlD-BamHI-R), cleaved with NdeI and BamHI, and insertion between the NdeI and BamHI sites of pMTL-YN1X. This positioned the *mtlD* gene under the transcriptional and translational (RBS) control of *fdx* promoter.

ACE Procedure

The procedure adopted was as previously described [9]. For inactivation of *pyrE*, *E. coli* CA434 donor cells carrying pMTL-YN18 were conjugated with R20291 and transconjugants selected on BHIS media supplemented D-cycloserine (250 μg/ml), cefoxitin (8 μg/ml), thiamphenicol (15 μg/ml) and uracil (5 μg/ml). A single transconjugant was re-streaked onto the same medium and then a 'large' representative colony streaked onto minimal agar medium agar supplemented with FOA (2 mg/ml) and uracil (5 μg/ml). The colonies that arose were re-streaked twice onto the same media, and analyzed by PCR to confirm deletion of *pyrE* (as detailed in Results), and Sanger sequencing was used to confirm the expected genotype. Confirmation that the plasmid had been lost, was obtained by patch plating onto BHIS agar supplemented with thiamphenicol and establishing that no growth occurred. For correction of the *pyrE* mutation (ie., plasmids pMTL-YN1, 1C, 1X, 2, 2C and 2X) transconjugants were streaked onto minimal media without uracil or FOA supplementation and those colonies that developed analysed as above.

Allelic exchange procedure

The adopted protocol resembles that used for the isolation of allelic exchange mutants using *codA* as a negative selection marker [8], except counter-selection was based on differential resistance to FOA rather than FC. It in essence involves the initial isolation of pure single crossovers clones, and their subsequent transfer onto selective plates to identify those cells in which the plasmid has excised. These clones correspond to colonies that are resistant to FOA as a consequence of loss of the heterologous PyrE+ allele and reversion to a *pyrE* minus phenotype. Depending which homology arm undergoes recombination, plasmid excision can result in either the desired double crossover mutant, or a wild type cell. In the absence of any bias, the wild type cells and desired mutants should be present in equal proportions, ie., 50:50.

Each plasmid was conjugated into the appropriate host and plated onto BHIS media (supplemented with cycloserine, cefoxotin, thiamphenicol and uracil) and the transconjugants restreaked onto fresh media to allow the identification of putative single crossover mutants. These were readily identified as visibly larger colonies after 16–24 hr. Such colonies were sequentially restreaked from single colonies and after 2–4 passages, their identify and purity confirmed by an appropriate PCR. Unlike ACE, where powerful selection allows the purification of the desired double crossover excision event, it is crucial that the single cross-over integrants are pure, and not contaminated with wild type cells, ie., cells in which integration has not occurred. Otherwise, the contaminating wild type cells, which are all resistant to FOA, will cause unacceptable high background when cells are plated on FOA media during the screen for the desired double crossover events. To establish that no wild type cells were present, an appropriate primer pair (spo0A-YN-F2/spo0A-YN-R2 in the case of *spo0A*, cwp84-F3/cwp84-R4 in the case of *cwp84*, and mtlD-F3/mtlD-R3 the case of *mtlD*) was used that annealed to the chromosome at positions that flanked the site of the in-frame deletion but that were both distal to the regions that comprised the homology arms of the plasmid borne allelic exchange cassette. The absence of a DNA product of a size consistent with the wild type indicated clonal purity, and that wild type cells were not present. To confirm that they were indeed single crossover integrants, two different primers (M13F and M13R) were used that annealed to plasmid specific sequences together with the appropriate flanking primers. The presence of a DNA fragment indicated that the clones were indeed single crossover integrants, while the size was diagnostic of at which homology arm (LHA or RHA) recombination had occurred.

Following the isolation of pure single crossover integrants of the two strains, a single colony was re-streaked onto minimal medium supplemented with FOA and uracil to select for cells in which the integrated plasmid had excised. The isolated FOA resistant colonies were then screened by PCR using primers that anneal to the upstream and the downstream sequence of the respective target. DNA products of the predicted size were subjected to Sanger sequencing to confirm their genotype.

