Development of high-throughput genome editing tools towards ethylene production in *Cupriavidus* species

Sophie Cécile Vaud, MSc.

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

January 2019

Declaration

Except where stated otherwise acknowledged, the work presented in this thesis is entirely my own. No part has been submitted for another degree at The University of Nottingham or any other institute of learning.

Sophie Vaud, January 2019.

Abstract

Depletion of natural hydrocarbon resources has catalysed research interest into sustainable routes for the production of bulk chemicals. *Cupriavidus necator* H16 has been extensively studied for the production of polyhydroxybutyrate (PHB), a biopolymer utilised as an alternative to petroleum-based plastics. However, the strain lacks efficient, fast, and user-friendly strain engineering tools. A single mutant is typically generated via a conjugation/counterselection method, which requires multiple steps and results in a maximum efficiency of 50%, necessitating extensive screening via colony PCR.

Here is presented the development of HTP (high-throughput) editing tools in *C. necator*. These tools were then employed for the metabolic engineering of *Cupriavidus metallidurans* (*C. metallidurans*), another chassis utilised within the Synthetic Biology Research Center (SBRC) of Nottingham along with *C. necator*. In particular, engineering efforts focused on implementing for the first time the ethylene-forming enzyme (EFE) pathway in *C. metallidurans* and improving production of ethylene, a platform chemical of the SBRC, in that strain.

The assessment of Lambda-Red (λ-Red) and RecET recombineering systems were inconclusive and highlighted the difficulty to adapt λ-Red outside of *Escherichia coli* (*E. coli*). The implementation of CRISPR/Cas9 required many optimisation steps before the emergence of a mutant, with an overall efficiency of 40%.

Additional HTP tools were further designed for introduction and optimisation of the Ethylene-Forming Enzyme (EFE) pathway in *Cupriavidus metallidurans* CH34. These HTP tools were first applied in *E. coli* as proof of concept and enabled a 6.3-fold increase in ethylene productivity, compared to the highest ethylene productivity reported to

date in *E. coli* (Lynch *et al.*, 2016). The global Transcriptional Machinery Engineering (gTME) technique involved the semi-automated creation of an *rpoD* mutant library and ultimately participating in the emergence of ethylene overproducing strains. To maximise the selection of mutants with desirable traits, ethylene synthesis was coupled to proline formation via a growth couple and cells were maintained in a proline-free growth medium during Adapted Laboratory Evolution (ALE) fermentation. The gTME and ALE engineering methods are readily available for transfer into *C. metallidurans* and by extension, to other *Cupriavidus* strains.

Altogether, the development of genomic, transcriptomic and metabolomic engineering tools described in this work will boost the strain engineering potential of these non-model chassis for both current and novel chemical production.

Acknowledgments

First and foremost, I would like to thank my supervisors Drs Stephan Heeb, Samantha Bryan, Klaus Winzer and Jamie Twycross. Stephan Heeb has provided very good advice on how to navigate in the academic research environment and showed me an enthusiastic approach to science. These last four years in the world of research have been challenging and intense, and I would have never been able to get this far without the immensely valuable supervision of Samantha Bryan, who has brought me the scientific and human competencies to be a good scientist. She has also constantly supported me in every success and difficult time one can encounter along a PhD course and I am looking forward to continuing collaborating with her after my PhD studies.

To the BBSRC DTP who funded this research work.

To Magda Jonczyk, who helped me a lot on the robotic-associated work on ethylene. To Ines Canadas who provided the riboswitch-controlled CRISPR vector and for the valuable conversations we had about CRISPR.

To my colleagues and friends from CBS, those long hours in the lab wouldn't have been the same without you. To many friends for their constant support.

To my new colleagues from Engineering, who have been incredibly supportive during the writing period and to my project supervisors Alex Conradie and Gill Stephens, who trusted me to continue the exciting DETOX project.

To Lorenzo, my 'beacon in the darkness', who never failed supporting me in spite of (maybe) too many week-ends spent in CBS.

And of course, to my parents, who I hope will be proud of their daughter.

Table of contents

List of figures

List of tables

Abbreviations

ºC	Degrees Celsius
2-OG	2-oxoglutarate
ACC	1-aminocyclopropane-1-carboxylate
ACO	ACC oxidase
ALE	Adaptive laboratory evolution
Amp	Ampicillin
AKG	Alphaketoglutarate
ATP	Adenosine triphosphate
AU	Arbitrary units
bp	Base pair
Cas	CRISPR associated
CDS	Coding domain sequence
cfu	Colony forming units
cm	Centimetre
Cm	Chloramphenicol
cPCR	Colony PCR
CRISPR	Clustered regularly interspaced palindromic repeats
CRMAGE	CRISPR-MAGE
crRNA	CRISPR RNA
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSB	Double strand break
ds	Double-strand
EDTA	Ethylene diamine-tetra-acetic acid
EFE	Ethylene forming enzyme
eYFP	Enhanced yellow fluorescent protein
F	Farad
g	Grams
GC	Gas Chromatography
GFP	Green fluorescent protein
gRNA	Guide RNA

gTME	Global transcription machinery engineering
h	Hour
HA	Homology arms
HDR	Homology-directed repair
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HiFi	High Fidelity
HR	Homologous recombination
HRP	Horseradish peroxidase
HTP	High-throughput
iGEM	International genetically engineered machine
IGg	Immunoglobin g
Kan	Kanamycin
kb	Kilobase
KMBA	2-keto-4-methylthiobutyric acid
kt	Kiloton
λ	Lambda
LB	Lysogeny broth
LBA	LB + Ampicillin
LBC	LB + Chloramphenicol
LBK	LB + Kanamycin
LBT	LB + Tetracyclin
LBTm	LB + Trimethoprim
LBCS	LB+Chloramphenicol+Spectinomycin
LBCT	LB+Chloramphenicol+Tetracyclin
LBCTm	LB+Chloramphenicol+Trimethoprim
LSLB	Low salts LB
MAGE	Multiplex automated genome engineering
Mbp	Mega base pair
MIC	Minimal inhibitory concentration
min	Minute
mL	Millilitre
mM	Millimolar concentration
MM	Minimum medium

mRNA	RNA messenger
MSM	Minimal salts medium
MTA	Methylthioadenosine
NAD	Nicotinamide adenine dinucleotide
NHEJ	Non-homologous end joining
ng	Nanogram
nm	Nanometre
No Theo	No theophylline
nt	Nucleotide
Ω	Ohm
OD	Optical density
P5C	Δ-1-pyrroline-5-carboxylate
PAM	Protospacer adjacent motif
PBS	Phosphate Buffer Saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PHA	Polyhydroxyalkanoate
PHB	Polyhydroxybutyrate
PHBV	Polyhydroxybutyrate-co-hydroxyvalerate
PHBHHx	Polyhydroxybutyrate-co-hydroxyhexanoate
pv	Pathovar
RBS	Ribosome binding site
RNA	Ribonucleic acid
RO	Reverse osmosis
RM	Restriction-modification
rpm	Revolutions per minute
RR	Response regulator
RT	Room temperature
s	Second
SAM	S-adenosylmethionine
SBRC	Synthetic biology research centre
SC	Stream cracking
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SG	Sodium gluconate
sgRNA	Single gRNA
SOB	Super optimal broth
SOBC	SOB + Chloramphenicol
SOC	Super optimal broth with catabolite repression
Spec	Streptomycin
spp.	Species
SpyCas9	Streptococcus pyogenes Cas9
ss	Single-strand
StmCas9	Streptomyces Cas9
t	Ton
TALEN	Transcription-activator like effector nuclease
TCA	Tricarboxylic acid cycle
Tet	Tetracycline
TF	Transformation frequencies
Theo	Theophylline
Tm	Trimethoprim
tracR	Transactivating RNA
Tris-HCl	2-amino-2-hydroxymethyl-propane-1,3-diol
w/v	Weight per volume
µg	Microgram
µL	Microliter
µM	Micromolar concentration
UV	Ultra violet
ZFN	Zinc-finger nuclease

Introduction

1.1 Cupriavidus necator H16, a versatile microorganism

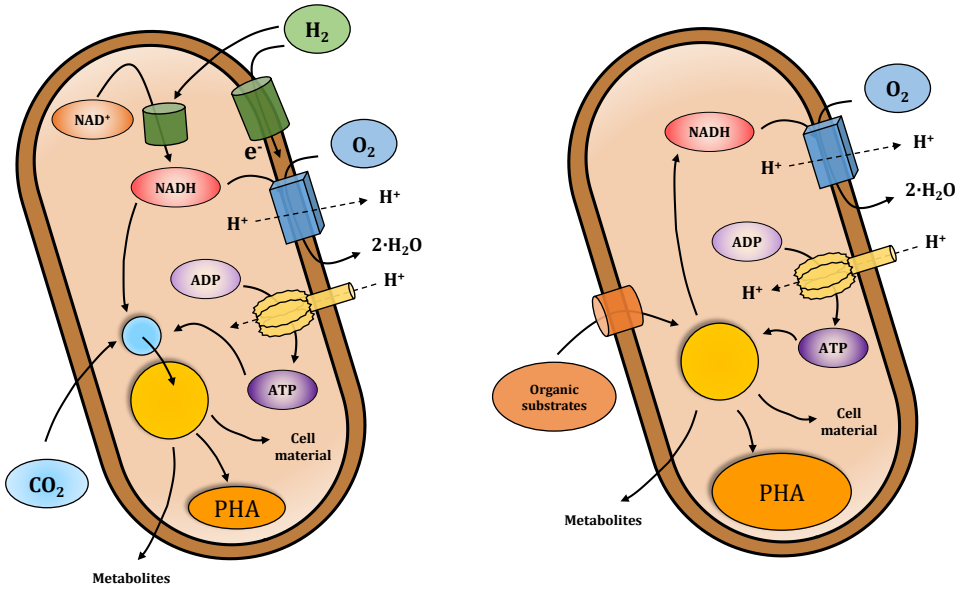
The Synthetic Biology Research Centre (SBRC) at the University of Nottingham aims to utilise non-model chassis, e.g. bacteria of the *Cupriavidus* and *Clostridium* genera, for the industrial production of commercially valuable platform chemicals from C1 carbon compounds. This study will focus on *Cupriavidus necator* and *Cupriavidus metallidurans*.

1.1.1 A brief history of C. necator H16

Originally isolated from sludge by Bovell and Wilde in 1961, *Cupriavidus necator* H16 was initially called *Hydrogenomonas eutrophus,* due to its ability to oxidise H_2 (Wilde, 1962). Subsequent phylogenetic reclassification led to several name changes, *Alcaligenes eutropha* (Davis *et al.*, 1969), *Ralstonia eutropha* (Yabuuchi *et al.*, 1995), *Wautersia eutropha* (Vaneechoutte *et al.*, 2004) and finally after almost half a century of debate, it was named *Cupriavidus necator* (Vandamme and Coenye, 2004).

1.1.2 Natural properties of C. necator H16

C. necator H16 is a gram-negative lithoautotrophic organism and belongs to the class of Betaproteobacteria, specifically the Burkholderiaceae family (Yabuuchi *et al.*, 1995). A natural inhabitant of soil and freshwater biotopes, it thrives on organic compounds such as sugars, but is also able to utilise CO_2 as a sole carbon source (**Figure 1.1**) (Gai *et al.*, 2014). The fixation of CO_2 is mediated through the Calvin-Benson-Bassham cycle, while H_2 oxidisation ensures replenishment of proton energy (Pohlmann *et al.*, 2006). *C. necator* H16 is also well adapted to transient anoxia (Cramm, 2009).



Lithoautotrophic metabolism Heterotrophic metabolism

Figure 0.1 – The versatile metabolism of C. necator H16

The yellow circles indicate central metabolism, while the blue circle is the Calvin-Benson-Bassham cycle. The green squares symbolise the two energy-conserving hydrogenases. The orange circles represent polyhydroxyalkanoate (PHA), mainly polyhydroxybutyrate (PHB), storage granules. Adapted from Pohlmann *et al*, 2006.

C. necator has been extensively studied as the model organism for polyhydroxybutyrate (PHB) biosynthesis, because this polyhydroxy-alkanoate (PHA) possesses many of the properties found in petroleum-based plastics (O P Peoples and A J Sinskey, 1989). *C. necator* can naturally accumulate up to 61-77% of its dry cell weight as PHB under specific conditions (high carbon, limiting nitrogen and phosphorus) (Wong *et al.*, 2012; Obruca *et al.*, 2014). PHB has been produced at scale in *C. necator* since the 1960s, utilising diverse carbon sources such as crop waste, sugar beets and vegetable oils and is commercialised under different trademarks (Biopol™, Minerv® , ENMAT™) (Kourmentza *et al.*, 2017).

Over the last decade, metabolic engineering of *C. necator* H16 has mainly focused on the optimisation of the native PHB pathway (Reinecke and Steinbüchel, 2009; Mohidin Batcha *et al.*, 2014; Obruca *et al.*, 2014; Aramvash *et al.*, 2015). Substantial efforts have also been made to broaden

the range of polymers currently synthesised, via incorporation of diverse monomers in the formation of PHAs, generating, for instance, poly (3-hydroxybutyrate-co-3-hydroxyvalerate, PHBV) (Ghysels *et al.*, 2018), poly (3-hydroxybutyrate-co-4-hydroxybutyrate)(Lee, Park and Huh, 1997) and poly (3-hydroxybutyrate-co-3-hydroxyhexanoate, PHBHHx) (Wong *et al.*, 2012) copolymers. Furthermore, efforts have been made to redirect cellular metabolism towards the production of high quality hydrocarbon biofuels from CO_2 (Lu *et al.*, 2012; Bi *et al.*, 2013; Kang and Yu, 2015).

1.1.3 Organisation and general features of the genome of C. necator H16

The genome of *C. necator* is composed of two chromosomes, chromosome 1 (4,052,032 base pairs (bp)) and chromosome 2 (2,912,490 bp), and one megaplasmid (452,156 bp), for a total of 7.4 Mbp, making it one of the largest genomes among the Burkholderiaceae family (Fricke, Kusian and Bowien, 2009). In 2006, both chromosomes were sequenced and analysed (Pohlmann *et al*, 2006). The sequence of the megaplasmid pHG1 had been previously reported (Schwartz *et al*, 2001). As predicted, genes involved in H_2 oxidation and CO_2 fixation are well represented amongst the 6,116 coding domain sequences (CDS). Multiple transport proteins have also been identified, an abundance that perhaps reflects the high concentration and the diversity of nutrients found in soils and freshwaters, from which the organism was originally isolated. The megaplasmid pHG1 contains genes essential for lithoautotrophy and facultative anaerobic growth e.g. genes coding for hydrogenases and enzymes involved in denitrification (Schwartz and Friedrich, 2001), as well as mobile elements, integrases and the necessary elements for conjugation.

For more than two decades, most of the genetic studies were conducted on the wild-type H16 strain and the PHB-negative mutant PHB$^-$4, obtained from chemical mutagenesis (Schlegel, Lafferty and Krauss, 1970). Recombinant strains expressed heterologous genes harboured on plasmids

but the chromosomes were not modified, limiting the study of the organism (Cook and Schlegel, 1978; Peoples and Sinskey, 1989; Fukui and Doi, 1997).

The first gene deletion in *C. necator* targeted 1.8 kb of the *hox*Xgene, which is involved in the regulation of hydrogen oxidation (Lenz *et al.*, 1994). The deletion was made by allelic exchange utilising double homologous recombination and sucrose-based counter-selection. A ColE1-based suicide plasmid was utilised, pLO1, with homology arms complementary to the genomic target (on megaplasmid pHG1) and the *sacB* gene from *Bacillus subtilis* (Simon, Priefer and Pühler, 1983). Transconjugants were selected on minimum medium supplemented with the appropriate antibiotic. The plasmid DNA was integrated into pHG1 by homologous recombination following a procedure previously published (Bernhard *et al.*, 1996) and selected clones were inoculated in Low Salt LB-MOPS, with no antibiotics to allow for a second homologous recombination to occur. Approximately 10^8 cells were spread onto LB plates containing 15% sucrose. The *sacB* gene codes for the exoenzyme levansucrase, which converts sucrose into levan polymer (Gay *et al.*, 1983). Though the toxicity mechanism is not completely understood, accumulation of this high-molecular fructose polymer in the periplasm of gram-negative cells is lethal (Steinmetz *et al.*, 1983), therefore selecting clones which achieved a second homologous recombination and the excision of the suicide plasmid from the chromosome or megaplasmid. This results in two possible outcomes, either a reversion to the wild-type chromosome/megaplasmid or the generation of a markerless mutant (**Figure 1.2**) (Nakashima *et al.*, 2014). Sucrose-resistant colonies were selected and screened by colony PCR (cPCR), for successful allelic exchange recombination.

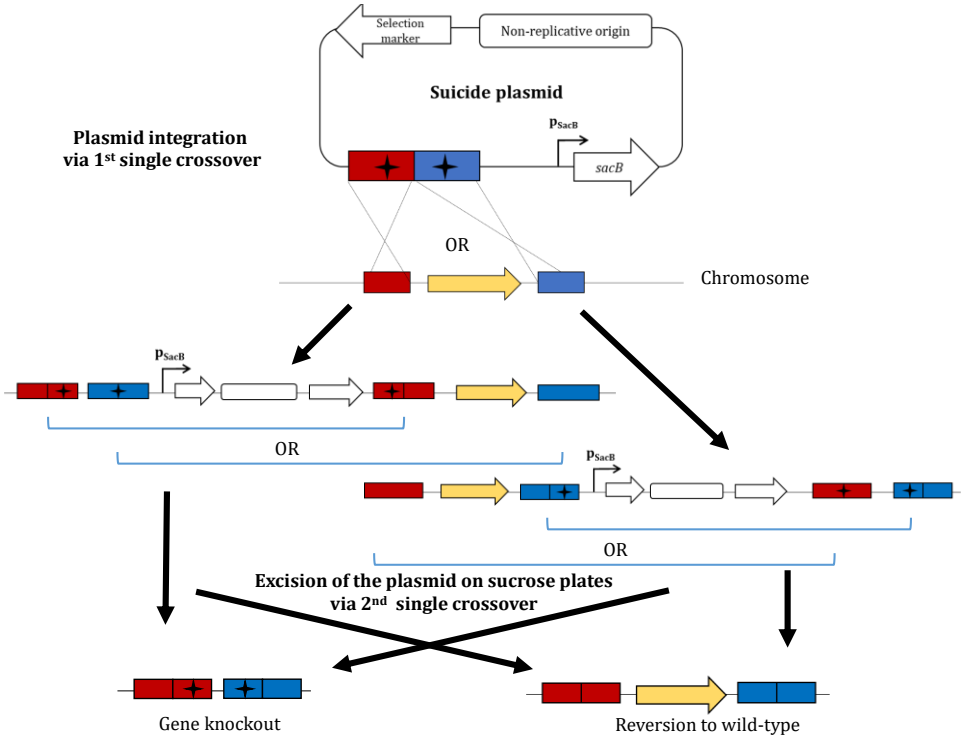


Figure 0.2 – Procedure for allele replacement mutagenesis via sucrose counterselection

The suicide vector contains the *sacB* gene conferring sucrose sensitivity and an antibiotic resistant selection marker. After conjugation, the deletion construct integrates into one of the chromosomes or the megaplasmid by homologous recombination. The resulting strain is selected for antibiotic resistance. Cells having excised the integrated suicide vector are selected on sucrose-enriched medium after a second crossover event, which results in either the generation of a deletion strain or reversion to wild type. Strains are selected for sucrose-resistance, followed by confirmation of plasmid loss by antibiotic sensitivity, and screening for allelic replacement by cPCR. Adapted from Nakashima *et al.*, 2014.

Several studies intended to develop a genetic toolbox for easy and efficient metabolic engineering of *C. necator*. This set of tools now includes: expression vectors having origins of replication of low-, medium- and high-copy-numbers (Bi *et al.*, 2013; Gruber *et al.*, 2014; Alagesan, Minton and Malys, 2018), constitutive and inducible promoters (Delamarre and Batt, 2006; Fukui *et al.*, 2011; Johnson *et al.*, 2018) and ribosomal binding sites derived from *Escherichia coli* (*E. coli)*, *C. necator*, and computational design (Bi *et al.*, 2013). Combinatorial assembly of these regulatory biobricks enables the fine-tuning of heterologous gene expression in H16 and the application of this strain in industrial production.