Cwp84 Analysis Methods

S-layer extracts from the WT, mutant and complemented strains were prepared by the glycin acid method [44]. Proteins were separated by SDS-PAGE and transferred onto a PVDF membrane for immunoblotting. Blocking of the membrane was followed by incubation with either anti-Cwp84 (1/2,000 dilution) or anti-HMW-SLP (1/10,000 dilution). Washing was done as previously described [45], and bound antibodies were detected using goat anti-rabbit immunoglobulin G alkaline phosphatase conjugate (1/20,000 dilution; Sigma) with the substrates nitroblue tetrazolium and 5-bromo-4-chloro-3-indolylphosphate (Sigma). Anti-HMW-SLP antibodies were a kind gift from Pr. Neil Fairweather, Imperial College, London.

Supporting Information

File S1 Supporting Information.
(DOCX)

Author Contributions

Conceived and designed the experiments: YKN AC STC NPM. Performed the experiments: YKN ME SP MMC CJ. Analyzed the data: YKN CJ AC NPM. Contributed reagents/materials/analysis tools: ME SP. Wrote the paper: YKN CJ SC NPM.

References

1. Poxton IR, McCoubrey J, Blair G (2001) The pathogenicity of *Clostridium difficile*. Clinical Microbiology and Infection 7: 421–427.
2. Rupnik M, Wilcox MH, Gerding DN (2009) *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. Nature Reviews Microbiology 7: 526–36.
3. Liyanage H, Kashket S, Young M, Kashket ER (2001) *Clostridium beijerinckii* and *Clostridium difficile* Detoxify Methylglyoxal by a Novel Mechanism Involving Glycerol Dehydrogenase. Applied Environmental Microbiology 67: 2004–2010.
4. O'Connor JR, Lyras D, Farrow KA, Adams V, Powell DR, et al. (2006) Construction and analysis of chromosomal *Clostridium difficile* mutants. Molecular Microbiology 61: 1335–1351.
5. Dineen SS, Villapakkam AC, Nordman JT, Sonenshein AL (2007) Repression of *Clostridium difficile* toxin gene expression by CodY. Molecular Microbiology 66: 206–219.

6. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP (2007) The ClosTron: a universal gene knock-out system for the genus *Clostridium*. Journal of Microbiological Methods 70:452–464.
7. Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, et al. (2010) The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. Journal of Microbiological Methods 80: 49–55.
8. Cartman ST, Kelly ML, Heeg D, Heap JT, Minton NP (2012) Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between *tcdC* genotype and toxin production. Applied Environmental Microbiology 78: 4683–90.
9. Heap JT, Ehsaan M, Cooksley CM, Ng Y-K, Cartman ST, et al. (2012) Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. Nucleic Acids Research 40(8):e59.
10. Boeke JD, Croute F, Fink GR (1984) A positive selection for mutants lacking orotidine-5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Molecular and General Genetics 197: 345–346.
11. Knipfer N, Seth A, Shrader TE (1997) Unmarked gene integration into the chromosome of *Mycobacterium smegmatis* via precise replacement of the *pyrF* gene. Plasmid 37: 129–140.
12. Sato T, Fukui T, Atomi H, Imanaka T (2003) Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. Journal of Bacteriology 185: 210–220.
13. Bitan-Banin G, Ortenberg R, Mevarech M (2003) Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene. Journal of Bacteriology 185: 772–778.
14. Deng L, Zhu H, Chen Z, Liang YX, She Q (2009) Unmarked gene deletion and host-vector system for the hyperthermophilic crenarchaeon *Sulfolobus islandicus*. Extremophiles 13: 735–746.
15. Tripathi SA, Olson DG, Argyros DA, Miller BB, Barrett TF, et al. (2010) Development of *pyrF*-Based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a *pta* Mutant. Applied Environmental Microbiology 76: 6591–6599.
16. Minton NP, Morris JG (1981) Isolation and partial characterization of three cryptic plasmids from strains of *Clostridium butyricum*. Journal of General Microbiology 127: 325–331.
17. Heap JT, Pennington OJ, Cartman ST, Minton NP (2009) A modular system for *Clostridium* shuttle plasmids. Journal of Microbiological Methods 78: 79–85.
18. Cartman ST, Minton NP (2010) A mariner-based transposon system for in vivo random mutagenesis of *Clostridium difficile*. Applied Environmental Microbiology 76: 1103–1109.
19. Chambers SP, Prior SE, Barstow DA, Minton NP (1988) The pMTL nic- cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. Gene 68: 139–149
20. Burns DA, Heap JT, Minton NP (2010) SleC is essential for germination of *Clostridium difficile* spores in nutrient-rich medium supplemented with the bile salt taurocholate. Journal of Bacteriology 192: 657–664.
21. Kirby JM, Ahern H, Roberts AK, Kumar V, Freeman Z, et al. (2009) Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of *Clostridium difficile*. Journal of Biological Chemistry 284: 34666–34673.
22. de la Riva L, Willing SE, Tate EW, Fairweather NF (2011) Roles of cysteine proteases Cwp84 and Cwp13 in biogenesis of the cell wall of *Clostridium difficile*. Journal of Bacteriology 193: 3276–3285.
23. Iwen PC, Booth SJ, Woods GL (1989) Comparison of media for screening of diarrheic stools for the recovery of *Clostridium difficile*. Journal of Clinical Microbiology 27: 2105–2106.
24. Postma PW, Lengeler JW, Jacobson GR (1993) Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiology and Molecular Biology Reveiws 57: 543–594.
25. Saier MH Jr, Reizer J (1992) Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. Journal of Bacteriology 174: 1433–1438.
26. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton NP, et al. (2006) The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nature Genetics 38: 779–786.
27. Henstra SA, Tolner B, ten Hoeve Duurkens RH, Konings WN, Robillard GT (1996) Cloning, expression, and isolation of the mannitol transport protein from the thermophilic bacterium *Bacillus stearothermophilus*. Journal of Bacteriology 178: 5586–5591.
28. Fabret C, Ehrlich SD, Noirot P (2002) A new mutation delivery system for genome-scale approaches in *Bacillus subtilis*. Molecular Microbiology 46: 25–36.
29. O'Callaghan D, Maskell D, Liew FY, Easmon CS, Dougan G (1988) Characterization of aromatic- and purine-dependent *Salmonella typhimurium*: attention, persistence, and ability to induce protective immunity in BALB/c mice. Infection and Immunity 56: 419–423.
30. Falkow S (1988) Molecular Koch's Postulates Applied to Microbial Pathogenicity. Reviews Infectious Disease 10 Suppl 2:S274–S276.
31. Emerson JE, Reynolds CB, Fagan RP, Shaw HA, Goulding D, et al. (2009) A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein. Molecular Microbiology 74(3):541–56.
32. Cheung JK, Keyburn AL, Carter GP, Lanckriet AL, Van Immerseel F, et al. (2010) The VirSR two-component signal transduction system regulates NetB toxin production in *Clostridium perfringens*. Infection and Immunity 78(7):3064–72.
33. Kennedy CL, Krejany EO, Young LF, O'Connor JR, Awad MM, et al. (2005) The alpha-toxin of *Clostridium septicum* is essential for virulence. Molecular Microbiology 57(5):1357–66.
34. Dingle TC, Mulvey GL, Armstrong GD (2011) Mutagenic analysis of the *Clostridium difficile* flagellar proteins, FliC and FliD, and their contribution to virulence in hamsters. Infection and Immunity 79(10):4061–7.
35. Li J, Chen J, Vidal JE, McClane BA (2011) The Agr-Like Quorum-Sensing System Regulates Sporulation and Production of Enterotoxin and Beta2 Toxin by *Clostridium perfringens* Type A Non-Food-Borne Human Gastrointestinal Disease Strain F560. Infection and Immunity 79(6): 2451–2459.
36. Rabinowitz JC (1972) Preparation and properties of clostridial ferredoxins. Methods in Enzymology 24:431–446.
37. Takamizawa A, Miyata S, Matsushita O, Kaji M, Taniguchi Y, et al. (2004) High-level expression of clostridial sialidase using a ferredoxin gene promoter-based plasmid. Protein Expression and Purification 36(1):70–5.
38. Theys J, Pennington O, Dubois L, Anlezark G, Vaughan T, et al. (2006) Repeated cycles of *Clostridium*-directed enzyme prodrug therapy result in sustained antitumour effects *in vivo*. British Journal of Cancer 95(9): 1212–1219.
39. Williams DR, Young DI, Young M (1990) Conjugative plasmid transfer from *Escherichia coli* to *Clostridium acetobutylicum*. Journal of General Microbiology 136: 819–826.
40. Hussain HA, Roberts AP, Mullany P (2005) Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630Δ*erm*) and demonstration that the conjugative transposon Tn*916*ΔE enters the genome of this strain at multiple sites. Journal of Medical Microbiology 54:137–41.
41. Smith CJ, Markowitz SM, Macrina FL (1981) Transferable tetracycline resistance in *Clostridium difficile*. Antimicrobial Agents and Chemotherapy 19:997–1003
42. Purdy D, O'Keeffe TA, Elmore M, Herbert M, McLeod A, et al. (2002) Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. Molecular Microbiology 46: 439–52.
43. Sambrook JRD. (2001) Molecular cloning: a laboratory manual. CSHL Press.
44. Calabi E, Ward S, Wren B, Paxton T, Panico M, et al. (2001) Molecular characterization of the surface layer proteins from *Clostridium difficile*. Molecular Microbiology 40(5):1187–99.
45. Janoir C, Péchiné S, Grosdidier C, Collignon A (2007) Cwp84, a surface-associated protein of *Clostridium difficile*, is a cysteine protease with degrading activity on extracellular matrix proteins. Journal of Bacteriology 189(20):7174–80.