Subsequent efforts to improve mutant generation in *C. necator* have been laborious and still rely on sucrose-based counterselection (Lenz and Friedrich, 1998). This method can be difficult and does not always result in effective gene replacement. As an example, the *fadD* gene, B1148, could not be deleted despite several attempts (Chen *et al.*, 2015). Furthermore, the *sacB* counterselection takes three weeks (Xiong *et al.*, 2018) to complete in *C. necator* and only has an efficiency rate of 50% (Lenz, Lauterbach and Frielingsdorf, 2018; Xiong *et al.*, 2018). Besides, the allelic exchange frequency generally decreases significantly as the size of the gene deletion increases (Posfai *et al.*, 1999). Indeed, intramolecular recombination efficiency depends on either the physical distance between the homologous regions and/or direct repeats on the chromosome (Posfai *et al.*, 1999). Consequently, cPCR-based screening must be deployed, with hundreds of colonies needing to be screened, to identify one successful mutant among many revertants to wild type. Although the synthetic toolbox is expanding, efficient high-throughput (HTP) tools still need to be developed to enable the generation of genome-scale mutant libraries.

1.2 *Cupriavidus metallidurans*, a multi-metal resistant bacterium

1.2.1 A brief history of C. metallidurans CH34

Cupriavidus metallidurans CH34 was initially isolated from a metal factory in Belgium in 1978 (Mergeay, Houba and Gerits, 1978). This bacterium thrives in toxic anthropogenic environments, which contain millimolar levels of heavy metals like copper, cobalt, cadmium, lead or mercury (Monchy *et al.*, 2007).

C. metallidurans has also undergone several name changes from *Alcaligenes eutrophus* CH34 (Collard *et al.*, 1994) to *Ralstonia metallidurans* (Goris *et al.*, 2001), *Wautersia metallidurans* (Munkelt, Grass and Nies, 2004) and finally *Cupriavidus metallidurans* (Vandamme and Coenye, 2004).

1.2.2 Natural properties of C. metallidurans CH34

Like *C. necator,* CH34 can also synthesise PHB, albeit at lower densities than *C. necator* (Pohlmann *et al.*, 2006). The membrane-bound hydrogenase of CH34 showed a higher catalytic activity, a better H_2 affinity, and a higher O_2 tolerance than the hydrogenase from H16 (Mergeay *et al.*, 1985), making CH34 an attractive platform for the development of microbial fuel-cell technology from CO_2 (Wrighton and Coates, 2009).

C. metallidurans remains the microorganism of choice for studying metal resistance (Janssen *et al.*, 2010). Indeed, CH34 has been widely utilised to understand the mechanism behind gold bio-mineralisation, allowing the conversion of gold chloride into 24-carat gold (Reith *et al.*, 2009; Lal *et al.*, 2013). Despite scale-up being currently unfeasible, studies have allowed the ecological significance and metal resistance to be fully investigated in this strain (Julian *et al.*, 2009; Rojas *et al.*, 2011; Bütof *et al.*, 2017). *C. metallidurans* possesses many inducible efflux pumps that

mediate the exportation of metal ions present in the cytoplasm. This resistance system was initially thought to be limited to cell detoxification (Nies, 2000). In fact, *C. metallidurans* is able to solubilise the metals and subsequently sequester the metal ions by bioprecipitation or bioadsorption in nuclei on the surface of outer membrane proteins (Diels *et al.*, 2009). Therefore, engineering efforts have been focused on optimising *C. metallidurans* for the degradation of xenobiotics and aromatic compounds in metal-rich waste water (Springael, Diels and Mergeay, 1994; Diels *et al.*, 2009) and for the indirect precipitation of heavy metals (Diels *et al.*, 2009). This could be the first step towards the scale-up of bioremediative processes in *C. metallidurans*.

1.2.3 Organisation and general features of the genome of C. metallidurans CH34

Sequencing of the genome of *C. metallidurans* CH34 resulted in the identification of 25 loci involved in heavy-metal resistance and a large number of genes coding for transporters or signal transduction systems (Janssen *et al.*, 2010). CH34 has two chromosomes CHR1 and CHR2, of 3,928,089 bp and 2,580,084 bp respectively, and two megaplasmids, pMOL28 and pMOL30, 171,459 bp and 233,720 bp in size, respectively. Most of the genes involved in cellular functions (cell cycle control, replication and translation) are located on CHR1, while CHR2 contains genes involved in cell motility, signal transduction and secondary metabolism. The plasmids carry genes involved in heavy-metal resistance and similarities have been found between pMOL28 and the megaplasmid pHG1 of *C. necator*. In addition, the presence of many insertion sequences and transposition elements indicates the genomic plasticity of CH34 and suggests that foreign genes are readily assimilated by this bacteria (Janssen *et al.*, 2010).

Genome engineering of *C. metallidurans* CH34 was achieved for the first time with the deletion of eleven sigma factor genes via allelic

replacement utilizing the Cre/lox excision. The suicide plasmid pECD889 harboured two 300 bp fragments complimentary to the sequences upstream and downstream of the target gene and a kanamycin-resistant cassette flanked with two *loxP* recognition sites (Grosse, Friedrich and Nies, 2007). The pECD889 plasmid was utilised to replace the target gene with a kanamycin resistance cassette in the genome of *C. metallidurans* via double crossover recombination. To create an in-frame markerless deletion, the Cre recombinase from the plasmid pCM157 (Marx and Lidstrom, 2002) was expressed, which catalysed *in vivo* excision of the resistance cassette at the *loxP* recognition sites. However, the Cre/lox excision still leaves a scar at the genomic locus and subsequent mutations could lead to undesirable intramolecular recombination between the scar sequences (Campo *et al.*, 2002; Suzuki *et al.*, 2005). This method has been extensively utilised to generate mutant strains in different bacteria (Suzuki *et al.*, 2005; Noskov *et al.*, 2015).

Developing a straightforward, efficient and HTP genome editing method suitable for *C. necator* and *C. metallidurans* is thus necessary, allowing these microorganisms to be developed as platforms for designer microbial chassis.

1.3 High throughput genetic engineering techniques

There is an increasing drive towards developing a sustainable route to produce bulk chemicals by fermentation. However, biotechnological processes must be economically viable to compete directly with unsustainable industrial practices. A number of chemical platform molecules are already manufactured from bacteria or yeast, like bioethanol, produced at industrial scale in *Saccharomyces cerevisiae* (maximum ethanol concentration 97.94 g/L and productivity 4.19 g/L/h) (Zhang *et al.*, 2011) and *E. coli* (40–55 g/L) (Koppolu and Vasigala, 2016). The gram-positive *Corynebacterium glutamicum* is the workhorse for industrial bioproduction of amino-acids, mainly glutamic acid (~3 million tons in 2014) (Wendisch *et al.*, 2016). In 1916, *Clostridium acetobutylicum* was one of the first organisms to be used in large-scale microbial fermentation processes for acetone and butanol production (Sauer, 2016). Imperial Chemical Industries Ltd (Great Britain) initiated the industrial production of poly-(3-hydroxybutyrate-co-3-hydroxyvalerate) in 1982 in *C. necator* and commercialised the polymer under the name Biopol™. A few companies continued the industrial fermentation of *C. necator* and extended the range of PHAs commercialised (Chen and Chen, 2010). However, PHB metabolism has been implemented in many other bacteria, such as the model *E. coli* chassis or the CO_2-fixing cyanobacterium *Synechocystis* (Reinecke and Steinbüchel, 2009), which are great candidates to replace *C. necator* in industrial applications. Therefore, more efforts are focused on lowering the PHA production cost in *C. necator* including the metabolic engineering of the strain to improve cell density, growth rate and formation of larger PHA granules for easy separation (Chen, 2009). Strain engineering could also enable the reshaping of cellular metabolism to direct carbon flux towards diverse metabolic products of high commercial value like ethylene (Harmsen, Hackmann and Bos, 2014), itaconic acid (Choi *et al.*, 2015) or butanol and increase the industrial potential of *C. necator*. As mentioned

above, *C. metallidurans* would also benefit from development of engineering approaches for the scale-up of environmental applications.

Nowadays, metabolic engineering relies on HTP genetic tools to generate a large number of mutants with superior or interesting phenotypic traits in a quick, efficient and cost-effective manner (Mougiakos *et al.*, 2018). Here are described the most characterised and efficient genetic tools for mainstream metabolic engineering.

1.3.1 Recombinases

One year after the landmark discovery of the DNA double helix structure, two pioneering studies established that modifications within the DNA of the bacteriophage lambda (λ) could be readily generated and transferred to the chromosome of *E. coli* (Wollman and Jacob, 1954; Kaiser, 1955), demonstrating the first targeted allelic-exchange procedure, commonly called recombination. The mechanisms behind this genetic recombination were gradually unravelled and the homologous recombination system Lambda-Red (λ-Red) was described for the first time (Echols and Gingery, 1968; Signer and Weil, 1968). This system has been extensively used in recombinational engineering or recombineering of bacteria (Murphy, Campellone and Poteete, 2000; Lesic and Rahme, 2008; Lim, Min and Jung, 2008). The λ-Red system and its application will be described in detail in **Chapter 2**. The combination of the λ-Red system with the Cre (Nagy, 2000) and FLP (Huang, Wood and Cox, 1997) site-specific recombinases, counter-selection genes (Murphy and Campellone, 2003; Warming *et al.*, 2005; Heermann, Zeppenfeld and Jung, 2008) and the I-SceI meganuclease (Kim *et al.*, 2014) has resulted in the development of genome-editing tools with significantly improved recombination efficiencies.

Following the archetype model of λ-Red discovery and application, other prophage-derived recombineering systems have been developed, like

the *E. coli* Rac prophage *recET* system (Clark *et al.*, 1984; Hall, Kane and Kolodner, 1993; Clark, Satin and Chu, 1994). The *recET* system involves the cooperation of two proteins, the endonuclease RecE and the recombinase RecT (Hall, Kane and Kolodner, 1993; Clark, Satin and Chu, 1994; Noirot and Kolodner, 1998; Muyrers *et al.*, 2000). RecET-based recombineering has been successfully applied to multiple hosts such as *E. coli* (Tolmachov *et al.*, 2006), *Pseudomonas syringae* (Swingle *et al.*, 2010) or *Zymomonas mobilis* (Wu *et al.*, 2017). Though the *recET* system is an interesting recombineering alternative to λ-Red, there is no report in literature of an HTP adaptation.

The Cre and Flp recombinases have both been utilised for efficient large-scale genome editing; for example, in *Mycoplasma mycoides* a synthetic gene successfully replaced a fragment of approximately 100 kb in the chromosome. The procedure necessitated three steps of Recombinase-Mediated Cassette Exchange, based on the activity of the Cre/lox system (Noskov *et al.*, 2015). The first example of multiplexing, the Flp-based Recombinase-Mediated Cassette Exchange system (Turan *et al.*, 2010), demonstrated that different mutated Flp recognition target sites could be recognized and recombined in parallel by the Flp recombinase. However, both the Cre and Flp recombinases require very specific target sites, which must be inserted into the chromosome prior to editing. Directed evolution studies have been utilised to generate alternative recombination sites in the chromosome and the DNA-binding domain has been engineered to broaden the recognition specificity of both enzymes (Santoro and Schultz, 2002). Despite these improvements, no ideal recombinase (e.g. with high efficiency, flexibility and low toxicity) has been identified to date.

Multiplex Automated Genome Engineering (MAGE) was developed about a decade ago in the Church lab (Wang *et al.*, 2009). MAGE relies on editing rounds, which include repeated cycles of cell growth, induction of the Red recombinase, electroporation of a pool of degenerate oligos and subsequent recovery in media (Wang *et al.*, 2009). Successive cycles enable

rapid and efficient generation of thousands of unique mutants, each representing a ubiquitous set of genomic alterations. Integration of real time monitoring, automation and software control have allowed the fine-tuning of microbial platforms towards improved productivity, e.g. the overproduction of L-DOPA (Wei, Cheng and Liu, 2016), generation of promoter libraries (Wang *et al.*, 2012) and improved ethanol tolerance (Zhang *et al.*, 2015). Many derivatives of MAGE have emerged like RAGE (Si *et al.*, 2015), Cos-MAGE (Wang *et al.*, 2012) and YOGE (DiCarlo *et al.*, 2013), accompanied by a suite of predictive algorithms (Bonde *et al.*, 2014; Quintin *et al.*, 2016). Thus, MAGE has gradually become part of mainstream metabolic engineering and enhances the close connection between biology and computing.

The application of λ-Red assisted recombination system was attempted in *C. necator* to develop a MAGE-like HTP method for large-scale generation of mutants and rapid metabolic engineering of *Cupriavidus* species.

1.3.2 Nucleases

1.3.2.1 Zing-Finger Nucleases (ZFN) and Transcription Activator-Like Effector Nucleases (TALEN)

The emergence of 'genome editing' or 'genome engineering', defined as a set of techniques that 'enable investigators to manipulate virtually any gene in a diverse range of cell types and organisms' (Gaj, Gersbach and Barbas Iii, 2013), was accomplished with the development of targeted nucleases. These engineered enzymes cleave at specific and programmable loci in the chromosome and offer a complete, efficient and quick alternative to homologous recombination (Boch *et al.*, 2009; Urnov *et al.*, 2010).

Both Zinc-finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs), induce mutagenesis, triggering gene

replacement (Carroll, 2011). Both ZFNs and TALENs are fusion DNA-binding proteins, composed of stackable motifs that individually recognise a specific sequence (DNA-binding domains) and the unspecific nuclease domain of the restriction enzyme Fok1, which causes a double-stranded break (DSB) at the targeted site (Bogdanove and Voytas, 2011; Carroll, 2011). DNA repair then progresses utilising either non-homologous end-joining (NHEJ) or homologous-directed repair (HDR). The application of these nuclease-directed engineering tools has been successful in a myriad of different hosts like fruit flies (Bibikova *et al*, 2003), zebrafishes (Ekker *et al*, 2008), worms (Wood *et al*, 2011) and rats (Tesson *et al*, 2011).

However, ZFNs are difficult to engineer and have very poor specificity, with each zinc finger domain recognising a 3- to 4-bp DNA sequence with varying specificity (Ramirez *et al.*, 2008; Gupta and Musunuru, 2014). TALENs are much easier to design and therefore tend to be more successful, since each TAL motif can recognise one single base pair (Moscou and Bogdanove, 2009). Their main advantage over ZFNs is their flexibility, targeting longer sequences through accumulation of nucleotide-specific TAL motifs (Guilinger *et al.*, 2014). TAL motifs are also more cost-effective to produce and less toxic to the cells. TALENs can, however, lead to unintentional mutations, specifically when binding domains are composed of very similar sequences (Hockemeyer *et al.*, 2011). Therefore, off-target cleavage remains a significant problem in both TALEN and ZFN systems (Joung *et al.*, 2013). Thus, both ZFNs and TALENs are not ideally suited to efficient multiplexgenome engineering.

1.3.2.2 The CRISPR/Cas9 System

Following the development of ZFNs and TALENs, the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas (CRISPR-associated) system has recently revolutionised the sphere of genomic

biology and is now firmly established as the ultimate choice for genome editing. The CRISPR/Cas9 system is efficient, fast to implement and functional in nearly all of the systems tested so far (Doudna and Charpentier, 2014).

CRISPR sequences were initially discovered in 1987 (Ishino *et al.*, 1987). In the analysis of the *iap* gene, coding for the isozyme alkaline phosphatase in *E. coli,* it was noted that the downstream sequence consisted of five dyad symmetrical repeats of 14 bp, interspaced with a variable 32-nucleotide (nt) sequence between each repeat (Ishino *et al.*, 1987). The team did not realise the significance of their discovery, since they had sequenced the first five of the 14 repeat sequences of the CRISPR locus, located downstream the *iap* gene and upstream the Cas complex. CRISPR was not fully discovered until 2005 with the onset of the genomic era. Two independent teams, Francisco Mojica in Spain and Christine Pourcel in France had simultaneously a Eureka! moment when they established the homology between these variable 32-nt sequences, named spacers, and plasmid or phage sequences (Mojica *et al.*, 2005; Pourcel, Salvignol and Vergnaud, 2005).

The next big discovery came from the dairy industry. Rodolphe Barrangou from Danisco® (now DuPont®) and Helene Deveau from the University of Laval in Canada, sequenced the genome of *Streptococcus thermophilus*, a workhorse for yoghurt and cheese production (Barrangou *et al.*, 2007; Deveau *et al.*, 2008). They proved that CRISPR activity could be assimilated to a bacterial 'immune system', with the spacer sequences acting as a 'memory' of previous encounters with exogenous phages. The CRISPR study subsequently became largely focused on the *Streptococcus* genus. With their publication, Jenifer Doudna and Emmanuelle Charpentier not only demonstrated that the dual-RNA/Cas9 ribonucleoprotein complex can cleave effectively, they also paved the way for the use of CRISPR/Cas9 as a powerful genome editing tool (Jinek *et al.*, 2012). It was initially utilised

in the rearrangement of human nerve and mouse kidney cells, directing the bacterial system against the genome of the host cell line (Cong *et al.*, 2013; Mali *et al.*, 2013). The CRISPR/Cas system is now regarded as a quick and easy-to-use method with high precision, efficiency, flexibility and multiplexing (Mali *et al.*, 2013). Furthermore, it is functional in a diverse range of species including fruit flies (Bassett and Liu, 2014), zebrafish (D'Agostino *et al.*, 2016), human cell lines (Mali *et al.*, 2013) and bacteria (Jiang *et al.*, 2015).

An increasing number of CRISPR/Cas systems have been identified and characterised and a classification has been progressively adopted (Ishino, Krupovic and Forterre, 2018). The CRISPR/Cas systems are organised in two classes including six types and over 20 subtypes (Shmakov *et al.*, 2017). The main classification criterion was based on the effector protein: class 1 includes multimeric nuclease complexes, while class 2 systems are based on a unique effector component (**Figure 1.3**) (Makarova *et al.*, 2015; Shmakov *et al.*, 2015). Each class includes three types (types I, III and IV for class 1 and types II, V and VI for class 2) which are distinguished by the signature *cas* genes (Makarova *et al.*, 2015). For example, type I is associated with the Cas3 nuclease (Westra *et al.*, 2012), type II with the Cas9 (Sapranauskas *et al.*, 2011) and type III with the Cas10 nuclease (Samai *et al.*, 2015). Each type is subdivided into a number of subtypes, identified with a letter code (for example type I includes the seven subtypes A, B, C, D, E, F and U) according to the organisation of the CRISPR loci, the CRISPR adaptation system or the sequence similarity between multiple shared Cas proteins (Makarova *et al.*, 2015).

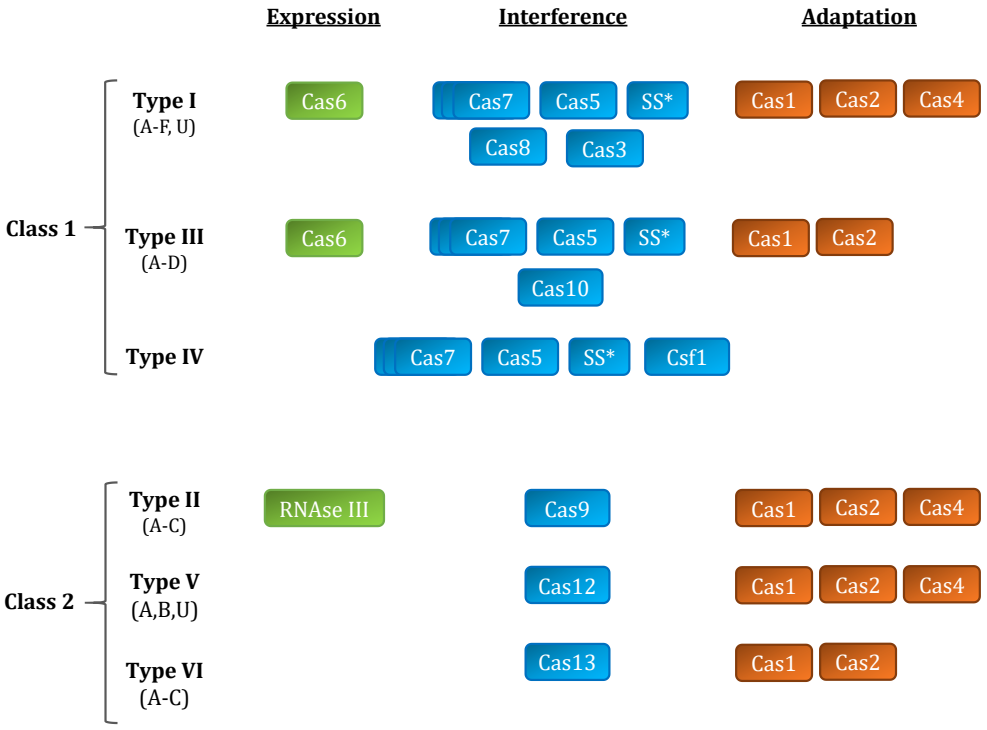


Figure 0.3 – Classification of CRISPR/Cas systems

CRISPR/Cas systems were organized into classes, types and subtypes based on the analysis of the sequence and gene organisation of the Cas proteins as they were being discovered. The number of effector components distinguished class 1 (multimeric) from class 2 (single nuclease) while the analysis of the nature and function of the interference complex (and sometimes the adaptation proteins) provided distinct types and subtypes of which the subtype II-A is the most characterised of them. Adapted from the great History of CRISPR (Ishino, Krupovic and Forterre, 2018).