Expanding the Repertoire of Gene Tools for Precise Manipulation of the *Clostridium difficile* Genome: Allelic Exchange Using PyrE Alleles.

SUPPORTING INFORMATION

Yen Kuan Ng[≠ 1], Muhammad Ehsaan[≠ 1], Sheryl Philip[1], Mark Collery[1], Clare Janoir[2], Anne Collignon[2], Stephen T. Cartman[1], Nigel P. Minton[1*]

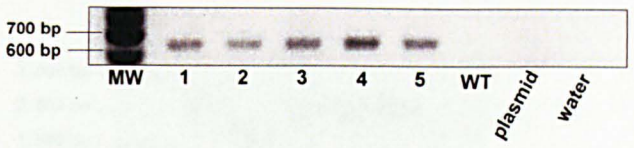
[≠] These authors contributed equally to this work.

Author affiliations:

[1]Clostridia Research Group, NIHR Biomedical Research Unit in GI Disease, Centre for Biomolecular Sciences, University Park, University of Nottingham, Nottingham, NG7 2RD, United Kingdom. [2]Université Paris-Sud, Faculté de Pharmacie, Département de Microbiologie, Unité EA 40-43, 92296 Châtenay-Malabry Cedex, France.

*To whom correspondence should be addressed: Nigel P. Minton
Clostridia Research Group, Centre for Biomolecular Sciences, University Park, University of Nottingham,
Nottingham, NG7 2RD, United Kingdom
Phone: +44 (0)115 84 67458
Fax: +44 (0)115 82 32120

(a)



(b)