The most studied Cas9 nuclease belongs to the type II-A of class 2 (Shmakov *et al.*, 2017) and the type II CRISPR/Cas9 system from *S. pyogenes* remains the most widely used genome editing tool (Jinek *et al.*, 2012). Nonetheless, advances in understanding the cleavage mechanism of the type V Cpf1 nuclease (Zetsche *et al.*, 2015) and recent demonstration of RNA cleavage driven by the enzyme has shed light on this other class 2 effector system. It enables strains to be engineered where the expression of Cas9 had previously been toxic and more strain development is expected with this alternative CRISPR/Cas system (Ungerer *et al.*, 2012; Jiang *et al.*, 2017; Li *et al.*, 2018).

The CRISPR/Cas9 system requires three components, assembled in a complex: a single endonuclease Cas9 and two small RNAs (**Figure 1.4**). The CRISPR-associated RNA (crRNA), includes the sequence of one of the spacers encoded in the CRISPR array (Brouns *et al.*, 2008). This sequence is the genomic portion of a viral or invader plasmid, previously encountered by the cell and stored in the CRISPR array as an 'immune memory' (Haurwitz *et al.*, 2010). The crRNA is transcribed from the CRISPR locus and matured via partially binding to the transactivating RNA (tracrRNA), followed with cleavage by the RNA III polymerase (Dugar *et al.*, 2013). The tracrRNA is typically formed of two palindromic loops, which offer a scaffold for the Cas9 protein. The ribonucleoprotein complex can target any sequence, named the protospacer, which needs to be complementary to the spacer. The ribonucleoprotein will cleave within the protospacer sequence as long as the sequence is directly followed at the 3'end by a Cas9-specific protospacer adjacent motif (PAM) (Gasiunas *et al.*, 2012). The *S. pyogenes* Cas9 specifically recognises the PAM sequence 5'-NGG-3' (Mali, Esvelt and Church, 2013).

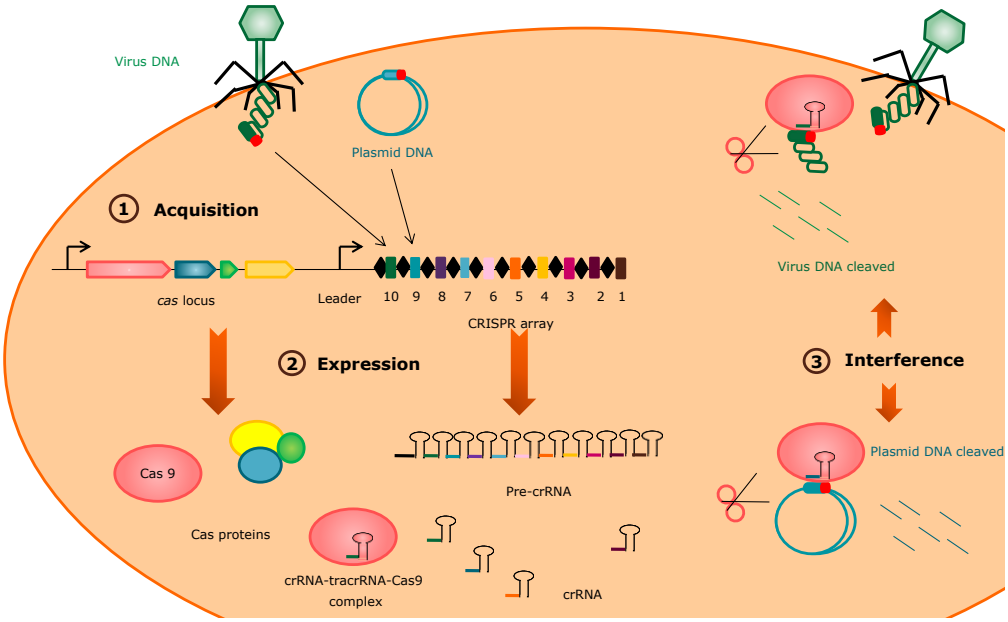


Figure 0.4 – Overview of the type II CRISPR native immune system

1. Acquisition: when a bacterial virus or an exogenous plasmid enters the cell, the bacterium triggers defence systems that degrade the invaders. Fragments of their DNA is stored in a CRISPR array as spacers (coloured boxes) separated from one another by repeats (black diamonds). The most recently integrated spacer (10) is located near the leader sequence.
2. Expression: in the array, the leader sequence is transcribed in a long pre-CRISPR-RNA (crRNA) that matures into an RNA duplex crRNA/transactivating RNA (tracrRNA) after cleavage by the RNase III. Simultaneously, the Cas proteins are expressed from the *cas* locus, including the endonuclease Cas9, which will form a ribonucleoprotein complex with the crRNA/tracrRNA duplex.
3. Interference: These newly assembled complexes are released into the cytoplasm and scan any piece of DNA encountered until their crRNA matches a complementary sequence. If this sequence is followed by a PAM motif (5'-NGG-3' for Cas9 of *S. pyogenes*), the nuclease protein cleaves the DNA, generating a DSB lethal for the invader.

To repurpose the system for genome engineering, the crRNA and tracrRNA can be fused into an engineered single guide RNA (sgRNA), a transcript of approximately 100 bp (Jinek *et al.*, 2012). The sgRNA specifically anneals to its complementary chromosomal sequence (spacer-protospacer binding) and recruits the Cas9 nuclease to cleave DNA, forming a DSB (Jinek *et al.*, 2012). The scar is repaired through either non-homologous end joining (NHEJ) or homology-directed repair (HDR). The first of these repair mechanisms, NHEJ, relies on a compact two-protein system: the homodimer Ku and the multifunctional polymerase/ligase, LigD

(Bowater and Doherty, 2006). It uses micro homologies at the end of DSBs to guide direct repair (Della *et al.*, 2004). However, it is an error-prone pathway, since the LigD protein often inserts or removes bases (resulting in indels), while filling the gap, thus leading to frameshift mutations and premature stop codons (Lieber, 2010). HDR requires a homologous template and results in precisely defined mutations, allowing knock-outs as well as knock-ins or single-base mutations to be generated (Wigley, 2013).

CRISPR/Cas9 has been successfully applied in gene deletion (Huang *et al.*, 2015), pathway insertion (Li *et al.*, 2015) and single point mutation (Arazoe, Kondo and Nishida, 2018). Multiplexgenome editing has been successful in both model chassis e.g. *E. coli, S. cerevisiae* (Generoso *et al.*, 2016), *Streptomyces* (Cobb, Wang and Zhao, 2015; Jiang *et al.*, 2015) and non-model chassis, like *Clostridium beijerinckii* (Wang *et al.*, 2015) and *Lactobacillus reuteri* (Oh and van Pijkeren, 2014). Additionally, the recent advent of HTP CRISPR-based technologies has considerably expanded the toolbox for synthetic and systemic engineering strategies. For instance, the no-SCAR (Scarless Cas9 Assisted Recombineering) system developed for genome editing in *E. coli,* combined both λ-Red recombination, to facilitate the chromosomal integration of a single-stranded (ss) or double-stranded (ds) DNA and CRISPR to counter select non-edited cells (Reisch and Prather, 2015). Similarly, the CRISPR Optimized MAGE Recombineering (CRMAGE) utilises CRISPR and MAGE to create a fast, precise and highly efficient (96.5-99.7%) multiplex cyclic editing method, which has been utilised in *E. coli* (Ronda *et al.*, 2016).

CRISPR derived systems like CRISPR interference, CRISPR activation and CRISPR/Cpf1 have also been adapted for multiplexing and fine-tuning gene expression and regulation and for screening large mutant libraries (Cheng *et al.*, 2013; Zetsche *et al.*, 2015; Fulco *et al.*, 2016; Li *et al.*, 2018).

CRISPR systems are found in about 45% of bacteria and 85-87% of archaea (Shmakov *et al.*, 2017; Ishino, Krupovic and Forterre, 2018), however they are thought to be less prevalent in environmental communities compared to laboratory cultivars (Burstein *et al.*, 2016). Both *C. necator* H16 and *C. metallidurans* CH34 lack any CRISPR systems. The implementation of the CRISPR/Cas9 technique was adapted for *C. necator,* as a proof of concept that *Cupriavidus* species are amenable for CRISPR/Cas9-based genome editing.

1.3.3 Transposons and Mobile Group II introns

Other genome editing techniques include transposons and group II introns. Transposons are mobile elements which are inserted into chromosomes via transposase activity (Simon, Priefer and Pühler, 1983). Transposons can be utilised to promote the stable insertion of foreign DNA into the host chromosome (Reznikoff, 2008; Martínez-García *et al.*, 2014). For example, Tn*5*-derived transposons have been developed to generate insertion mutants via the random introduction of genes specifying resistance to selected antibiotics into the chromosome of a variety of gram-negative bacteria (de Lorenzo *et al.*, 1990). Because insertion is random and often occurs in a coding sequence, the use of mini-Tn*5* is helpful to investigate gene essentiality in an organism, which has been exploited in the HTP analysis technique of Transposon Directed Insertion Sequencing (TraDIS) (Barquist *et al.*, 2016). If site- and orientation-specific is desired, the family of mini-Tn*7* vectors allows insertion at a unique intergenic site in the chromosome. Nonetheless, transposon-based tools are more dedicated to investigation of gene function (de Lorenzo *et al.*, 1990), study of protein-DNA interaction (Reznikoff, 2008) and genomic engineering of strains where antibiotic selection is not feasible (in biofilms, for example) (Choi *et al.*, 2005) than engineered to elicit phenotypes of interest in microorganisms.

Similarly, mobile group II introns are capable of chromosomal transposition by intron-encoded protein-dependent DNA target site recognition and the utilisation of a reverse splicing mechanism (Lambowitz and Zimmerly, 2011). The insertion site is determined by base-pairing, therefore, it is possible to re-target the intron to any desired chromosomal location (Karberg *et al.*, 2001). To ensure correct insertion, introns must be combined with gene modifying enzymes e.g. recombinases, allowing high editing efficiencies to be achieved (Jia *et al.*, 2011). The GETR (Genome Editing via Targetrons and Recombinases) platform was utilised in *E. coli*, allowing the insertion of a 12 kb operon, the deletion of 120 kb of DNA and the inversion of a 1.2 Mb sequence in *E. coli* (Enyeart *et al.*, 2014). A group II intron-based knockout system, RalsTron, has been developed in *C. necator* and allowed the knockout of the *phaC* gene (Park *et al.*, 2010) but the identification of a successful intron integration site remains difficult and such a system would require intensive optimisation of both the system and the algorithm to predict insertion sites that would result in efficient gene disruption (Yao and Lambowitz, 2007). In addition, the intron off-target rate and the requirement for recombinases limits the use of these elements as a robust genomic editing strategy (Enyeart *et al.*, 2014).

Finally, both transposons and introns have the potential to complement HTP technologies in the genomic editing field (introns have recently been used to express multiple sgRNAs in a Cas9 and Cpf1-based editing system (Ding *et al.*, 2018)) but they cannot be utilised effectively as HTP editing tools, individually.

In summary, there are a range of validated engineering tools utilised in a broad scope of model organisms but none of them is suitable for fast and efficient genome engineering of *Cupriavidus* species. The utilisation of prophage-derived recombinases is often restricted to strains closely related to *E. coli* while ZFNs and TALENs require extensive and costly design. Transposons are mostly dedicated to the discovery of essential genes and

mobile group II introns still show low editing efficiencies. However, some of these tools could be repurposed to be exploited in *Cupriavidus* spp. In particular, recombinases and CRISPR/Cas systems, that are the most amenable techniques for automated and multiplexgenome engineering, could greatly contribute to the development of advanced tools for the engineering and reshaping of metabolic networks in *Cupriavidus* species.

1.4 Towards ethylene production in *Cupriavidus* species

Developing HTP tools for the genome engineering of *Cupriavidus* species would enable the implementation of non-native metabolic pathways to turn these chassis into highly valuable industrial platforms to produce chemicals, which are traditionally produced via processes deleterious for the environment.

1.4.1 Ethylene, a major bulk molecule in the chemical industry

Ethylene is the simplest member of the alkene family, C_2H_4 or $CH_2=CH_2$ (Zimmermann and Walzl, 2009). It is a colourless, volatile gas with a density close to that of dioxygen. The presence of the double bond makes it a very reactive gas species; it is also used in many chemical transformations catalysed by transition metals (Ziegler and Rauk, 1979).

Ethylene is the most widely produced monomer in the petrochemical industry with a production culminating slightly above 150 M metric tons in 2016 and expected to keep climbing up to 200 M tons by 2020 (Eramo, 2012; Lewandowski, 2016). Ethylene derivatives (e.g. polyethylene, ethylene dichloride, ethylbenzene, ethylene glycol, ethylene oxide and (poly)vinyl chloride) are heavily used as bulk components in the manufacture of many polymers and plastics for textile, packaging, electronic and construction industries (**Figure 1.5**).

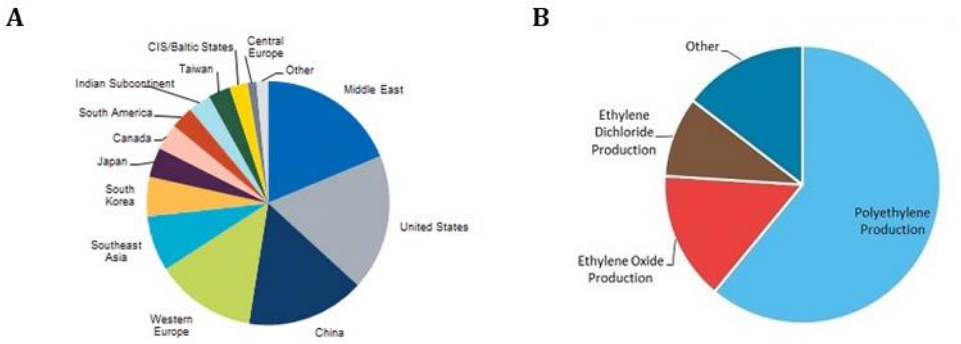


Figure 0.5 – Global ethylene market

A. World consumption of ethylene in 2016, from the Ethylene Chemical Economics Handbook (HIS Markit® study).
B. Global ethylene demand by application, from the report *Ethylene (ET): 2018 World Market Outlook and Forecast up to 2027* (Merchant Research & Consulting Ltd©).

1.4.2 Ethylene production

1.4.2.1 Petrol to ethylene: steam cracking

Ethylene is produced commercially by the steam cracking (SC) of hydrocarbon feedstocks such as naphtha and ethane (Johansson, 2014). The SC process includes heating a mix of steam and hydrocarbons up to 750-950°C, which enables the cracking of hydrocarbons into smaller block molecules like ethylene. After gas quenching, several repetitive cycles of compression and distillation are necessary to purify ethylene (Kniel, Winter and Stork, 1980).

Therefore, SC releases large quantities of CO_2 into the atmosphere, which is detrimental to the environment (Ghanta, Fahey and Subramaniam, 2014). Fuel burning, necessary to produce steam, is responsible for 78-93% of the environmental impact of the whole process (Ghanta, Fahey and Subramaniam, 2014). SC-generated ethylene is responsible for greenhouse gas releases, estimated at 1.2-4.4 tCO_2eq/t (ton of CO_2 equivalent/ton) ethylene (Patel, 2006), air acidification and metal emissions and represents 1.5% of the US carbon footprint (Worrell *et al.*, 2000). Furthermore, in the presence of solar UV (e.g. in daylight), ethylene has been shown to

contribute to depletion of the ozone layer, contributing to global warming (De Gouw and Warneke, 2005). Though process optimisation is being continuously conducted (He and You, 2015), increasing ethylene demand will continue to put pressure on existing production facilities. Although SC generates considerable levels of CO_2, it remains a cheap and economically viable process, which has only recently seen the implementation of alternative techniques (Amghizar *et al.*, 2017). These techniques aim to convert methane catalysed from syngas into olefins as methane prices are generally low and facilitate cost-competitive processes (Kee, Karakaya and Zhu, 2017). However, these techniques still employ drastic manufacturing conditions (heat, pressure, controlled steam dilution rate, etc) and their efficiency remains low compared to that of SC, which makes their commercialisation still challenging (Amghizar *et al.*, 2017).

1.4.2.2 Ethanol to ethylene: catalytic dehydration

Sustainable routes to ethylene production are imperative and the catalytic dehydration of bioethanol (from biomass) to ethylene, could provide such a solution (George W. Huber, Sara Iborra, and Corma 2006). However, this partially greener alternative still requires elevated temperatures (300−500°C) high pressure (0.1−0.2 MPa) and the presence of alumina catalysts (Morschbacker, 2009). Yields can reach 94−99%, the process has been scaled up to an industrial production level and 0.3% of the global capacity of ethylene is produced from bioethanol (Fan, Dai and Wu, 2012; Broeren, 2013). Bioethylene facilities have been set up in India and China but the biggest production unit (Braskem®) is located in Brazil with a production capacity of 200 kt per annum, benefiting from the regional availability of low-cost sugar cane, the main feedstock of bioethanol production (Broeren, 2013). Despite an undeniable economic viability, this process still releases significant levels of CO_2 into the atmosphere, estimated between 0.7 and 2.5 t CO_2eq/t ethylene (Patel, 2006). The extensive culturing of sugar cane also has detrimental environmental

consequences in terms of air pollution due to straw burning, soil pollution resulting from the overuse of pesticides and depletion of rainforest land (Thorburn *et al.*, 2011; Liboni, 2012).

1.4.2.3 Biosynthesis of ethylene: the potential of biotechnological processes

Ethylene is also an important plant hormone involved in maturation, ripening, loss of chlorophyll and leaf abscission (Li *et al.*, 2017). Three biological ethylene pathways have been identified (**Figure 1.6**).

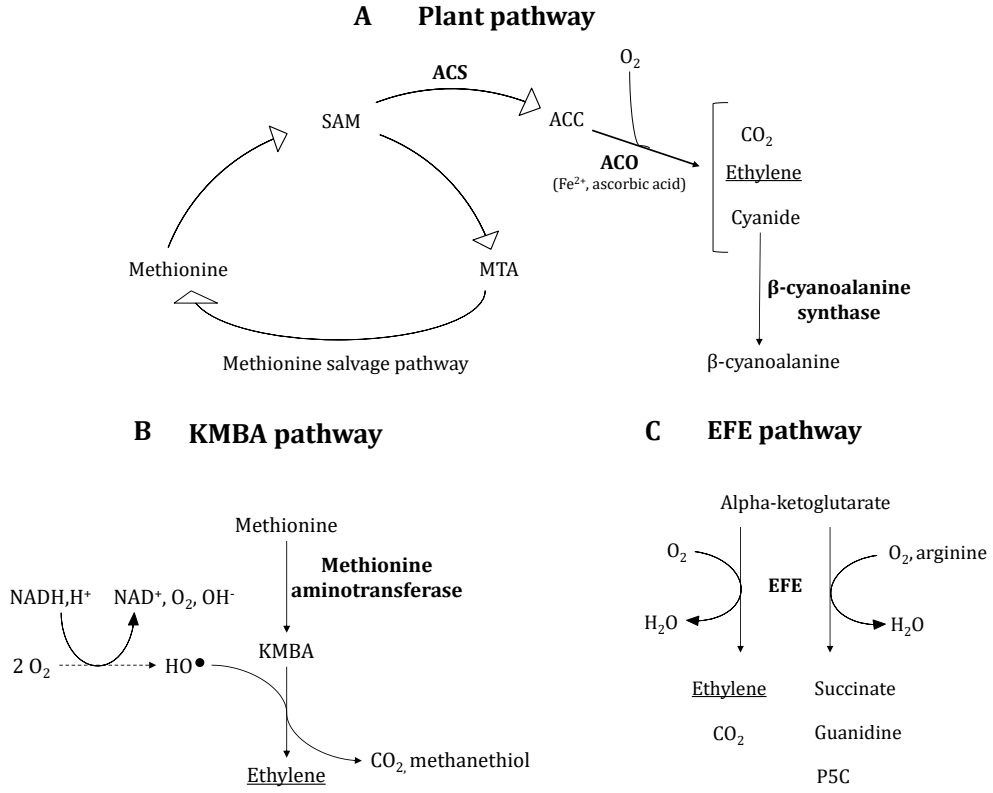


Figure 0.6 – Biological ethylene synthesis pathways

A. The plant pathway producing ethylene from methionine.
B. The 2-keto-4-methylthiobutyric acid (KMBA) pathway producing ethylene from methionine.
C. The Ethylene-Forming Enzyme (EFE) pathway producing ethylene from alpha-ketoglutarate.
SAM: S-adenosylmethionine; ACC: 1-aminocyclopropane-1-carboxylate; MTA: methylthioadenosine; ACO: ACC oxidase; ACS: ACC synthase and P5C: L-Δ-1-pyrroline-5-carboxylate.