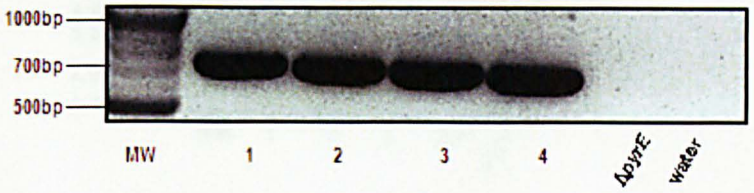
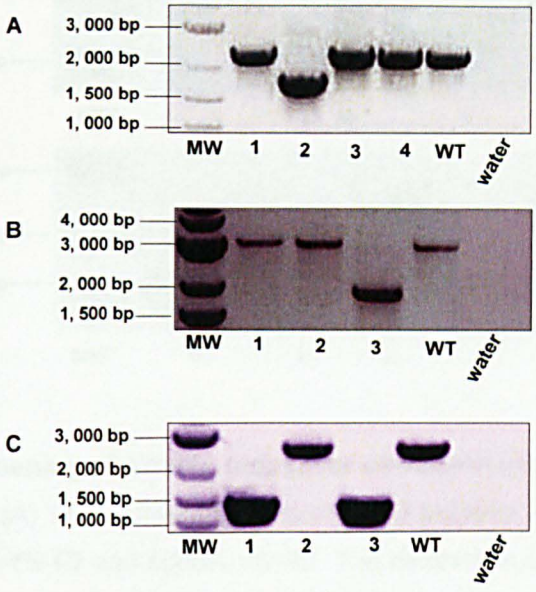


Figure S1. PCR screening of ACE generated *pyrE* mutant and wildtype strains in *C. difficile* strains 630Δ*erm* and *C. difficile* R20291. (a) five random FOA^R, *pyrE* mutants (lanes 1-5) made in R20291 using pMTL-YN18, and; (b) five random *pyrE* repaired clones (lanes 1-5) from R20291 made using pMTL-YN2. MW is a 2-Log DNA Ladder (NEB) molecular weight marker, plasmid is a pMTL-YN18 DNA control, WT is a wild-type *C. difficile* DNA control, and Δ*pyrE* is a mutants control taken from the strain in lane 1 of (a). The primers used, and expected size of the DNA fragments generated were: (a) primers pyrE-F2/pyrE-R2 and a 565 bp DNA fragment, and; (b) primers cdi630-pyrD-SF1/pyrE-R2 and a 750 bp DNA fragment.

42 FIGURE 10a



43

44 **Figure S2. PCR screening of double crossover candidate clones of *C. difficile* 630Δ*erm***
45 **for *spo0A*, *cwp84* and *mtlD*.** (A) PCR screening of four (1-4) putative *spo0A* in-frame deletion
46 clones using primers spo0A-YN-F2 and spo0A-YN-R2. The desired mutant should give a DNA
47 product of 1,845 bp in size (clone 2), while the wild type allele (clones 1, 3 and 4) generates a
48 product of 2,331 bp in size. (B) PCR screening of three (1-3) putative *cwp84* in-frame deletion
49 clones using primers cwp84-F3 and cwp84-R4. A double crossover mutant (clone 3) should give
50 a PCR product of 1,418 bp in size, as opposed to the 2,636 bp fragment predicted for the wild
51 type allele (clones 1 and 2). (C) PCR screening of three (1-3) putative in-frame deletions of *mtlD*
52 using primers mtlD-F3 and mtlD-R3. The in-frame deletion mutant is predicted to yield a PCR
53 DNA product 1,418 bp in size (clones 1 and 3), whereas the wild-type allele (clone 2) gives a
54 DNA product of 2,582 bp in size. PCR product from wild-type or double-crossover wild-type
55 revertants. MW is a 2-Log DNA Ladder (NEB) molecular weight marker and WT is a wild-type *C.*
56 *difficile* DNA control.