A. The plant pathway

The first pathway to have been characterised is the plant pathway producing ethylene from methionine (Lieberman *et al.*, 1965). Methionine is converted to S-adenosylmethionine (SAM) by the ATP-dependent SAM synthetase (Catoni, 1953) then SAM is degraded into 1-aminocyclopropane-1-carboxylate (ACC) and methylthioadenosine (MTA) by the ACC synthase (ACS) (Adams, Yang and Stumpf, 1979). Finally, ACC is oxidised into ethylene, CO_2 and cyanide by the ACC oxidase (ACO) in presence of O_2, Fe^{2+} ions and ascorbic acid. Cyanide is detoxified into beta-cyanoalanine by the beta-cyanoalanine synthase (Miller and Conn, 1980). Moreover, MTA is recycled into the methionine salvage pathway (Wang, Adams and Lieberman, 1982) which allows methionine recycling through ethylene biosynthesis.

B. The KMBA pathway

Some plant-associated fungi and bacteria utilise the alternative 2-keto-4-methylthiobutyric acid (KMBA) ethylene synthesis pathway, initially characterised in *Cryptococcus albidus* (Fukuda *et al.*, 1989) and *E. coli* (Ince and Knowles, 1986). The first reaction of this two-step pathway involves the conversion of methionine to KMBA via a specific transaminase (Ince and Knowles, 1986). KMBA is then spontaneously oxidised into ethylene depending on the generation of a reactive hydroxyl radical (OH^-) by a NADH:Fe(III)EDTA oxidoreductase (Ogawa *et al.*, 1990).

C. The EFE pathway

The third route for bio-ethylene is 2-oxoglutarate (2-OG)- or α-ketoglutarate (AKG)-dependent pathway and requires the presence of arginine and Fe^{2+} ions as co-factors (Nagahama *et al.*, 1991). It has been studied thoroughly in bacteria and in cell-free systems extracted from the fungus *Penicillum digitatum* (Fukuda, Fujii and Ogawa, 1986) and the

bacterium *Pseudomonas syringae* pv. *phaseolicola* PK2 Kudzu strain isolated from *Pueraria lobate* (a root plant commonly named Kudzu) (Goto and Hyodo, 1987). Bioproduction of ethylene was also noted in a member of the *Ralstonia* (now *Cupriavidus*) genus, *R. solanacerum,* but *P. syringae* produced ethylene at a rate (7 to 100×10^{-9} nL.cell^{-1}.h^{-1}) 500 to 1,000 times higher than that of *R. solanacerum* and several times higher than that of *P. digitatum* (Goto *et al.*, 1985).

Therefore, the *P. syringae* ethylene-forming enzyme (EFE) was purified from the organism and carefully studied, resulting in the establishment of a dual-circuit mechanism for the simultaneous formation of ethylene and succinate from AKG (Fukuda *et al.*, 1992). The enzyme was associated to the Fe(II)- and AKG-dependent oxygenase superfamily (Hausinger, 2015) and catalyses the hydroxylation of arginine and the AKG decarboxylation that forms ethylene, succinate, guanidine, L-Δ-1-pyrroline-5-carboxylate (P5C) and CO_2. In the last 2/3 years, crystallisation of EFE at high resolution (Martinez and Hausinger, 2016; Martinez *et al.*, 2017; M. Li *et al.*, 2018) and mutagenesis investigation enabled biochemical characterisation of the reaction mechanism, while calorimetric studies looked at the thermodynamics of the catalytic activity of EFE. These landmark structural analyses revealed significant insights into the mechanism of this fascinating enzyme, which could pave the way towards directed enzyme evolution approaches.

Meanwhile, the *P. syringae* EFE enzyme has been expressed in diverse chassis (*E. coli* and *S. cerevisiae*) aiming at implementing the bioprocess at industrial scale. Promising development is being conducted in cyanobacteria, which can use light to fix CO_2 into ethylene, offering a sustainable strain platform towards carbon-neutral production (Ungerer *et al.*, 2012; Xiong, Morgan and Ungerer, 2015; Mo *et al.*, 2017; Puthan Veetil, Angermayr and Hellingwerf, 2017). In the same perspective, exploiting the

lithoautotrophy of *Cupriavidus* species to develop an ethylene production could renew the industrial and commercial potential of these organisms.

1.5 Aim and objectives

C. necator has been utilised in large-scale PHAs fermentation for the last 40 years and efforts are being made to expand the synthetic biology toolbox for this organism and divert its natural carbon flux potential towards different molecules of industrial value. *C. metallidurans* shows a fascinating multi-resistance to heavy metals at high concentrations, and early development of medium-scale equipment dedicated to bioremediation of wastewater utilising the metal sequestration properties of the bacterium is being conducted. However, the engineering of both *Cupriavidus* species remains difficult, time-consuming and inefficient. Therefore, there is a great need for novel, fast, efficient and user-friendly HTP genomic tools suitable to these non-model chassis. The present work describes the development of novel methods of genome editing to facilitate the metabolic engineering of *Cupriavidus* spp. To illustrate the application of these methods, this study also presents the implementation and optimisation of the EFE ethylene pathway in *C. metallidurans.*

The development of HTP editing tools were conducted in *C. necator* only, since this species is better characterised and benefits from a more extensive range of genetic tools than *C. metallidurans*; the knowledge established with this species is expected to speed up the method development process. Nonetheless, the genomic modification techniques used here were designed with the perspective of extending their application to other *Cupriavidus* species, especially to *C. metallidurans*. **Chapter 3** presents the assessment of the λ-Red and RecET-assisted recombineering combined with the expression of the I-SceI endonuclease in a two-step scarless DNA editing method.

Despite many attempts to adapt the method to *C. necator*, it was not possible to generate a mutant strain using the phage recombinase-based systems and alternative editing tools were explored. **Chapter 4** describes

the development and optimisation of a second type of genome editing technology with great HTP potential: the CRISPR/Cas9 system.

Chapter 5 illustrates the implementation and optimisation of the ethylene biosynthesis pathway in *C. metallidurans*. To boost the strain engineering efficiency, the semi-HTP global Transcription Machinery Engineering (gTME) was established along with the Adaptive Laboratory Evolution (ALE). Both methods were first applied to the optimisation of ethylene production in *E. coli* as proof of concept.

In summary, this study aims to develop novel HTP genome editing tools for the implementation and optimisation of the EFE ethylene synthesis pathway. The specific objectives of this work are to:

- Develop λ-Red and *recET*-based homologous recombination systems in *C. necator*
- Adapt the CRISPR/Cas9 system in *C. necator*
- Engineer *C. metallidurans* for optimal ethylene production via semi-automated gTME and ALE techniques.

Materials and Methods

2.1 Media and antibiotic concentrations

Lysogeny broth (LB, Sigma-Aldrich®) medium (Bertani, 1951) was used for general laboratory growth supplemented, when required, with antibiotics at concentrations suitable for the particular bacteria being utilised (**Table 2.1**). When necessary, plates and liquid cultures were supplemented with 10 mM arabinose, 0.2% (w/v) rhamnose, 2 mM salicylic acid or 5 mM theophylline. For solid medium, 15 g/L agar was added.

Table 0.1 - Working concentrations of antibiotics for the culture of the organisms studied in this work

Antibiotic	*E. coli*	*C. necator*	*C. metallidurans*	Abbreviation
Ampicillin	100 μg/mL	100 μg/mL	_	LBA
Chloramphenicol	25 μg/mL	50 μg/mL	250 μg/mL	LBC
Kanamycin	50 μg/mL	300 μg/mL	_	LBK
Spectinomycin	100 μg/mL	300 μg/mL	300 μg/mL	LBS
Tetracycline	15 μg/mL	15 μg/mL	_	LBT
Trimethoprim	10 μg/mL	60 μg/mL	_	LBTm

2.1.1 Super Optimal Broth (SOB) or Hanahan's Broth

Super Optimal Broth (SOB, Hanahan's Broth, Sigma-Aldrich®) was utilised to make electrocompetent *Cupriavidus* cells (Tee *et al.*, 2017).

2.1.2 Minimum medium

Minimum medium (MM) (Schlegel, Kaltwasser and Gottschalk, 1961) was used to select *Cupriavidus* transconjugants after conjugation with *E. coli* S17-1 λ*pir* and contained 25 mM Na_2HPO_4 x 12 H_2O, 11 mM KH_2PO_4, 18.7 mM NH_4Cl, 0.8 mM $MgSO_4$ x 7 H_2O, 0.14 mM $CaCl_2$ x 2 H_2O, 4.6 μM Fe (III) NH_4-Citrate and 1 mL/L SL7 solution. SL7 was composed of 0.0325% (w/v) HCl, 1 mM H_3BO_3, 0.8 mM $CoCl_2$ x 6 H_20, 0.1 mM $CuCl_2$ x 2 H_2O , 0.5 mM $MnCl_2$ x 4 H_2O, 0.15 mM Na_2MoO_4 x 2 H2O, 0.1 mM $NiCl_2$ x 6

H_2O and 0.5 mM $ZnCl_2$ (Biebl and Pfennig, 1981). MM was supplemented with 0.4% (w/v) sodium gluconate (0.4% SG-MM).

2.1.3 Low Salts Lysogeny Broth (LSLB)

Low Salts Lysogeny Broth (LSLB) (Jugder *et al.*, 2016) was used for the counterselection of transconjugants following conjugation in *Cupriavidus* strains using a mobilizable suicide plasmid. It contained 5 g/L NaCl, half the salt of LB medium, and was supplemented with 15% (w/v) sucrose for *sacB* counterselection.

2.1.4 Minimal Salts Medium (MSM)

Minimal Salts Medium (MSM) was used to grow *Cupriavidus* species for ethylene detection (Lindenkamp, Volodina and Steinbüchel, 2012). It was composed of 3.746 g/L K_2HPO_4, 1.156 g/L KH_2PO_4, 0.962 g/L NH_4Cl, 0.702 g/L NaCl, 66 mg/L citric acid, 16.68 mg/L $FeSO_4$ x $7H_2O$, 0.1 mg/L $ZnCl_2$, 0.03 mg/L $MnCl_2$ x $4H_2O$, 0.05 g/L $CoCl_2$ x$6H_2O$, 0.07 mg/L $CuCl_2$ x $2H_20$, 0.1 mg/L $NiCl_2$ x $6H_2O$, 0.03 mg/L Na_2MoO_2 x $2H_2O$, 0.05 mg/L $CrCl_3$ x $6H_2O$, 0.3 mg/L H_3BO_3, 11 mg/L $CaCl_2$ and 240 mg/L $MgSO_4$. The following carbon sources were tested in this medium for the growth of *Cupriavidus* strains expressing *efep*: sodium acetate 5, 10 and 15 mM (Im *et al.* 2017; Lenz, 1998), sodium benzoate 5 mM (Ampe and Lindley, 1995), sodium gluconate 0.4% (w/v) (Nies *et al.*, 1987), sodium lactate 5 mM (Yan, Du and Chen, 2003) and sodium succinate 0.3% (Rojas *et al.*, 2011). When required, arginine and alphaketoglutarate (AKG) were supplemented to the medium at 3, 10 or 20 mM and 2, 10 or 20 mM respectively (Lynch *et al.*, 2016).

2.1.5 M9 minimal medium

M9 minimal medium was used to grow *E. coli* K-12 MG1655 and *ΔproB* strains during the gTME and ALE experiments. It contained 40 mM Na_2HPO_4 x $2H_2O$, 22 mM KH_2PO_4, 8.6 mM NaCl, 18.7 mM NH_4Cl, 2 mM $MgSO_4$, 100 μM $CaCl_2$ and 0.4% (w/v) glucose.

2.1.6 Determination of minimal inhibitory antibiotic concentrations (MICs) using the E-test

C. necator H16 was tested for sensitivity to trimethoprim and tetracycline (used as a control) in LB and 0.4% SG-MM. Five colonies were isolated and grown in 3 mL of either LB or 0.4% SG-MM at 30°C and 200 rpm until the inoculum reached an $OD_{600nm} \sim 0.1$. When the OD value was higher than 0.1, the culture was diluted in PBS to a final OD of 0.1 according to the following formula:

$$Vinoc = ODfinal\ x\ Vfinal \Big/ (OD\ inoc - ODfinal)$$

A sterile swab was then soaked in the inoculum suspension. The excess fluid was removed by pressing it against the inside wall of the test tube and the entire agar surface was carefully streaked three times, rotating the plate 60°C each time to evenly distribute the inoculum. The surface was completely dry before application of the Etest gradient strips. The plates were then incubated at 30°C for 48 h before the MIC were calculated. The experiment was carried out in triplicate.

2.1.7 General chemicals

Except otherwise stated, all chemical compounds were purchased from Sigma-Aldrich.

2.2 Biological manipulation

2.2.1 Bacterial strains, plasmids and strains

All strains and plasmids used in this study are listed in **Table 2.2**. All primers used in this study are given in **Appendices**.

Table 0.2 – Bacterial strains and plasmids used in this study

* indicates that the associated strain is a product of this study

Name	Relevant genotype or description	Source or reference
E.coli		
DH5α	F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsdR*17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1*gyr*A96 *rel*A1	CRG culture collection
Top10	F– *mcrA* Δ(*mrr-hsdR*MS-mcrBC) Φ80*lacZ*ΔM15 Δ*lacX*74 *rec*A1 *araD*139 Δ (*ara leu*) 7697 *gal*U *gal*K *rpsL* (StrR) *end*A1 *nup*G	CRG culture collection
C2987	*fhu*A2 D(*argF-lacZ*) U169 *pho*A *gln*V44 f80D(*lacZ*)M15 *gyrA*96 *rec*A1 *rel*A1 *end*A1 *thi*-1 *hsdR*17	NEB, Ipswich, USA
S17-1 *λpir*	*recA, thi, pro*, *hsdR*-M+RP4: 2-Tc: Mu:Km Tn7 *λpir*, TpR SmR	CRG culture collection
K12 MG1655	*K-12 F– λ– ilvG– rfb-50 rph-1*	Joshua Petch, University of Nottingham
Δ*proB* K12-MG1655	*K-12 F– λ– ilvG– rfb-50 rph-1 ΔproB*	Alexander Van Hagen, SBRC
***Cupriavidus* spp.**		
H16	H16 wild type	DSM-428 DSMZ, Braunschweig, Germany
Δ*phaC*	H16, Δ*phaC*	Christian Arenas
Δ*recJ*	H16, Δ*recJ*	*
CH34	CH34 wild type	Samantha Bryan, SBRC Nottingham
General plasmids		
pME6000	pBBR1 *tetR/tetA lacZa*	Stephan Heeb, University of Nottingham
pME6031	pVS1-p15A RK2 *tetR/tetA* MCS T4 term	Stephan Heeb, University of Nottingham
pUC19	pMB1 *bla lacZa*	CRG culture collection
pMMG	pVS1-p15A *tetR/tetA gfp*	Stephan Heeb, University of Nottingham
pMTL71101	pBBR1 *catP* mob *C. gilardii* CR3 *recE recT C. metallidurans CH34 dnaM*	Samantha Bryan, SBRC Nottingham
pMTL71401	pBBR1 *aph(3')-II* mob	Muhammad Ehsaan, SBRC Nottingham

pEH042	pBBR1 *catP mob nahG* (H16_RS08130) pSal-rfp	Erik Hanko, SBRC Nottingham
Recombineering plasmids		
pSLTS	repA101(Ts) *bla araC* pBAD-*Red bet, exo and gam* ptet-*I-SceI*	(Kim *et al.*, 2014)
CmR-pSLTS	repA101(Ts) *catP araC* pBAD-*Red bet, exo and gam* ptet-I-*SceI*	*
pSLVS1	pVS1-p15A *catP araC* pBAD-*Red bet, exo and gam* ptet-*I-SceI*	*
pSLBBR1	pBBR1 *catP araC* pBAD-*Red bet, exo and gam* ptet-*I-SceI*	*
p*recET-C*	pVS1-p15A *catP araC* pBAD-*recET-dnaM (C.necator* JMP134) ptet-*I-SceI*	*
precET-P	pVS1-p15A *catP araC* pBAD-*recET(P.syringae pv. syringae B728a*) ptet-*I-SceI*	*
pSV200	pMB1 *bla* MCS *HsdR(H16)* HA (100/50bp) pamp-*dhfR* I-SceI site	*
pSV210	pBBR1 *tetR/tetA HsdR*(H16) HA(100/50bp) pamp-*dhfR I-SceI site*	*
pSV242	pBBR1 *tetR/tetA phaC* (H16) HA (950/50bp) *aadA*- I-SceI site	*
pSV442	pMB1 *tetR/tetA phaC* (H16) HA (950/50bp) *aadA*- I-SceI site	*
pLO3	pBR322 *tetR/tetA sacB*	(Lenz and Friedrich, 1998)
pLO3_Δ*recJ*	pBR322 *tetR/tetA sacB HA_ΔrecJ*	*
CRISPR/Cas plasmids		
pMTL-casd-1339-ack	colE1 oriT *Spycas9* gRNA scaffold	Pete Rowe, SBRC Nottingham
pSV500	pVS1-p15A *catP araC* pBAD-*λ Red gam-Spycas9* ptet-*I-SceI*	*
pCRISPomyces-2	colE1-pSG5(Ts) oriT prpsL(xc) *Stmcas9* gapdh(EL)-*lacZ* sgRNA-*redN*	Addgene #61737
pSV501	pVS1-p15A *catP araC* pBAD-*Stmcas9* ptet-*I-SceI*	*
pCRMAGE	pBBR1 *aph(3')-II araC* pBAD -*Stmcas9* pnahG-*recET-dnaM (C.necator* JMP134)	*
pMTL8315_CRISPR _*Cpast_spoIIE*_HA	pCB102 *catP* P*fdx-Rb3 -Spycas9* p1339-sgRNA_*spoIIE_Cpa*-fdx term HA_*Cpa*	Ines Canadas, SBRC Nottingham
pCRISPR	pBBR1 *catP* pfdx-RB3-*Spycas9* p1339-sgRNA_*spoIIE_Cpa*-fdx term HA_*Cpa*	*
pCRISPR_gRNA0	pBBR1 *catP* pfdx-RB3-*Spycas9* p1339-sgRNA_*cadA (E. coli)*-fdx term HA_*phaC*	*
pCRISPR_574 (+)	pBBR1 *catP* pfdx-RB3-*Spycas9* p1339-sgRNA_*phaC_574(+)*-fdx term HA_*phaC*	*
pCRISPR_737 (-)	pBBR1 *catP* pfdx-RB3-*Spycas9* p1339-sgRNA_*phaC_737(-)*-fdx term HA_*phaC*	*
pTargetF-*cadA*	pMB1 *aadA* pij23119-sgRNA-*cadA*	(Jiang *et al.*, 2015)
pSV600	pBBR1*aadA* pij23119-sgRNA-*cadA*	*
pTarget	pVS1-p15A *tetR/tetA* pRhaB-sgRNA *cadA*	*
pTarget_*hsdR*	pBBR1 *aadA* pij23119-sgRNA_*hsdR*	*
pTarget_HA_*hsdR*	pBBR1 *aadA* pij23119-sgRNA_*hsdR*_HA_*hsdR*	*
pTarget_*odhA*	pBBR1 *aadA* pij23119-sgRNA_*odhA*	*
pTarget_ HA_*odhA*	pBBR1 *aadA* pij23119- sgRNA_*odhA*_HA_*odhA*	*
pTarget_ *phaC*	pVS1-p15A *tetR/tetA* pij23119-*sgRNA_phaC*_rrnBT2	*

pTarget_ HA _ *phaC*	pVS1-p15A *tetR/tetA* pij23119-*sgRNA_phaC*_rrnBT2_HA2_*phaC*	*
pij23119_EYFP	pBBR1 *aph(3')-II* plac lacO M13 pT3 rrnBT2 pj23119-*eyfp*	*
p_EYFP	pBBR1 *aph(3')-II* plac lacO M13 pT3 rrnBT2 *eyfp*	*
pBBRR33	pBBR1 *aph(3')-II* plac lacO M13 pT3 rrnBT2 pRham-*eyfp*	Samantha Bryan, SBRC Nottingham
Ethylene-related plasmids		
pEFE(P)	pBBR1 *catP* pj5-*efep*-λtl3	*
pHokSok	pBBR1 *catP* p*phaC-efep*-λtl3 hok/sok	*
gTME		
pGEM-*efep*	f1 *bla* p15-*efep* pSP6 M13 pT7 plac lacO	Alexander Van Hagen, SBRC Nottingham
p*LacZα*	pVS1-p15A *tetR/tetA* T4 term plac lacO M13 pUC19 MCS lacZα	*
p*rpoD*	pVS1-p15A *tetR/tetA* T4 term plac lacO M13 rpoD (*E.coli* MG1655)	*

2.2.2 Culture growth

E. coli strains were grown at 37°C unless otherwise stated while *Cupriavidus* strains were propagated at 30°C. Liquid cultures were agitated at 200 rpm and typically grown at 10-20% of the volume of a 250 mL baffled flask equipped with a membrane screw cap to allow optimal oxygenation. When required, the OD of cultures was measured at 600 nm (OD_{600}) with a BioMate™ 3S Spectrophotometer (ThermoFischer Scientific, Waltham, MA, United States) or with an automated spectrometer (Tecan Infinite 200).