57

58

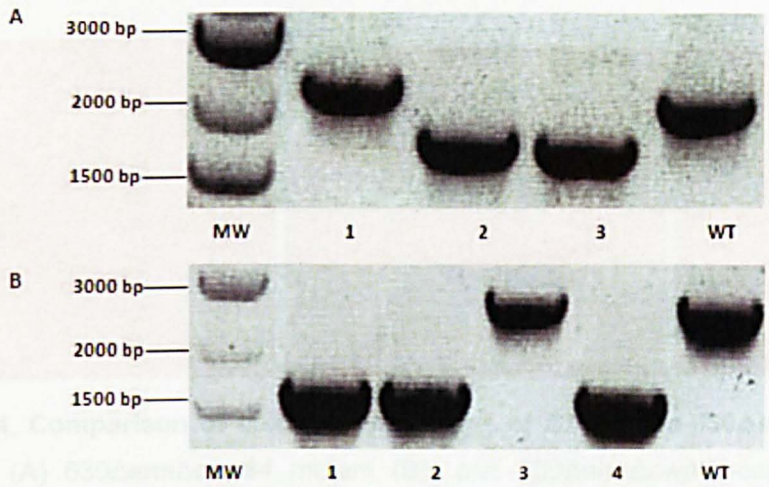


Figure S3. PCR screening of double crossover candidate clones of *C. difficile* R20291 for *spo0A* and *cwp84*. (A) PCR screening of three (1-3) putative *spo0A* in-frame deletion clones using primers spo0A-YN-F2 and spo0A-YN-R2. The desired mutant should give a DNA product of 1,845 bp in size (clones 2 and 3), while the wild type allele (clone1) generates a product of 2,331 bp in size. (B) PCR screening of four (1-4) putative *cwp84* in-frame deletion clones using primers cwp84-F3 and cwp84-R4. Double crossover mutants (clones 1, 2 and 4) should give a PCR product of 1,418 bp in size, compared to the 2,636 bp fragment predicted for the wild type allele (clone 3). MW is a 2-Log DNA Ladder (NEB) molecular weight marker and WT is a wild-type *C. difficile* DNA control.

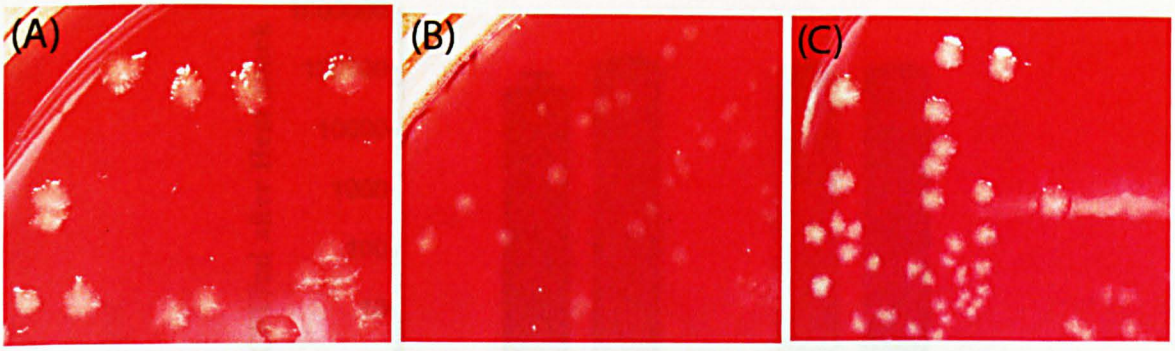


Figure S4. Comparison of colony morphology of *C. difficile* 630Δerm strains. *C. difficile* 630Δerm (A) 630ΔermΔ*cwp84* mutant (B), and 630ΔermΔ*cwp84*-complemented (C) were streaked onto anaerobic blood agar plate and incubated overnight to observe colony morphology. Colonies of *C. difficile* 630Δerm (A) and 630ΔermΔ *cwp84*-complemented (C) are indistinguishable and showed irregular edged colonies compared to *C. difficile* 630Δerm cwp84 mutant (B) showed more rounded colonies and are noticeably smaller due to its slower growth rate.