2.2.3 Bacterial storage

For long-term storage, a traffic light system was adopted: a new bacterial strain was stored in triplicates in one green, one yellow and one red vial to be used subsequently in this order. 1 mL of overnight cultures were mixed with Cryobeads from the Pro-Lab Diagnostics™ Microbank™ Bacterial and Fungal Preservation System. ~ 900 μL of cryopreservative fluid was removed from each vial and the remaining liquid was used to resuspend the culture. The vial was then aseptically closed and inverted a few times. The vials were immediately stored at -80 ºC.

2.3 Molecular procedures

2.3.1 Extraction and purification of chromosomal DNA

Genomic DNA was extracted from *E. coli* MG1655*, C. necator* and *C. metallidurans* using a Sigma-Aldrich® GenElute™ Bacterial Genomic DNA Kit, according to manufacturer's instructions. The protocol for DNA extraction from gram-negative bacteria was utilised.

2.3.2 Extraction and purification of plasmid DNA from E. coli

Plasmid DNA was extracted from *E. coli* cells using either the Qiagen Plasmid Miniprep Kit or the NEB® Monarch® Plasmid Miniprep Kit according to the manufacturer's instructions.

2.3.3 Extraction and purification of plasmid DNA from Cupriavidus spp.

Cupriavidus species have a high lipid content, which can make DNA extraction from conventional plasmid kits particularly challenging. Therefore a protocol (Feliciello and Chinali, 1993), originally developed for plasmid purification from *E. coli*, was used for plasmid extraction from *Cupriavidus* species. The final pellet was resuspended in 100 μL TE and stored at -20°C.

2.3.4 Amplification of DNA via PCR

All polymerase chain reactions were performed in Sapphire 0.2 mL thin-walled, flat cap PCR tubes using a Biometra Trio (AnalytikJena©, Jena, Germany) PCR thermocycler. All PCR-amplified DNA was analysed via agarose gel electrophoresis.

2.3.4.1 Q5 PCR for cloning purposes

PCR was performed utilising either NEB® Q5® polymerase or Q5® 2X Master Mix in 25 or 50 μL reactions **(Table 2.3)** and adapted

thermocycler conditions **(Table 2.4)**. Q5® is a high-fidelity DNA polymerase and was preferentially utilised given its proof-reading capacity.

Table 0.3 – Composition of Q5® -based PCR reaction

Component	**50 µL reaction**	**25 µL reaction**
5× Q5 Polymerase Buffer	10 µL	5 µL
dNTPs (10 µM)	1 µL	0.5 µL
Forward primer (10 µM)	2.5 µL	1.25 µL
Reverse primer (10 µM)	2.5 µL	1.25 µL
Template DNA	~20 ng	~10 ng
5X High-GC content Enhancer (optional|)	5 µL	2 .µL
Q5® DNA Polymerase	0.5 µL	0.25 µL
Nuclease-free water	To 50 µL	To 25 µL
Total reaction volume	**50 µL**	**25 µL**

Table 0.4 – Thermal cycling conditions for Q5® PCR

Step		**Temperature**	**Duration**
Initial denaturation		98°C	2 min
30 cycles	Denaturation	98°C	10 s
	Annealing	Tm +1°C	30 s
	Extension	72°C	30 s/kb
Final extension		72°C	5 min
Hold		10°C	∞

2.3.4.2 DreamTaq colony PCR (cPCR) for screening

cPCR was performed to screen selected transformants from successful cloning, DNA transformation or genomic mutation. Thermo-Fisher Scientific® DreamTaq Green® PCR 2X Master Mix was used routinely in 15 µL reactions. For *Cupriavidus* clones, a single colony was picked using a toothpick and resuspended in 10 µL of nuclease-free water in a PCR tube, then boiled for 10 min. 5 µL of the boiled material was subsequently added to 10 µL of the pre-made master mix composed of DreamTaq Green® PCR 2X Master Mix and the adequate primer pair (**Table 2.5**). For screening of *E. coli* colonies, a single colony was selected and directly mixed in a 15 µL reaction. Thermocycling conditions are shown in **Table 2.6**.

Table 0.5 – Composition of DreamTaq Green-based cPCR reaction

The template was either a single colony picked from the agar plate and directly mixed in the reaction or 5 µL boiled material.

Component	
DreamTaq Green® PCR Master Mix (2x)	7.5 µL
Forward primer (~17 µM)	0.5 µL
Reverse primer (~17 µM)	0.5 µL
Template DNA	Single colony or 5 µL
Nuclease-free water	Variable
Total reaction volume	**15 µL**

Table 0.6 – Thermal cycling conditions for cPCR

Step		**Temperature**	**Duration**
Initial denaturation		95°C	5 min
30 cycles	Denaturation	95°C	30 s
	Annealing	Tm -5 °C	30 s
	Extension	72°C	60 s/kb
Final extension		72°C	10 min
Hold		10°C	∞

2.3.5 Restriction endonuclease digestion of DNA

Restriction enzymes supplied by Thermo-Fisher Scientific® were all from the FastDigest® range and were used in 10XGreen Buffer. NEB® enzymes were incubated in the universal 10X CutSmart® Buffer. Restriction digests were set up as 25 µL reactions in 1.5 mL Eppendorf®

microcentrifuge tubes (**Table 2.7**) and incubated at 37°C for 1 h. Reactions run in the Thermo-Fisher Green Buffer contained a density and tracking dye to allow analysis via agarose gel electrophoresis allowing direct application to the gel, while reactions incubated in the CutSmart® Buffer required the addition of 5 μL 5X DNA Gel Loading Dye before loading.

Table 0.7 – Composition of restriction endonuclease digest reaction

Component	
Green or CutSmart® Buffer (10×)	2 μL
DNA	1μg
Restriction enzyme # 1	1 μL
Restriction enzyme # 2 (if required)	1 μL
Nuclease-free water	Variable
Total reaction volume	**20 μL**

2.3.6 Manipulation of 5' DNA ends

2.3.6.1 De-phosphorylation of DNA

Vector dephosphorylation was conducted to prevent recircularisation of the vector backbone and reduce the number of false positive clones obtained. NEB Shrimp Alkaline Phosphatase was used in the following reaction (**Table 2.8**) to catalyse the 5' end dephosphorylation of pre-digested vector DNA. 30 μL reactions were incubated at 37°C for 30 min.

Table 0.8 – Composition of dephosphorylation reaction

Component	
DNA	1 µg
CutSmart Buffer (10X)	3 µL
rSAP	1 µL
Nuclease-free water	Variable
Total reaction volume	**30 µL**

2.3.6.2 Phosphorylation of DNA

Phosphorylation of DNA was carried out to add inorganic phosphate (P_i) to the 5'hydroxyl terminus of the PCR-amplified fragments allowing subsequent ligation. NEB® T4 Polynucleotide Kinase was used in T4 DNA Ligase Buffer (**Table 2.9**) and 10 µL reactions were incubated at 37°C for 30 min prior to ligation.

Table 0.9 – Composition of phosphorylation reaction

Component	
DNA	500 ng-1 µg
10X T4 DNA Ligase Buffer	3 µL
10X T4 Polynucleotide Kinase	1 µL
Nuclease-free water	Variable
Total reaction volume	**10 µL**

2.3.7 Analysis of DNA via agarose gel electrophoresis

Agarose gels (1%) were prepared by mixing 1 g technical grade agarose (Sigma) per 100 mL 1X TAE buffer (40 mM Tris base; 0.1% (v/v) glacial acetic acid; 1 mM EDTA) and melting the mixture at high heat for 2/3 min. SYBR® Safe DNA gel stain (Thermo-Fisher Scientific®) was added at a final concentration of 10 μL/mL to visualise DNA bands on the gel. The samples in Green Buffer or supplemented with NEB® 6X Purple Gel Loading Dye were loaded on the gel and run in 1X TAE buffer at 100 V, (400 mA) for 40-60 min and visualised using a UV transilluminator (Gel Doc™). DNA size was estimated by comparison with a NEB® 2-log DNA ladder.

2.3.8 Gel extraction and purification of DNA

2.3.8.1 Extraction and purification of DNA from agarose gel

DNA bands of appropriate size were excised from the agarose gel under blue light on a transilluminator (BT Lab Systems©). Each slice was subsequently weighed in a 1.5 mL Eppendorf® tube. DNA was extracted using the either the QIAquick® or Monarch® Gel Extraction Kit according to the manufacturer's instructions. DNA was eluted in 13 μL of nuclease-free water.

2.3.8.2 Extraction and purification of DNA from reaction mixtures

DNA was directly extracted from both PCR and restriction digest reaction mixtures, using either the QIAquick PCR purification kit or the Monarch® PCR & DNA Clean-up kit according to the manufacturer's instructions. Elution was performed in 13 μL nuclease-free water.

2.3.9 Ligation of DNA fragments

2.3.9.1 Traditional ligation of two DNA fragments

The ligation of two DNA fragments between the 5´-phosphate and the 3´-hydroxyl groups of their compatible cohesive or blunt ends was performed using NEB® T4 DNA Ligase. Ligation reactions were usually set up in a 3:1 insert: vector ratio as stated (**Table 2.10**). Ligation reactions were incubated at RT for 2 h, or at 4°C overnight. The amount of insert DNA was calculated using the following equation (Engler M.J. and Richardson, 1982):

$$\frac{ng\ of\ vector\ \times size\ of\ insert\ (kb)}{size\ of\ vector\ (kb)} \times molar\ ratio\ of\ \frac{insert}{vector} = ng\ of\ insert$$

Table 0.10 – Composition of two fragment-ligation reaction

Component	
Vector DNA	50-100 ng
Insert DNA	Variable
10× T4 DNA Ligase buffer	1 µL
T4 DNA ligase	1 µL
Nuclease-free water	To 10 µL
Total reaction volume	**10 µL**

2.3.9.2 Ligation of DNA fragments using HiFi Assembly

HiFi assembly was utilised to assemble multiple fragments generated via PCR and restriction digest. PCR products were amplified with primers designed utilising the NEBuilder® web tool (NEB®) and included 20-25 bp overlapping ends to facilitate DNA annealing. Reactions were

typically set up in a 2:1 insert: vector ratio (**Table 2.11**) and were incubated for 1 h at 50°C. The amount of insert was calculated as described in *2.3.9.1.*

Table 0.11 – Composition of HiFi Assembly reaction

Component	2-4 fragment assembly
Vector DNA	50-100 ng
Insert(s) DNA	Variable
HiFi® Assembly Master Mix (2x)	5-10
Nuclease-free water	Variable
Total reaction volume	**10-20 µL**

2.3.10 Quantification of nucleic acids

DNA and RNA concentration were measured using a BioMate™ 3S Spectrophotometer (ThermoFischer Scientific, Waltham, MA, United States). The device was blanked with high purity water or elution buffer from the extraction kits. One µL of each sample was loaded onto the device and absorbance at 260 nm and 280 nm was measured. DNA maximal absorbance is at 260 nm whereas proteins absorb preferentially at 280 nm. The 260/280 ratio was used to measure sample purity. A ratio of ~ 1.8 was generally accepted as ''pure'' for DNA whilst a ratio of ~2.0 was considered 'pure' for RNA (Sambrook, Fritsch and Maniatis, 1989).

2.3.11 RNA isolation and reverse transcription

2.3.11.1 RNA isolation

Ten millilitres cultures were set up in LB medium, supplemented with the appropriate antibiotic and inoculated with a single colony of *C. necator* cells. Cultures were incubated overnight at 30°C, with shaking.

Cultures were then centrifuged and lysed in 2 mL TRI Reagent® (Sigma-Aldrich®) by repeated pipetting and transferred to a fresh 15 mL tube. To ensure complete dissociation of nucleoprotein complexes, samples were left at RT for 5 min. A total of 0.4 mL of chloroform was added to each sample and samples were shaken vigorously for 15 s. After 15 min at RT, the resulting mixture was centrifuged at 12,000 xg for 15 min at 4°C. Centrifugation separates the mixture into 3 phases: a red organic phase (containing protein), an interphase (containing DNA), and a colourless upper aqueous phase (containing RNA). The aqueous phase was transferred to a fresh tube and 1 mL of isopropanol was added per mL of TRI Reagent used. Samples were kept at -20°C overnight. Samples were then centrifuged at 12,000 xg for 30 min at 4°C. The RNA precipitate formed a pellet on the side and bottom of the tube. The supernatant was removed, and the RNA pellet was washed by adding 2 mL of 75% ethanol. Samples were vortexed then centrifuged at 7,500 xg for 15 min at 4 °C. The RNA pellet was air-dried for a maximum of 1 h-1 h 30 min. Pellets were not allowed to completely dry out as this greatly decreases solubility. Pellets were then resuspended in 100 μL nuclease-free water. The final preparation of RNA should be free of DNA and proteins, RNA purity was measured on the SimpliNano™ (GE Healthcare®, Chicago, IL, USA) and an A_{260}/A_{280} ratio of ~2.0 was accepted for qualifying a sample as 'pure' RNA.

2.3.11.2 DNAse treatment

DNase treatment of RNA samples was conducted prior to Reverse Transcriptase PCR (RT-PCR). The DNase digestion reaction was set up with RQ1 RNase-Free DNase (Promega®) and incubated at 37°C for 30 min (**Table 2.12**). One microlitre of RQ1 DNase Stop Solution (Promega®) was added and samples incubated at 65°C for another 10 min to inactivate the DNase and terminate the reaction.

Table 0.12 – Composition of DNase digestion reaction

Component	
RNA in water	1-8 µL
10X RQ1 RNase-Free DNase Reaction Buffer	1 µL
RQ1 RNase-Free DNase	1 u/µg RNA
Nuclease-free water	Variable
Total reaction volume	**10 µL**

2.3.11.3 Two-step RT-PCR

10 µL of the treated RNA was used to carry on to the two-step RT-PCR. The first step is the reverse transcription of RNA into complementary DNA (cDNA) transcripts that are then used in the second step as a template for PCR amplification of the sequence of interest using specific primers. The RT step is set up in a sterile microfuge tube **(Table 2.13)** and run in two steps. First, RNA, Go Script™ Oligo (dT) (Promega®), dNTPs and nuclease-free water were mixed together in a 16µL final volume then heated for 3-5 min at 65-80°C. Samples were briefly centrifuged and promptly placed on ice, RT buffer, RNAse inhibitor and M-MuLV Reverse Transcriptase were then added to a 20 µL final volume. The mixed samples were incubated for 1 h at 42°C before the enzyme was inactivated at 90°C for 10 min. All reactions were stored at -20°C or directly utilised in the PCR reaction.

Table 0.13 – Composition of reverse transcription reaction

Component	
Total RNA solution	0.5-2 µg
Oligo (dT)	2 µL
dNTPs mix (2mM)	4 µL
Nuclease-free water	To 16 µL
10X RT buffer	2 µL
RNAse inhibitor	1 µL
M-MuLV Reverse Transcriptase	1 µL
Total reaction volume	**20 µL**

2.4 DNA transfer into bacterial cells

2.4.1 Preparation of *E. coli* electrocompetent cells

A single colony of *E. coli* DH5α was inoculated in 10 mL LB and incubated overnight at 37°C with agitation. 1 mL of the overnight culture was then inoculated into 100 mL of pre-warmed LB in a 1L flask and incubated for ~2 h at 37°C with shaking. Once the OD_{600} had reached ~0.8, the flask was stored on ice for 15 min. The 100 mL culture was split into four parts by pouring about 25 mL into ice-cold centrifuge tubes, before cells were centrifuged at 6,000 rpm for 15 min at 4°C. The pellet was resuspended in 100 mL ice-cold sterile water and resuspended cells were centrifuged again. The pellet was then washed with ice-cold 10% glycerol and centrifuged one last time. Cell pellets were finally resuspended in 500 μL ice-cold glycerol. 50-μL aliquots were transferred into sterile 1.5 mL microfuge tubes and stored in the -80°C freezer or immediately used for transformation (M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, 1989).

2.4.2 Preparation of *E. coli chemically competent cells*

To prepare chemically competent *E. coli*, a frozen glycerol stock of bacterial cells (Top10, DH5α, etc.) was streaked out onto an LB plate and grown overnight at 37°C. A single colony of *E. coli* from the fresh LB plate was selected and used to inoculate a 10 mL starter culture of LB. The culture was grown overnight at 37°C in a shaker. 100 mL of LB media was then inoculated with 1 mL of the starter culture and grown in a 37°C shaker. When the OD_{600} reached 0.35-0.4, the cells were immediately put on ice for 30 min. The 100 mL culture was split into four parts by pouring about 25 mL into ice-cold centrifuge tubes. The cells were harvested by centrifugation at 3000 xg for 15 min at 4°C. The supernatant was removed, and each pellet was gently resuspended in 10 mL of ice-cold 100 mM $MgCl_2$.

The cells were harvested by centrifugation at 2000 xg for 15 min at 4°C. The supernatant was decanted, and the pellets were combined in about 2 x 10 mL of ice-cold $CaCl_2$. The suspension was kept on ice for at least 20 min. The cells were harvested by centrifugation at 2000 xg for 15 min at 4°C. The supernatant was discarded, and the pellet was resuspended in ~2 mL of ice cold 85 mM $CaCl_2$, 15% glycerol. Fifty microlitres aliquots were transferred into sterile 1.5 mL microfuge tubes and stored in the -80°C freezer or immediately used for transformation (M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, 1989).

2.4.3 Transformation of *E. coli competent cells*

Typically, either 100 ng of plasmid DNA or 10-20 μL of the ligation or HiFi assembly reaction, was mixed gently with competent cells by pipetting up and down. In the case of electrocompetent cells, the mixture was transferred to a pre-chilled electroporation cuvette of 0.2 cm gap width and chilled on ice for 5 min. The electroporation was performed at 2.5 kV, 25 μF, and 200 Ω in a MicroPulser™ Electroporator (Bio-Rad). Cells were transferred into 1.5 mL-Eppendorf tubes and supplemented with 950 μL of SOC media (ThermoFischer®). For chemically competent cells, the mixture was placed on ice for 30 min then heat-shocked at 42°C for 30 s. Tubes were transferred back onto the ice for 2 min then 950 μL of RT SOC media (ThermoFischer®) (M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, 1989).