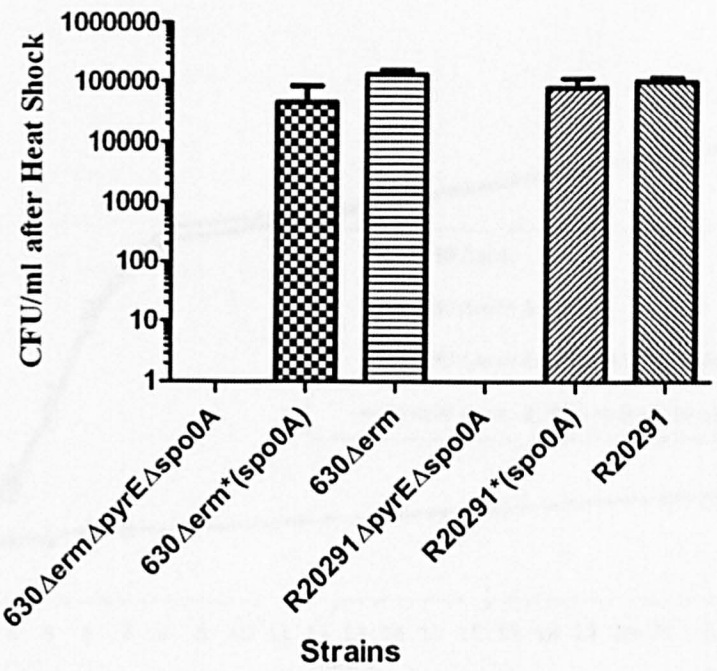
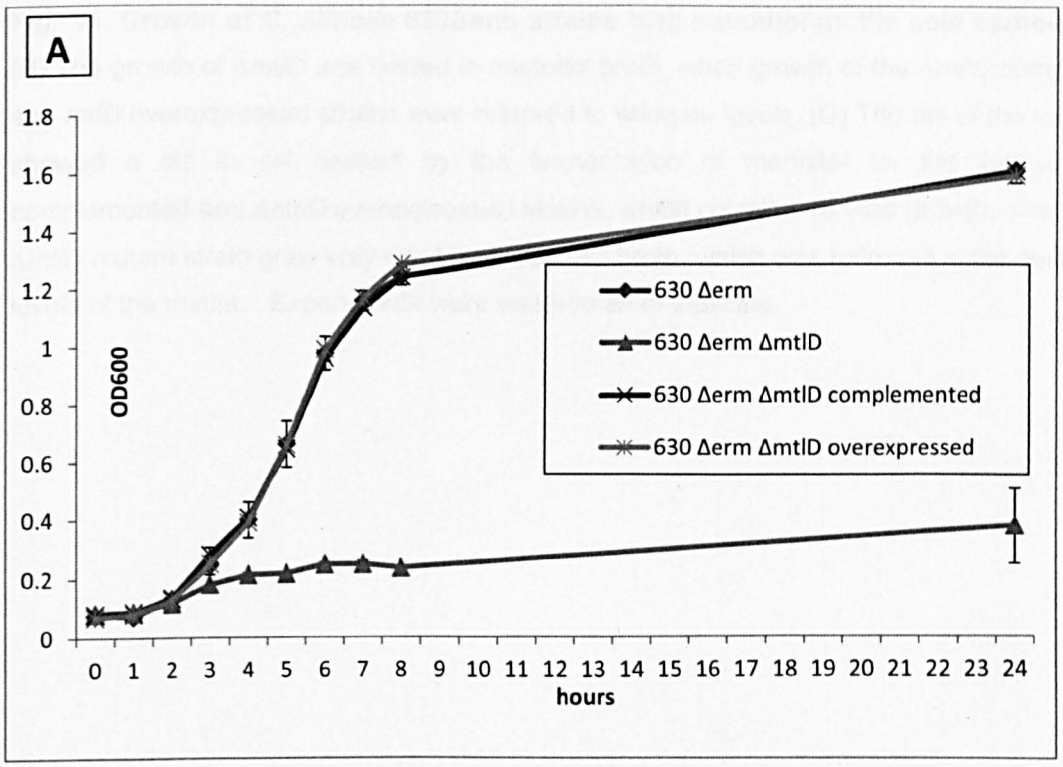
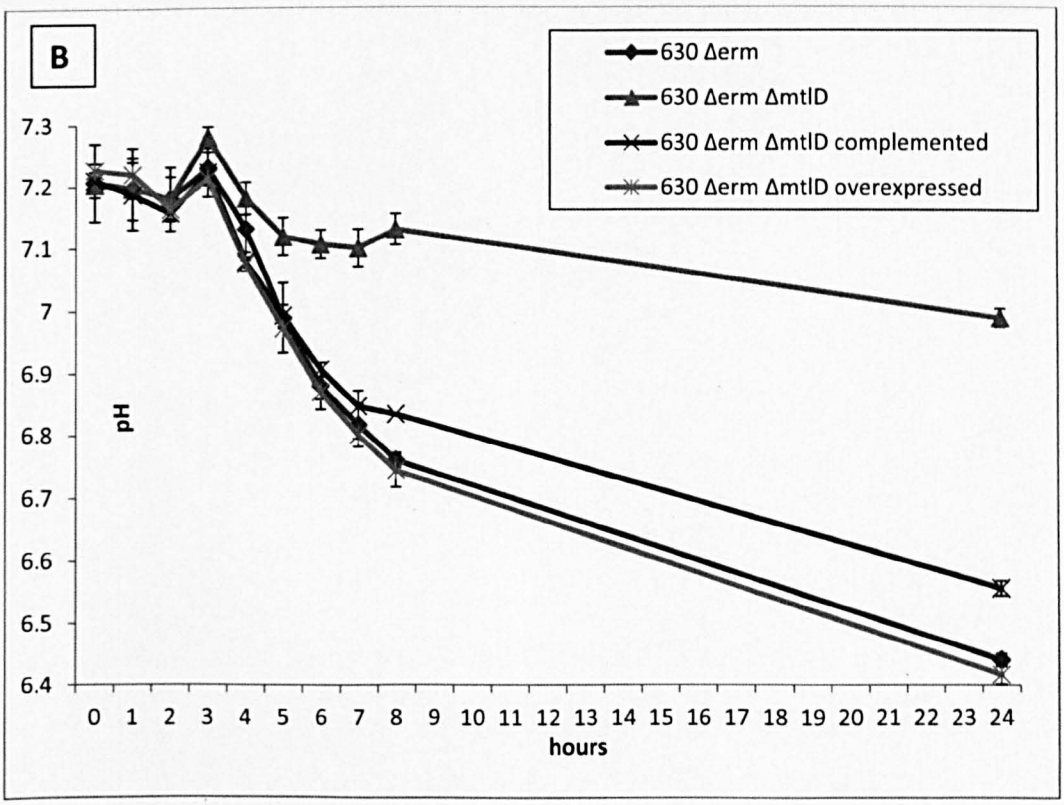


Figure S5. Complementation of *spo0A* mutants of *630Δerm*Δ*pyrE* and R20291Δ*pyrE*. The sporulation phenotype of the mutants and complemented mutants were compared to the wildtype strains by assaying colony forming units (CFU) on BHIS supplemented with 0.1% [w/v] sodium taurocholate before and after heat shock (65°C for 30 min) following growth of each strain in BHIS media for 120 h. Assayed strains are: 630Δ*erm*Δ*pyrE*Δ*spo0A*; 630Δ*erm**(*spo0A*); 630Δ*erm*; R20291Δ*pyrE*Δ*spo0A*; R20291*(*spo0A*), and; R20291, where * indicates the *spo0A*-complemented strain (*spo0A*) which has had a copy of the *spo0A* gene inserted into the chromosome, concomitant with the correction of the *pyrE* allele back to wildtype. The detection limit for the assay was 50 CFU/ml. All experiments were undertaken in triplicate. Phase contrast microscopy confirmed that phase-bright spores were absent in the mutant cultures but present in both the complemented and wildtype strains.


A
OD600
hours
630 Δerm
630 Δerm ΔmtlD
630 Δerm ΔmtlD complemented
630 Δerm ΔmtlD overexpressed

117


B
pH
hours
630 Δerm
630 Δerm ΔmtlD
630 Δerm ΔmtlD complemented
630 Δerm ΔmtlD overexpressed

118

Fig. S6. Growth of *C. difficile* 630Δerm strains with mannitol as the sole carbon source.. (B) The growth of Δ*mtlD* was limited in mannitol broth, while growth of the Δ*mtlD* complemented and *mtlD* overexpressed strains were restored to wildtype levels. (C) The pH of the media broth showed a dip in pH caused by the fermentation of mannitol for the wildtype, Δ*mtlD* complemented and Δ*mtlD* overexpressed strains, which correlate to their growth. The 630 Δ*erm* Δ*mtlD* mutant strain grew very weakly in mannitol broth, which was reflected in the sustained pH levels of the media. Experiments were undertaken in triplicate.