Both electrocompetent and chemically competent cells were recovered at 37°C for 1 h with shaking. Cells were harvested by centrifugation at 13,000 rpm for 1 min and the pellet was resuspended in 100 μL. Resuspended cells were spread on to pre-warmed LB plates with the appropriate antibiotic. Plates were incubated overnight at 37°C and subsequent clones were screened by cPCR.

2.4.4 Preparation of *Cupriavidus* spp. electrocompetent cells

C. necator and *C. metallidurans* cells were prepared for transformation utilising the electroporation method developed by Ehsaan *et al*, unpublished. Ten milliliters of SOB medium, supplemented with the appropriate antibiotic, was inoculated with a heavy loop of fresh *Cupriavidus* spp. cells and incubated overnight at 30°C with shaking. Overnight cultures were then used to inoculate 50 mL of SOB media to a final OD_{600} ~ 0.065-0.085. The culture was incubated with shaking for about two hours at 30°C, when the OD_{600} reached 0.3, the cells were harvested by centrifugation at 7000 rpm and washed twice with ice-cold 1 mM HEPES pH 7.0. After a final centrifugation step, the pellet was resuspended in 10% glycerol/1 mM HEPES pH 7.0 to an OD_{600} of 5. Approximately 100 μL cell aliquots were either stored at -80°C or utilised directly for electroporation.

2.4.5 Transformation of *Cupriavidus* spp. electrocompetent cells

One hundred nanograms of plasmid DNA was mixed with 100 μL of electrocompetent *Cupriavidus* spp. cells and transferred into a 0.2 cm gap cuvette, before electroporation at 2.5 kV, 25 F and 200 Ω. Immediately following electroporation, 900 μL of RT SOC medium (Fischer Scientific) was added to the cells and the whole mixture was transferred into a 1 mL Eppendorf tube. Electroporated cells were recovered for 2 h at 30°C, shaking, before being plated onto selective medium.

2.4.6 Conjugative plasmid transfer and gene replacement in Cupriavidus necator using a mobilisable suicide plasmid

2.4.6.1 Construction of a mobilisable suicide plasmid

To construct a mobilisable suicide plasmid for the knock-out of the *recJ* gene (H16_RS05815) in *C. necator*, two 800 bp fragments containing the left and right homology arms, complementary to flanking regions upstream and downstream of the *recJ* gene, were amplified using the primer

pairs *recJ*_OUT.F/R, utilising *C. necator* H16 genomic DNA as a template. PCR fragments were assembled by HiFi in the pLO3 vector (Lenz and Friedrich, 1998), pre-digested with SacI and XbaI to generate pLO3_*recJ*. The backbone pLO3 vector contained a high-copy number pBR322 origin of replication, an IncP origin of transfer, the *B. subtilis sacB* gene, under its native constitutive promoter and a tetracycline-resistance cassette.

2.4.6.2 Conjugative plasmid transfer (First single-crossover)

About 100 ng of pLO3_*recJ* were used to transform chemically competent *E. coli* S17-1 λ*pir* and transformants were selected on LBT. Successful transformants were screened by cPCR. Single colonies were incubated overnight (~16-18 h) at 37°C in 10 mL of LSLB supplemented with tetracycline 15 μg/mL. *C. necator* H16 was also incubated overnight at 30°C in LB. 1 mL of each culture was centrifuged for 4 min at 5000 rpm. The cells were then washed consecutively with 1 mL and then 500 μL of LB, the cells were then mixed together and centrifuged again. The supernatant was discarded, and the pellet was resuspended in 30 μL of LB and spotted onto a fresh LB plate, mating was performed at 30°C for 6 h. After 6 h, the cells were scrapped off the plate with a sterile loop and resuspended in 1 mL of PBS. Transconjugants were selected on 0.4% SG-MM + Tet (15 μg/mL) agar plates, incubated for 48-72 h at 30°C. Tetracycline-resistant transconjugants were picked and purified by streaking (2x) on 0.4% SG-MM + Tet (15 μg/mL) agar plates. Purified transconjugants were screened by cPCR to confirm single crossovers and insertion of the suicide plasmid into the chromosome.

2.4.6.3 Gene replacement by *sacB*-based counter selection (Second crossing-over)

Selected transconjugants were inoculated in LSLB + 15% sucrose without antibiotic and grown overnight at 30°C. 100 μL of appropriate dilutions were spread onto LSLB plates containing 15% sucrose. Sucrose–

resistant survivors appeared after 48-72 h of incubation at 30°C. At least 25 sucrose-resistant clones were picked and plated onto LB + Tet (15 µg/mL) and LB agar plates. Cells that did not grow on LB + Tet (15 µg/mL) were tetracycline-sensitive and therefore potentially fully segregated mutants, in which the tetracycline-resistant suicide plasmid should have been excised along with the *recJ* gene. Colonies from the replica LB plate were used to screen tetracycline-sensitive clones for gene deletion by cPCR and Sanger sequencing.

2.5 Calculation of plasmid retention and transformation efficiencies

2.5.1 Plasmid segregational stability of the pVS1 replicon in C. necator (modified from (Simpson, Skurray and Firth, 2003))

Plasmid stability of the pVS1 origin of replication was compared to stability of the pBBR1 replicon, already widely used in *C. necator* (Bi *et al.*, 2013). Two seed cultures of *C. necator* H16 harbouring the pVS1-replicative pMMG plasmid and the pBBR1-based pME6000 vector (Maurhofer *et al.*, 1998) were individually plated onto LBT and incubated at 30°C for 48 h. Three single colonies were harvested from each agar plate and grown independently in 10 mL LBT at 30°C and 250 rpm for 16 h. One hundred microlitres of the overnight cultures was used to inoculate 10 mL of LBT and the strains were incubated at 30°C for 4 h. One milliliter was extracted from each culture and washed twice in PBS. The cultures were serially diluted and 100 µL of the 10^{-3}-10^{-6} dilutions were spread onto non-selective LB technical replica plates (x2), for viable counts. The number of colonies from each plate was recorded two days later and the less confluent dilution was selected for further plating. One hundred microlitres of the 10^{-4} dilution was used to inoculate 10 mL LB with no antibiotic at 30°C and 250 rpm for 16 h. This corresponds to the start of the experiment (t=0 h). On subsequent days the serial dilution of each individual culture was plated onto LB agar, colonies were counted after 48 h and 100 colonies (50 per technical

replicate) were re-streaked onto LB and LBT to quantify the proportion of cells retaining each plasmid. A subculture was inoculated every day in 10 mL LB with 100 µL of the 10^{-4} dilution (final dilution: 10^{-6}; number generations needed to return to stationary phase: ~20). This procedure was repeated over 4-6 days. When necessary a lower dilution than that determined initially, was used to spread onto LB agar plates to maintain a similar number of viable cells counted on LB plates across the whole experiment.

The ratio of plasmid-bearing cells was measured as follows (Kramer, 2016):

$$\%\ plasmid - bearing\ cells = \frac{Number\ of\ colonies\ on\ LB\mathrm{T}}{Number\ of\ colonies\ on\ LB} \times 100$$

Then the percentage of plasmid retention in each *C. necator* strain was plotted against the number of cell generations.

2.5.2 Transformation efficiencies

Transformation efficiencies (TE) in *C. necator* H16 were determined after counting single colonies on selective agar plates and utilizing the following formula (Hanahan, 1983):

$$TE\ (CFU/ug) = \frac{Colonies\ on\ selective\ agar\ plate/amount\ of\ DNA\ used\ (\mu g)}{dilution\ of\ cell\ suspension\ before\ plating}$$

The survival percentage was estimated by calculating the CFUs obtained after second plating by the CFUs enumerated on transformation plates and that figure was normalized to a percentage. Each transformation was performed in triplicate.

2.6 Homologous recombination-based generation of deletion mutants in *C. necator*

2.6.1 Preparation of λ-Red or RecE/T electrocompetent C. necator cells

A λ-Red or RecE/T plasmid was introduced into *C. necator* by electroporation. The transformants were spread on LB agar medium containing 50 µg/mL chloramphenicol (LBC). A single colony was used to inoculate 5 mL of SOB supplemented with 50 µg/mL chloramphenicol (SOBC) and grown overnight at 30°C. The cultures were used to inoculate 50 mL of SOBC supplemented with 10 mM arabinose (to induce expression of the λ-Red or RecE/T system) to a final OD ~ 0.065-0.085. The culture was incubated with shaking for about 2 h at 30°C. When the OD_{600} reached 0.2-0.3, the cells were harvested by centrifugation at 7000 rpm and washed twice with ice-cold 1 mM HEPES as described previously. The pellet was resuspended in 10% glycerol-1 mM HEPES and either directly transformed or stored at -80°C.

2.6.2 Preparation of donor DNA

The donor DNA required for allelic exchange was either a suicide circular plasmid or a linear double stranded DNA cassette generated by restriction digestion or PCR amplification. To delete the *hsdR* gene, the plasmid pSV200 was synthesised by Biomatik© (Cambridge, Ontario, Canada). It is constituted of the high-copy-number ColE1 origin of replication that is not recognised by the replication machinery of *C. necator* (Kleihues *et al.*, 2000)). The suicide vector pSV200 also contained the *bla* gene that confers resistance to ampicillin and the mutation cassette C1 targeting the *hdsR* gene. The cassette includes the *dhfR* gene under the constitutive *ampR* promoter and a modified *ampR* RBS, the I-SceI recognition site preceded with two transcriptional terminators (Kim *et al.*, 2014) and the homology regions HR1, HR2 and HR3. To generate the *phaC*-targeting suicide plasmid pSV442, pSV200 was digested with XbaI and

assembled via HiFi isothermal reaction to four fragments that formed the mutation cassette C2: LHA, *sceI*, *aadA* and RHA. Primers used to construct the vector can be found in Appendix **Table A.1**. The *C. necator*-compatible plasmid pSV242 was built from the XbaI-linearised C2 fragment of pSV442. The PCR-generated pBBR1 origin of replication was ligated to the C2 fragment to form pSV242.

The suicide plasmid was transformed and extracted from either DH5α or Top10 *E. coli* cells. To produce the linear cassette, the pSV242 plasmid was transformed in *C. necator*. A single colony was incubated overnight in LB supplemented with 15 μg/mL tetracycline at 30°C with shaking. One microgram of the plasmid was then digested with XbaI to produce the mutation cassette C2. The cassette was gel purified and resuspended in 8 μL nuclease-free water. Alternatively, 10 ng of plasmid DNA was used as a template for PCR amplification of the same cassette using primers C1. F/R or C2.F/R. The products of 6 individual PCR reactions were gel purified and resuspended in 8 μL nuclease-free water producing approximately 10 μg DNA, to maximise the amount of donor DNA used in λ-Red or RecE/T homologous recombination (Lesic and Rahme, 2008).

2.6.3 Genome editing procedure (adapted from the Kim method (Kim et al., 2014))

A total of 100 μL of λ-Red or RecE/T-encoding electrocompetent cells were freshly prepared and resuspended in 10% glycerol-1 mM HEPES. The cells were transformed with 100 ng-10 μg donor DNA by electroporation as described previously. After electroporation, cells were outgrown in 900 μL SOC supplemented with 50 μg/mL chloramphenicol and 10 mM arabinose for 3 h, with shaking at 30°C. The cells were then spread onto plates containing LB plus chloramphenicol 50 μg/mL and trimethoprim 60 μg/mL (LBCTm) or spectinomycin 300 μg/mL (LBCS) to

select for cells in which the mutation cassette had been integrated into the chromosome.

2.6.4 Selection and screening of λ-Red or RecE/T recombinants

After incubation at 30°C for 48-72 h, colonies were counted and all or at least 20 individual colonies were picked and streaked onto fresh LBCTm or LBCS plates. After incubation overnight at 30°C, survivors were screened for trimethoprim or spectinomycin-resistance and tested via colony PCR for correct integration of the mutation cassette with both internal and external primers.

2.7 CRISPR-based generation of deletion mutants

2.7.1 CRISPR-based plasmid curing system in C. necator

2.7.1.1 Plasmid construction

The plasmid pSV500 was HiFi-assembled from the PCR-amplified fragments pSLVS1 and Cas9 using primers pSLVS1.Bb. F/R and Cas9.F/R using plasmids pSLVS1 and pMTL_Cas_1339_ack (Pete Rowe, personal communication) as templates. The p15A-pVS1 shuttle vector contained the *S. pyogenes cas9* gene (*Spycas9*), which expression is being controlled by the arabinose-inducible pBAD promoter. The sequence of a second *cas9* gene, codon-optimised for *Streptomyces* species, (*Stmcas9*) was assembled utilising the backbone of pSV500 to create an 11.6 kb plasmid named pSV501.

The sgRNA-expressing vector pSV600 was generated using pME6000 as a template. The pBBR1 fragment was cloned into pTarget_F, previously digested with MluI/NdeI. The isothermal assembly method was utilised to assemble the construct. pSV600 was then used as a template to create pTarget, quick change PCR was used to switch the gRNA sequence with primers gRNA F/R. The linear PCR product was treated with DpnI, purified, then its 5'ends were phosphorylated to allow for self-ligation.

To test the promoter activity of pij23119, a series of plasmids were built from the template pBBRR33, kindly provided by Dr Samantha Bryan. Pij23119_EYFP was used to evaluate the expression of *eyfp* under the pij23119 promoter while pBBR33 was used as a benchmark for EYFP fluorescence controlled by the rhamnose-inducible promoter pRham. pEYFP carries *eyfp* with no promoter, allowing the estimation of transcriptional reflux pf the pBBRR33 backbone. The region including the rhamnose-inducible pRham promoter and the *eyfp* gene was cut out of pBBRR33 using the enzymes NsiI and NdeI. It was replaced by the fragment pij23119-*eyfp*, obtained by PCR using pBBRR33 as a template and the primers pij23119_EYFP.F/R, with

pij23119_EYFP.F including the sequence of the constitutive *E. coli* promoter (Ke*lly et al*, 2009)To construct pEYFP, pBBRR33 was digested with NsiI and NdeI then treated with T4 DNA Polymerase (NEB®) supplemented with dNTPs, as described in section *2.3.6*, to fill in 3' and 5' ends. The fragment was then recircularised using T4 DNA Ligase.

The expression of eYFP was assessed in cells harbouring each of the three plasmids, supplemented with/without 0.2% w/v rhamnose. The experiments were performed in parallel in both *C. necator* and in *E. coli,* allowing expression levels to be assessed between the two species. Fluorescence and OD_{600} were monitored on a Tecan Infinite 200 over 111 min in *E. coli* and 668 min in *C. necator*. As *C. necator* has a growth rate lower than that of *E. coli*, it was necessary to extend the length of the monitoring period to make sure cells had reached stationery phase and the whole exponential phase was covered in the experiment.

2.7.1.2 Preparation of CRISPR/Cas9-encoding electrocompetent C. necator cells

The Cas9-expressing plasmid was introduced into *C. necator* by electroporation. The transformants were spread on LBC and a single colony was used to inoculate 5 mL of SOBC supplemented with 10 mM arabinose and grown overnight at 30°C. The culture was used to inoculate 50 mL of SOBC supplemented with 10 mM arabinose, at a final OD ~ 0.065-0.085. The culture was incubated with shaking for about two hours at 30°C. When the OD_{600} reached 0.3, the cells were harvested by centrifugation at 7000 rpm and washed twice with ice-cold 1 mM HEPES as described previously. The pellet was resuspended in 10% glycerol-1 mM HEPES and either utilised directly or stored at -80°C.

2.7.1.3 Genome editing procedure

One hundred microlitres of CRISPR/Cas9 electrocompetent cells were transformed with either water or ~100 ng sgRNA-transcribing plasmid, by electroporation as described previously. After electroporation, 900 µL SOC supplemented with 10 mM arabinose was added to the cells and the mixture was incubated for 3 hours, shaking at 30°C. Cells were harvested by centrifugation, resuspended in 150 µL LB and spread onto LBC (total cell number), LBS (total transformed cells) and LBCS (surviving transformants) plates to select for CRISPR/Cas9-induced mutant cells.

2.7.1.4 Screening for deletion mutants

Cells were counted and up to 16 colonies from each LBCS were patched onto fresh LBCS plates for colony purification and onto LBS to screen for cells sensitive to chloramphenicol, reflecting the correct cleavage of the Cas9 vector. Screening was performed by cPCR using the primer pairs I-SceI.F/R or Cas9.Strep.IN.F/R to confirm the loss of the Cas9-expressing plasmid. Transformation efficiencies were calculated as described in *2.5.2*.

The number of cells on LBC was established as 100% survival rate and the efficiency rate was the rate of cells that did not survive after colony purification on LBCS. cPCR was used to confirm the loss of the Cas9 plasmid.

2.7.2 CRISPR-based generation of deletion mutants in C. necator

2.7.2.1 Plasmid construction

pCRMAGE contains both the Cas9 protein and the *recET* system on the same backbone. The plasmid pMTL71101, from the SBRC culture collection, was digested with NheI and SacI to generate a fragment with the *Cupriavidus gilardii recET* homologues genes and the *C. necator* JMP134 *dnaM* gene. The intermediate pCasRecBB plasmid was created by the isothermal assembly (Jiang *et al.*, 2015) of the *recET/dnaM* fragment to the

PCR–generated sequences pBBR1 and KanR. The pBBR1 fragment was produced using pCRMAGE_pBBR1.F/R primers and the plasmid pME6000 as a template whilst the generation of KanR was performed with the primer pair pCRMAGE_KanR.F/R and the template pMTL71401. Then pCasRecBB was linearised with SpeI and PsiI and ligated via HiFi assembly to the salicylic-acid regulated promoter p_{nahG} and its associated transcription regulator *nahG* (H16_RS08130) to create pCRMAGE.

Finally, pMTL8315_CRISPR _*Cpast_spoIIE*_HA was digested with SbfI and AatII to produce a 4.9 kb backbone. The fragments pBBR1 + CatP and HA were generated via PCR using the primer pairs pCRISPR_pBBR1 + catP_F/R and pCRISPR_HA_F/R and the plasmids pSLBBR1 and pTarget_HA_4 as templates. Both fragments were assembled with the SbfI-AatII backbone via HiFi reaction to create pCRISPR.

For each gene targeted with the two-plasmid CRISPR/Cas9 system, a similar plasmid assembly protocol was utilised to construct the sgRNA-transcribing vector. A series of up to 4 plasmids was built per target gene. A gRNA targeting the *E. coli cadA* gene was considered as a non-targeting gRNA vector control (gRNA0) in the CRISPR/Cas9 experiments. The three other members harboured each a unique guide sequence targeting the region of interest (**Table 2.14**). The plasmids all carried the same 1.6 kb homology region, flanking the gene or the specific region to be knocked out. The pTarget_*hsdR* and pTarget_*odhA* series were built by assembling the left and right homology arms amplified from *C. necator* total DNA into the backbone vector pSV600, previously digested with EcoRI and BglII. The pTarget_*phaC* series was built using the low-copy number-plasmid pTarget. The pCRISPR series was built using the pCRISPR plasmid in which the *spoIIE* spacer was replaced by the *E. coli cadA* spacer (pCRISPR_gRNA0) or one of the two *phaC*-targeting gRNAs, either pCRISPR_574(+) or pCRISPR_737(-). The homology sequences were assembled in the SpeI-digested pTarget. The sgRNA targets were constructed utilising inverse PCR. Each forward primer

had a 10 bp sequence complementary to the 3'end of the gRNA, whilst each reverse primer had a 10 bp tail complementary to the 5'end of the gRNA sequence. The numbers used to name each gRNA correspond to the position of the Cas9-induced cleavage site, 3 bp upstream of the protospacer adjacent motif (PAM) sequence and defining position 1 as the first nucleotide of the coding sequence.

Table 0.14 – gRNA sequences used in the generation of C.necator CRISPR/Cas9 mutants

gRNA	Target	Strand	Sequence	PAM
p_Target_*hsdR* series				
354 (-)	*hsdR*_N	-	GAATGTCCCAGTCAATGTAG	CGG
431 (+)	*hsdR*_N	+	TTGCCCGCCTGGGTTCAATT	CGG
387 (-)	*hsdR*_N	-	ACTGGTTGACCACTGTGAAG	CGG
p_Target_*odhA* series				
1007 (+)	*odhA*	+	CAACCCGGTGGTCGAAGGCT	CGG
1038 (-)	*odhA*	-	CCTTGTGGCCGACTTCGCCG	CGG
1137 (+)	*odhA*	+	CTGAACCTCGCGCAGACCCG	CGG
p_Target_*phaC* series				
737 (-)	*phaC_N*	-	TGATGCACGGTGATGCACGG	CGG
574 (+)	*phaC_N*	+	ATGATGGAAGACCTGACACG	CGG
337 (-)	*phaC_N*	-	CGGCAGCGAAGCGATATGGG	AGG

2.7.2.2 Preparation of CRISPR/Cas9-encoding electrocompetent C. necator cells

The Cas9-expressing plasmid was introduced into *C. necator* by electroporation. The transformants were spread on LBC and a single colony was used to inoculate 5 mL of SOBC supplemented with 10 mM arabinose and grown overnight at 30°C. The culture was used to inoculate 50 mL of SOBC supplemented with 10 mM arabinose, at a final OD ~ 0.065-0.085. The culture was incubated with shaking for about two hours at 30°C. Then 10 mM salicylic acid was added to the cells to induce the expression of the heterologous *recET*/*dnaM* system in *C. necator* pCRMAGE. When the OD_{600} reached 0.3, the cells were harvested by centrifugation at 7000 rpm and washed twice with ice-cold 1 mM HEPES as described previously. The pellet

was resuspended in 10% glycerol-1 mM HEPES and either utilised directly or stored at -80°C.

2.7.2.3 Genome editing procedure

Approximately 100 µL of CRISPR/Cas9-encoding electrocompetent cells were transformed with either water or ~100 ng sgRNA-transcribing vector by electroporation, as described previously. After electroporation, the cells were grown in 900 µL of SOC media supplemented with 50 µg/mL chloramphenicol and 10 mM arabinose, the cells were incubated for 5 h. *C. necator* cells harbouring pCRMAGE, were supplemented after 1 h recovery, with 10 mM salicylic acid to induce the *recET*/*dnaM* genes. Cells were harvested by centrifugation, resuspended in 150 µL LB, diluted and 50 µL was spread onto either LBC (total cell number), LBS (total transformed cells) or LBCS plates to select for CRISPR/Cas9-induced mutant cells. Transformations were conducted three times and the averages ± standard deviations of their efficiencies were calculated as colony forming units per µg of plasmid DNA.

For strains harbouring the pCRMAGE plasmid, cells were recovered for 5 h after transformation in SOC media supplemented with 300 µg/mL kanamycin (to maintain pSV500), 10 mM arabinose (Cas9 induction) and 2 mM salicylate (*recET* and *dnaM* induction). CRISPR-edited mutants were selected on LB plates supplemented with 300 µg/mL kanamycin and 15 µg/mL tetracycline (Km + Tet) at a dilution of 10^1-10^3. Transformations were done in duplicate.

2.7.2.4 Screening for deletion mutants

After 2-3 days incubation, transformants appeared and were counted then re-streaked onto fresh LBCS plates. Surviving colonies were screened via cPCR using primers internal and external to the targeted region.

For genome editing using the pCRISPR series, each vector was transformed into *C. necator* electrocompetent cells and transformants were screened via cPCR to confirm the presence of the plasmid. A single colony was incubated in LB supplemented with 60 µg/mL chloramphenicol for four days at 30°C. This recovery time allowed native recombination mechanisms to promote allelic exchange between the plasmid-borne homology arms and the genomic sequence flanking the *phaC_N* domain. A total of 100 µL of each culture was diluted in PBS and 100 µL of appropriate dilutions were spread on two sets of LB agar plates supplemented with chloramphenicol, one with 5 mM theophylline (Theo) for induction of *cas9* expression and the other one without inducer (No Theo). Plates were incubated for 48 h at 30°C and colonies were enumerated.

The *phaC_N* mutants were also screened for their phenotype. They were re-streaked onto 0.4% SG-MM plates supplemented with Nile-Red® (Sigma-Aldrich®). This lipophilic stain reacts with lipids and produces an intense fluorescence when excited under blue light (Gorenflo *et al.*, 1999). Potential mutants were plated alongside a wild-type control and the *ΔphaCAB* mutant strain. Transformation efficiencies were calculated as described in *2.5.2.* The number of cells on LBC was established as 100% survival rate and the efficiency rate was the rate of cells that did not survive after colony purification on LBCS and mutation was confirmed via cPCR.

2.8 Protein techniques

2.8.1 Sample preparation

C. necator H16 and *C. metallidurans* CH34 expressing genes encoding the proteins of interest were cultured overnight in 10 mL LB containing the appropriate antibiotics at 30°C and 200 rpm. OD_{600} nm was recorded and 1 mL of the culture was centrifuged, and the pellet was resuspended in the appropriate amount of Bugbuster® solution. The following calculation was

utilised to calculate the amount of Bugbuster required for efficient lysis (http://www.emdbiosciences.com/html/NVG/home.html):

$$Bugbuster\ (\mu L) = \frac{\frac{OD600nm}{0.2} \times 45}{2}$$

The Bugbuster solution was prepared from 10X Bugbuster Protein Extraction Reagent (Novagen®, **Table 2.15).**

Table 0.15 – Composition of the Bugbuster solution

Component	
10X Bugbuster Protein Extraction Reagent	100 µL
Protease Inhibitor Cocktail (Merck-Millipore)	20 µL
rLysozyme™ Solution (Merck-Millipore)	1 µL
Benzonase® Nuclease (Merck-Millipore)	0.1 µL
PBS	879 µL
Total reaction volume	**1 mL**

The reaction was incubated at RT and with shaking at ~100 rpm for 20 min. Then 50 µL of each reaction was transferred to a fresh Eppendorf® microcentrifuge tube and centrifuged at 13,000 rpm for 5 min. Then, 10 µL of the supernatant (soluble fraction) was transferred to a fresh tube containing 10 µL Sample Buffer (**Table 2.16**). Two additional 10 µL supernatant samples were stored at -20°C, as additional replicates. The rest of the supernatant was discarded and the pellet (insoluble fraction) was resuspended in 50 µL of PBS, then incubated at 4°C for ~30 min; 10 µL of each resuspended pellet was transferred to a fresh tube containing 10 µL of Sample Buffer.

Table 0.16 – Composition of the Sample Buffer solution

Component	
4X SDS Sample Buffer (Merck-Millipore)	2.5 µL
10X Reducing agent	1 µL
Nuclease-free water	6.5 µL
Total reaction volume	**10 µL**

2.8.2 Electrophoresis

Soluble and insoluble samples were briefly centrifuged, boiled for 10 min and briefly centrifuged again. Then 10-20 µL of each fraction (10-20 µL of a 1:10 dilution of each fraction) were loaded onto a pre-cast Thermo Fisher® NuPAGE® 4-12% SDS mini gel, in a ThermoFischer® XCell SureLock® Mini-Cell gel electrophoresis chamber, filled up with MES SDS running buffer. The samples and 5 µL of ThemoFischer® PageRuler™ Prestained Protein Ladder were run for 1 h 30 min at 120 V.

2.8.3 Staining of SDS gels

After electrophoresis, the protein gel was removed from the tank and placed in a square petri dish. Gels were initially washed with ~15 mL water for 5 min with agitation then ~20 mL (a sufficient amount to submerge the gel) of Life Technologies® SimplyBlue™ Safe Stain were added to the dish and the gel was stained for 1 h at 60 rpm. Gels were de-stained by subsequent washes in RO water, finally gels were imaged with a BioRad® Geldoc™ imaging system.

2.8.4 Western Blotting

Following electrophoresis, the gel was removed from the tank and immediately transferred onto the membrane of a BioRad® Trans-Blot®

Turbo™ Mini PVDF Transfer Pack, using a Trans-Blot® Turbo™ protein blotting system. The membrane was blocked with 5% non-fat dried milk dissolved in PBS for 1 h at RT and 60 rpm. A second replica gel was separately processed for staining. The membrane was washed three times in PBS with 0.1% Tween 20 for 10 min at RT/60 rpm. The blot was then incubated with the Primary Antibody (1:5000) in 1% BSA PBS, overnight at 4°C with gentle shaking, (**Table 2.17**). For EFE(P) detection, the anti-EFE(P) antibody was utilised in *C. metallidurans.* For the detection of either Cas9 expressed in *C. necator*, the rabbit monoclonal anti-Cas9 antibody [EPR18991] from Abcam® was used, as the primary antibody. Following overnight incubation, the membrane was washed three times in PBS with 0.1% Tween 20 for 10 min at RT and 60 rpm. The secondary antibody was added at (1:10000) in PBS +0.1% Tween 20 for 1 h at RT. As both the anti-Cas9 and the anti-EFE(P) primary antibodies were rabbit monoclonal IgGs, a unique goat monoclonal anti-rabbit IgG antibody conjugated with a horseradish peroxidase (HRP), was used as the secondary antibody. The membrane was washed three times in PBS with 0.1% Tween 20 for 10 min at RT and 60 rpm. Finally, 5 mL of ThermoScientific® One-Step™ TMB-Blotting, a one-component HRP substrate, was added to the membrane and development was allowed to proceed for about 5-10 min. The substrate is oxidised by the HRP and a blue-purple precipitate is formed, coloured bands represent proteins of interest. When bands started to appear, water was gently but immediately added into the dish to stop the oxidation reaction and the membrane was imaged with a BioRad® Geldoc™ imaging system.

Table 0.17 – Composition of the antibody-containing solutions

Component	**Primary Antibody (1:5000)**	**Secondary Antibody (1:10000)**
Primary Antibody	2 µL	_
Secondary Antibody	_	1 µL
PBS	10 mL	10 mL
BSA	100 µL	_
Total reaction volume	**10 mL**	**10 mL**

2.9 Microscopy

The plasmid pMMG was introduced *in C. necator* competent cells by electroporation. The transformed cells were spread on LBT_{15} and incubated at 30°C for 48 h. Colonies from one or two selected transformants were added to 4 µL of dH_2O on a microscope slide and a cover slip was added to the slide. As a negative control, pME6000 transformants were used, which produce no GFP. Both pMMG and pME6000 cells were observed on the Olympus IX71 inverted microscope using when required a light filter (λ = 400 nm) for visualisation of GFP during image acquisition. Pictures of the agar plates were taken with the 40x oil objective (1.4 NA). The pMMG plasmid contains the *Pseudomonas* pVS1 origin of replication (RepA and StaA), the *E. coli* p15A replicon and a repressible tetracycline resistance (TcR) determinant and constitutively expresses GFP.

2.10 Ethylene measurement

2.10.1 Culture growth

The selected strains were tested for ethylene production by gas chromatography (GC). A single colony of each strain was used to inoculate liquid M9 medium, supplemented with 100 µg/mL ampicillin and 15 µg/mL tetracycline, if required, and the cultures was grown over 48 h at 30°C. Cultures were diluted to an OD_{600nm} of 0.08 in 10 mL fresh M9 medium supplemented with antibiotics. Three millilitres aliquots were grown overnight in triplicates in 10 mL rubber-capped GC serum bottles at 30°C, 200 rpm.

2.10.2 Ethylene measurement by Gas Chromatography (GC)

Approximately 2 mL of the headspace was collected with a gas syringe after 24 h growth and analysed using a Trace™ 1300 gas chromatograph (Thermo Scientific™) under the following conditions:

column size: 0.53 mm × 40 mm; solid phase: Porapak N column; column temperature: 60°C; carrier gas: helium and detector: TCD. The amount of ethylene present in the sample was measured using the Ethylene Faster running method (**Figure 2.1**). Each vial was used only once for sample collection. OD values were determined using a spectrophotometer set at the wavelength λ = 600 nm). Quantity of ethylene (nmol) detected in 2 mL was rationalised to the full headspace (7 mL), OD_{600} was measured immediately after gas sampling for GC analysis and ethylene yield were obtained by rationalising the quantity of ethylene in 7 mL headspace to the sample OD_{600} and the culture volume (3 mL).

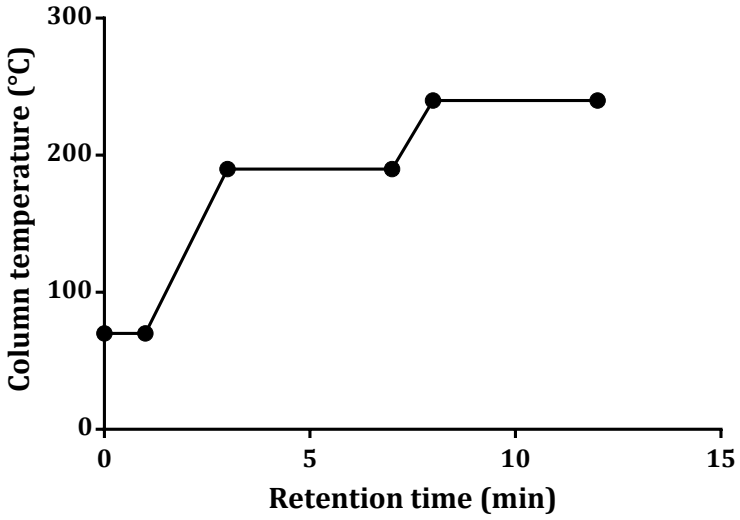


Figure 0.1 – Elution gradient for detection of ethylene on the Trace™ 1300 GC

2.10.3 Plasmid segregational stability of a Hok/Sok toxin/antitoxin vector in C. metallidurans (modified from (Simpson, Skurray and Firth, 2003)).

2.10.3.1 Construction of a Hok/Sok vector

To stabilise the *efep*-encoding pEFE(P) plasmid in *C. metallidurans* CH34 and to establish a stable, antibiotic free expression system in *Cupriavidus*, the Hok/Sok toxin/antitoxin system was added onto the pEFE(P) plasmid. The 580 bp Hok/Sok cassette was amplified by PCR with

the primer pair HokSok.F/R from the synthetic *hoksok* gene (Eurofins) and the *catP* gene was re-amplified from pEFE(P) with the primer pair pHS.*catP*.F/R. Both fragments were ligated using HiFi assembly into pre-digested pEFE(P) vector cut with AatII and ApaLI, forming pHokSok.

2.10.3.2 Evaluation of the contribution of the Hok/Sok system to the stability of an *efep*-expressing plasmid in *C. metallidurans*

The stability of pHokSok in CH34 was compared to that of pEFE(P) to assess whether the toxin/antitoxin system is a better addiction system compared to the chloramphenicol resistance conferred by *catP*.

Two seed cultures of *C. metallidurans C*H34 harbouring either the pEFE(P) or pHokSok plasmid were individually plated on LBC_{250} and incubated at 30°C for 48 h. Three single colonies of each strain were harvested from each agar plate and grown independently in 10 mL LBC_{250} at 30°C and 250 rpm for 16 h. Then, 100 µL of the overnight cultures was used to inoculate 10 mL of LBC_{250} and were incubated at 30°C for 6 h. One millilitre of each culture was sampled and washed twice in PBS. Then, 100 µL of 10^{-2}-10^{-5} serial dilutions were spread on non-selective LB technical replica plates (x2) for viable counts. The number of viable colonies were recorded two days later, and the most suitable dilution was selected for further plating. 100 µL of the 10^{-3} dilution was also used to inoculate 10 mL LB with no antibiotic at 30°C and 250 rpm for 16 h. This stage corresponds to the start of the experiment (t = 0 h). Serial dilutions from each individual culture were plated onto LB agar plates on a daily basis, colonies were counted after 48 h and 100 colonies (50 per technical replicate) were re-streaked onto LB and LBC_{250}, to quantify the proportion of cells retaining each plasmid. A subculture was inoculated every day in 10 mL LB with 100 µL of the 10^{-3} dilution. This procedure was repeated over 6 days (#50 generations). When required a lower dilution was used to spread onto LB agar plates, thus maintaining a constant number of viable cells,

which could be easily counted on LB plates across the whole experiment. Afterwards, the generation time of each strain was calculated using the formula (Simpson, Skurray and Firth, 2003):

$$G = \frac{t}{3.3 \log b/B}$$

With:

t = time interval (h),

B = number of bacteria at the beginning of a time interval

b = number of bacteria at the end of the time interval.

The ratio of plasmid-bearing cells was measured as follows:

$$\%\, plasmid - bearing\ cells = \frac{Number\ of\ colonies\ on\ LBC250}{Number\ of\ colonies\ on\ LB} \times 100$$

For evaluation of plasmid stability in fermentation culture, cells were grown in 750 mL working volume in continuous culture mode and were fed with a CO_2-rich gas mixture and DSMZ 81 supplemented with 0.8% glycerol. Dilution rate was initially set up at D = 0.01 h^{-1} (7.5 mL/h) and 6 mL samples were collected at regular time points, diluted appropriately and spread onto selective (for evaluation of plasmid segregational stability) and non-selective plates (for viable counts). Colonies were enumerated, and survival ratio was calculated as above.

2.11 Global Transcription Machinery Engineering (gTME)

2.11.1 Library construction

DNA cloning and plasmid preparations were performed according to standard methods. To construct the template for the gTME mutant library, a 550-bp fragment containing the LacZα fragment of β-galactosidase was amplified using the primer pair LacZα.F and LacZα.R, utilising pUC19 (Yanisch-Perron *et al*, 1985) as a template. The PCR fragment was assembled by HiFi into the backbone of the pME6031 vector pre-digested with MluI and XhoI to form pLacZα. The *rpoD* gene (*b3067*) and the 197 bp intergenic region between *dnaG* (b3066) and *rpoD,* including its native promoter, was amplified from the chromosomal DNA of *E. coli* K-12 MG1655, using the primers *rpoD*_E.F/R. The 2.1-kb fragment generated was digested with HindIII and SacI and ligated into the pLacZα vector, resulting in the generation of the vector p*rpoD*_E.

Error-prone PCR was then performed on the *prpoD*_E vector using the same primer pair *rpoD*_E.F and *rpoD*_E.R and the Genemorph II Mutagenesis Kit (Agilent Technologies). Three PCR reactions were set up with increasing initial amounts of DNA, to achieve a low (0-4.5 mutations/kb), medium (4.5-9 mutations/kb), and high (9-16 mutations/kb), mutation frequency. The higher the initial DNA template concentration is, the lower mutation frequency is expected for a fixed number of cycles (Agilent Technologies® Instruction Manual Catalog #200550).

The size of the PCR products was checked on an agarose gel. Both the p*LacZα* vector and the PCR products were digested overnight with HindIII and SacI. The p*LacZα* plasmid was then treated with recombined Shrimp Alkaline Phosphatase (rSAP) (NEB, Ipswich, USA) before ligation with each of the mutated *rpoD* pools, the ligations were transformed into

DH5α. Cells were selected on LB + Tet 15 μg/mL supplemented with X-Gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) 0.2% (w/v).

A robot colony picker was used to select blue/white colonies and selected white clones were grown overnight in LB for glycerol stock preparation and plasmid pool extraction. Clones obtained from the low, medium and high mutant library were stored separately. The total library size was approximately 10^3.

2.11.2 Phenotype selection

Fifty microlitres of electrocompetent *ΔproB* MG1655+pGEM-*efep* cells were transformed with 4 x 8 μL of each of the plasmid preps from the low, medium and high library or 100 ng of p*rpoD*_E as control and selected on LB supplemented with 100 μg/mL ampicillin and 15 μg/mL tetracycline. Cells were picked using the robot and inoculated for both glycerol stocks and into M9 medium supplemented with 100 μg/mL ampicillin and 15 μg/mL tetracycline in a 96-well plate format. These cultures were passaged once in fresh M9 medium supplemented with antibiotics and sub-cultured at an OD_{600nm}= 0.1. Cells were grown for 4 days and the OD_{600nm} was measured every 2 h, using the robotics platform. After 24 h, ODs were compared to the OD of the control strain *(*Δ*proB* MG1655+pGEM-*efep+* p*rpoD*_E) and all clones showing a higher or equal OD to the control were isolated and stored in glycerol.

2.11.3 Sequence analysis

Sanger sequencing of plasmids and PCR fragments was run by Source BioScience© (Nottingham, UK) or Eurofins Genomics© (Ebersberg, Germany). Whole genome sequencing and RNA-seq of *E. coli* strains were contracted out to Genewiz© (South Plainfield, NJ, USA).

2.12 Adaptive Laboratory Evolution (ALE)

2.12.1 Fermentation process

Two 100 mL cultures were prepared in 500 mL baffled shake flasks by inoculating glycerol stocks from the *rpoD* library into LB medium. The cells were subcultured in 0.8% (w/v) glycerol M9 medium supplemented with 15 µg/mL tetracycline (to maintain the p*rpoD* vector) at 30°C. The fermentations were carried out in a DASGIP® bioreactor system, with a working volume of 750 mL. Both the pH and dissolved oxygen (DO) were controlled by the DASGIP controller where 5% (w/v) NH_4OH was used to control pH at 7.0 and DO was controlled at 30% air saturation.

Batch growth was initiated by inoculating the precultures from the rich medium. Chemostat culture mode was started after 24 h batch growth with a dilution rate starting at 0.025 h^{-1} this was then reduced to 0.01 h^{-1}. When cells had grown enough (OD_{600} = 5) the dilution rate was increased stepwise increasing from 0.02, 0.05, 0.07 and finishing at 0.1 h^{-1}.

Online sampling was performed every 24 h to measure the OD_{600} of the culture. At each pseudo-steady state (when five volume equivalents had been pumped through the system), 6 mL of culture were collected, diluted in PBS and spread twice onto selective LB agar plates (Amp 100 + Tet 15). Singles colonies were then tested on the GC to identify the best ethylene producer(s) at each pseudo-steady state. Controls include Δ*proB* + pGEM-*efep* + p*rpoD*_wt and the pGEM-*efep* plasmid expressed in either the wild-type MG1655 *E. coli* strain or in the proline-auxotroph Δ*proB* strain. Measurements were taken in duplicate or triplicate after 24 h at 30°C, 200 rpm. Each sample name includes the associated dilution rate (D= 0.02h^{-1} to D = 0.07 h^{-1}) followed by the colony identification number.

2.12.2 96-well format adaptive evolution

Approximately 50 µL of glycerol stock of each mutant from the Δ*proB* MG1655+pGEM-*efep*+p*rpoD* library was revived in 300 µL final volume LB before adaptation in 0.4% (w/v) glucose M9 medium supplemented with 15 µg/mL tetracycline. Adaptation was initiated with cells that had recovered in M9 medium. Those cells were diluted to OD_{600} = 0.08 into 96-well plates in 300 µL final volume and incubated at 30°C for 24 h. OD_{600} was monitored at frequent intervals and cultures were diluted 1:10 into fresh medium every 24 h. A sample of the cells were stored at -80°C after the 5^{th} and the 12^{th} passage.

2.12.3 Sample analysis

2.12.3.1 Growth rate and ethylene production

OD_{600nm} was monitored daily during the fermentation process. At each pseudo-steady state (e.g. the dilution rate D was equal to the population growth rate µ), a 1 mL sample was serially diluted and 100 µL of both the 10^{-4} and 10^{-5} dilutions were plated on LB supplemented with 100 µg/mL ampicillin and 15 µg/mL tetracycline (LBAT). All colonies were re-patched onto fresh LBAT and single colonies were tested for ethylene production by GC. Measurements was taken in triplicate for the best-performing clones at each dilution rate.

2.12.3.2 Whole genome sequencing and transcriptomics

The best performing strain D0.07_14 and the control strains Δ*proB*/pGEM-*efep* and Δ*proB*/pGEM-*efep* + p*rpoD*_wt were subject to fermentation as described previously at a dilution rate of D = 0.07. When the steady-state was reached, samples were collected and prepared for analysis. Two millilitres from each sample was collected and used for total DNA extraction and sent for whole-genome sequencing at Genewiz (South

Plainfield, NJ). Another 2 mL sample was washed with PBS and stored at 4°C overnight in RNA*later* (ThermoFischer Scientific®). Then the samples were centrifuged, the supernatant was discarded, and the pellet was stored at -80°C before being shipped to Genewiz (South Plainfield, NJ) for RNA extraction and sequencing.

Lambda Red and *recET*-assisted homologous recombination

3.1 Introduction

Lambda-Red (λ-Red)-assisted recombineering has been extensively used in *E. coli* (Yu *et al.*, 2000; Murphy and Campellone, 2003; Yosef *et al.*, 2004; Meynial-Salles, Cervin and Soucaille, 2005) and adapted to prophage, bacteria and fungi engineering (Chaveroche, Ghigo and d'Enfert, 2000; Serra-Moreno *et al.*, 2006; Karlinsey, 2007a; Sawitzke *et al.*, 2007; Lesic and Rahme, 2008). Therefore, the recombination system was assessed in *C. necator* to develop a genome engineering method more efficient, rapid and user-friendly than the *sacB*-based counterselection technique.

In 2014, Kim *et al* described a versatile scarless DNA editing method based on the allelic exchange of a mutation cassette into the chromosome of *E. coli* and *Salmonella enterica*, mediated by homologous recombination, utilising the λ-Red system and scarless excision of the cassette by I-SceI cleavage (Kim *et al.*, 2014). The recombinase system from bacteriophage λ was named after a recombination-defective mutant (Echolas and Gingery, 1968; Signer and Weil, 1968). The λ phage utilises the Red system to introduce its DNA into the host chromosome and replicates via the cell replication machinery (Sawitzke *et al.*, 2007). The system includes three proteins: Exo, Beta and Gam (**Figure 3.1**) but does not require the RecA function (Ellis *et al.*, 2001). Recombination starts with the degradation of double-strand DNA (dsDNA) by Exo. Exposed single-strand DNA (ssDNA) ends are protected by Beta, enhancing annealing between the exogenous DNA and the chromosomal DNA, with a strand bias for the 'lagging' strand at the replication fork (Ellis *et al.*, 2001). Finally, the Gam protein prevents the nuclease complex RecBCD cleaving the recombineering DNA. The exogenous phage DNA is then inserted into the chromosome of the host (Boyle *et al.*, 2013).

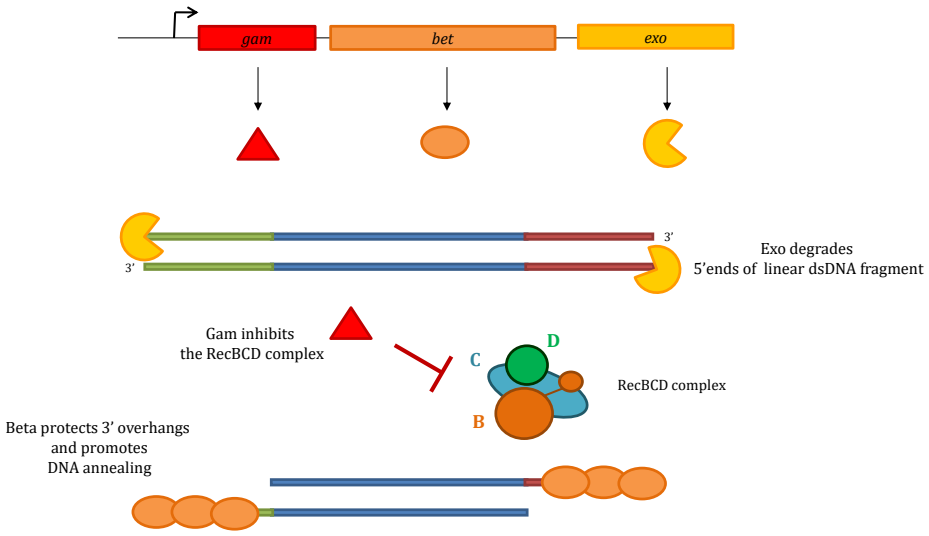


Figure 0.1 – Overview of bacteriophage λ recombination system.

Exo generates 3'-overhangs on linear dsDNA in a 5'-to-3' fashion. Beta binds the ssDNA and promotes single-strand annealing. An additional protein, Gam, prevents the RecBCD complex from degrading dsDNA fragments and improves the efficiency of dsDNA recombineering. RecBCD is a complex formed of three proteins (RecB, RecC and RecD) which unwinds blunt dsDNA ends and digests 3' and 5' strands (Wigley, 2013). (Adapted from Sharan *et al*, 2009).

Datsenko and Wanner adapted this mechanism to disrupt chromosomal genes in *E. coli*, generating PCR products with homology arms flanking regions of the target gene and co-transforming them with the plasmid-borne λ-Red genes (pKD46). An antibiotic selection marker was inserted into the host chromosome that allowed selection between wild type and mutant cells (Datsenko and Wanner, 2000). The antibiotic marker was eliminated by excision following the subsequent expression of the Flp recombinase, leaving a genetic scar.

λ-Red-assisted recombineering has been used to promote gene replacement in various bacteria and fungi (Chaveroche, Ghigo and d'Enfert, 2000; Lesic and Rahme, 2008; Jung *et al.*, 2012). For example, gene replacement was achieved in *E. coli* and *S. enterica* using a PCR-generated fragment as donor DNA (Lu *et al.*, 2002; Murphy and Campellone, 2003; Karlinsey, 2007b). It is interesting to note that plasmid-derived DNA fragments generated 50-to-60-fold higher replacement efficiencies

(recombinants/μg of DNA transformed), probably due to differences in the dsDNA ends of the fragments produced by PCR versus plasmid digestion (Murphy, Campellone and Poteete, 2000). Finally, the single expression of the Lambda Beta protein mediated recombination of ssDNA in *the E. coli* chromosome, enabling a simplified way to produce recombinant DNA molecules (Ellis, 2001).

I-SceI is a 235 amino acid homing endonuclease, encoded by the mitochondrial DNA of *Saccharomyces cerevisiae* (Watabe *et al.*, 1983). It recognises the 18 bp sequence 5'-TAGGGATAACAGGGTAAT-3', cutting it and generating a DSB (Kuijpers *et al.*, 2013).

Kim described a two-step recombineering system combining recombination promoted by the λ-Red system and dsDNA cleavage introduced by the meganuclease I-SceI, generating a scarless editing method in *E. coli* K-12 MG1655 and *S. enterica* (Kim *et al.*, 2014) (**Figure 3.2**). The editing method allows for gene deletion, insertion and point mutations to be carried out effectively and precisely and could be considered for multiplexgene engineering.

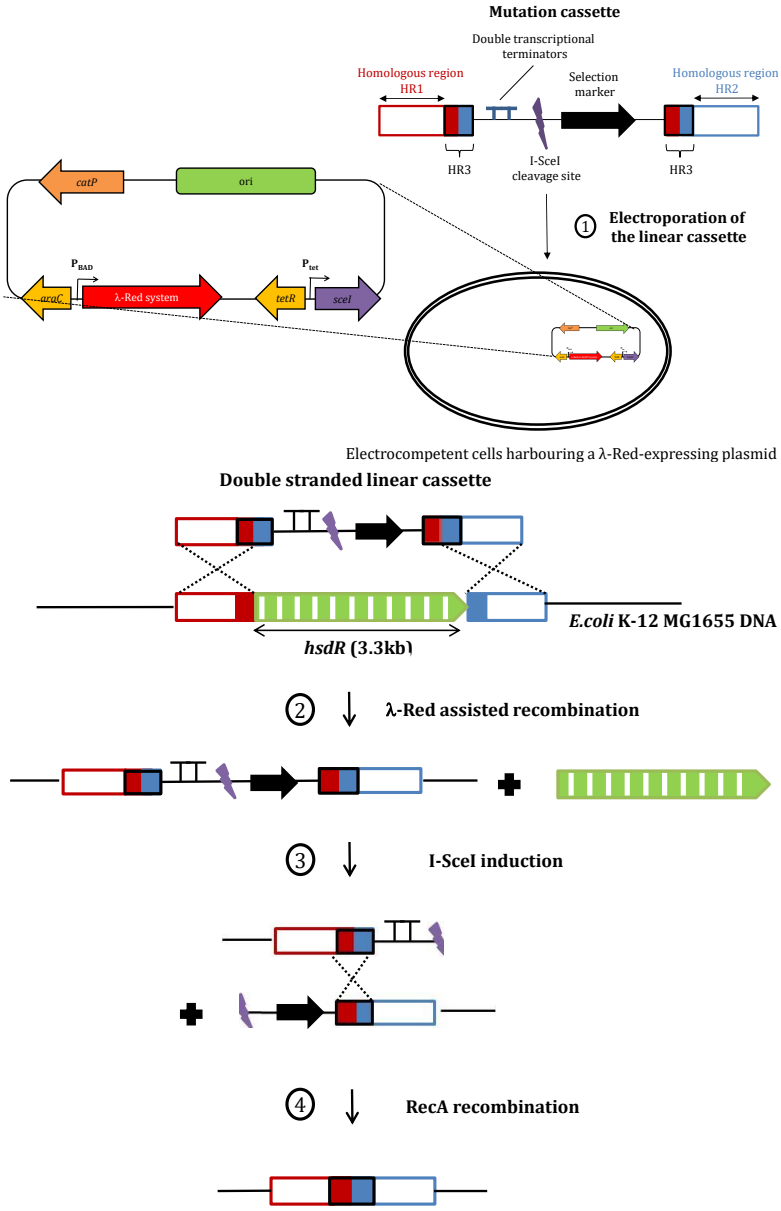


Figure 0.2 – General strategy for scarless genome editing (Kim et al., 2014)

1. The mutation cassette is introduced into the λ-Red-expressing strain. The cassette consists of a fragment containing 100 bp homology regions (HR1 and HR2), which directs the cassette into the desired chromosomal location. HR3 provides the region where homologous recombination can occur after I-SceI cleavage. Two HR3 sequences flank a sequence encoding a double transcriptional terminator in front of the I-SceI recognition site and a selection marker.
2. A double cross-over event enhanced by the λ-Red system allows the introduction of the mutation cassette. Recombined cells are selected on plates containing the appropriate selective marker.
3. Subsequently, I-SceI is induced and a DSB occurs, triggering a second homologous recombination event.
4. In a small minority of cells, the native RecA protein promotes recombination between the HR3 sequences, leaving no scar. The surviving cells have the desired chromosomal modification.

The RecET system, encoded by the Rac prophage, is also widely utilized for *in vivo* recombineering (Zhang *et al.*, 1998). It involves the cooperation of two key proteins: RecE, a type VIII exonuclease, that degrades dsDNA in a 5'-to-3' manner (similar to the λ Exo protein) and the RecT protein, a ssDNA-annealing and strand invasion protein, which facilitates recombination of homologous DNA (similar to the λ Beta protein (**Figure 3.3**) (Hall, Kane and Kolodner, 1993; Clark, Satin and Chu, 1994). The RecET system is a RecA-independent homologous recombination pathway (Kolodner, Hall and Luisi-DeLuca, 1994), which does not require an additional inhibiting protein, such as the λ Gam protein, utilised to suppress the activity of RecBCD, in the λ Red system (Lesic and Rahme, 2008).

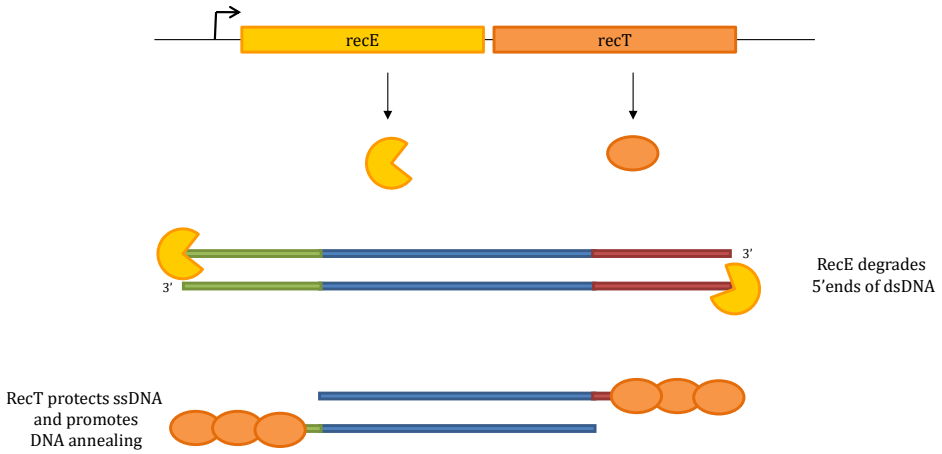


Figure 0.3 – Overview of the Rac prophage RecET recombination system.

RecE generates 3'-overhangs on linear dsDNA in a 5'-to-3' fashion. RecT binds the ssDNA and promotes ss-DNA annealing, resulting in the recombination of the exogenous linear DNA into the chromosome.

Both the λ-Red and RecET homologous recombination-based phage systems were assessed in *C. necator*.

3.2 Establishment of a λ Red-based system in *C. necator*

3.2.1 Construction of a functional λ-Red helper vector for C. necator

The λ-Red based plasmid pSLTS, has a temperature-sensitive origin of replication from the pSC101 vector, which has only been shown to be functional in *E. coli* and some other closely related Enterobacteriaceae (Felton and Wright, 1979; Ely and Wright, 1985; del Solar *et al.*, 1998). Despite repeated attempts, it was not possible to transform *C. necator* with the pSLTS plasmid, probably due to instability of the origin of replication, pSC101, in *C. necator*. Thus, it was necessary to replace pSC101 with another origin of replication. The replication and partitioning loci from the *Pseudomonas* plasmid pVS1 (Stanisich, Bennett and Richmond, 1977) was chosen for this. The copy number of pVS1 is around six to eight copies of plasmid per chromosome equivalent, which is equivalent to pSC101 (usually five copies per cell) (Sugiura, Ohkubo and Yamaguchi, 1993).

The plasmid pMMG (**Figure 3.4**), kindly provided by Dr S. Heeb, is a derivative from pME6032 (Heeb *et al*, 2002). The pVS1 replicon does not replicate in *E. coli,* because its origin of replication is not recognised by the replication machinery of *E. coli* (Heeb *et al.*, 2000). Therefore, a p15A-pVS1 shuttle vector was constructed; the p15A replicon can replicate in *E. coli* while the pVS1 loci are essential for replication in *Pseudomonas* and other species.

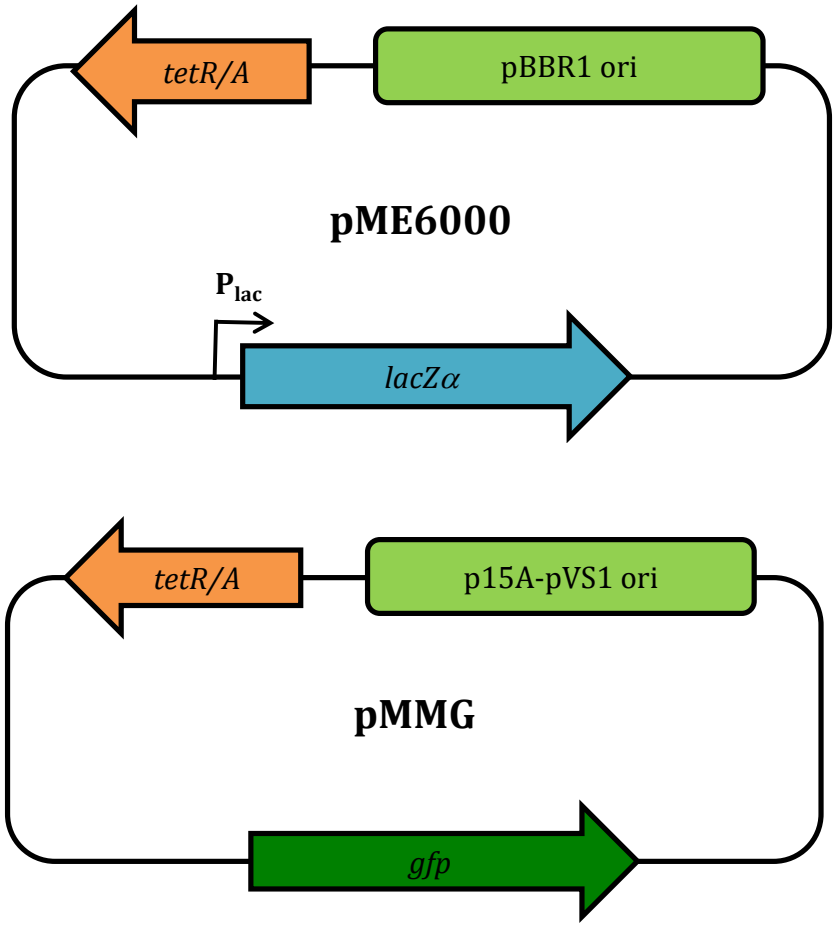


Figure 0.4 – Schematic representation of the medium-copy pME6000 and low-copy pMMG plasmids used for replication, fluorescence microscopy and stability assay in C. necator H16

tetR/A: confers resistance to tetracyclin; *gfp*: green fluorescent protein; *lacZα*: LacZα fragment of β-galactosidase.

C. necator was transformed with the plasmid pMMG and subsequent transformants were subject to imaging utilizing a fluorescent microscope (**Figure 3.5**), thus proving that the replication machinery encoded by the pVS1 loci is also functional in *C. necator*. However, replicon stability needed to be confirmed in *C. necator*.

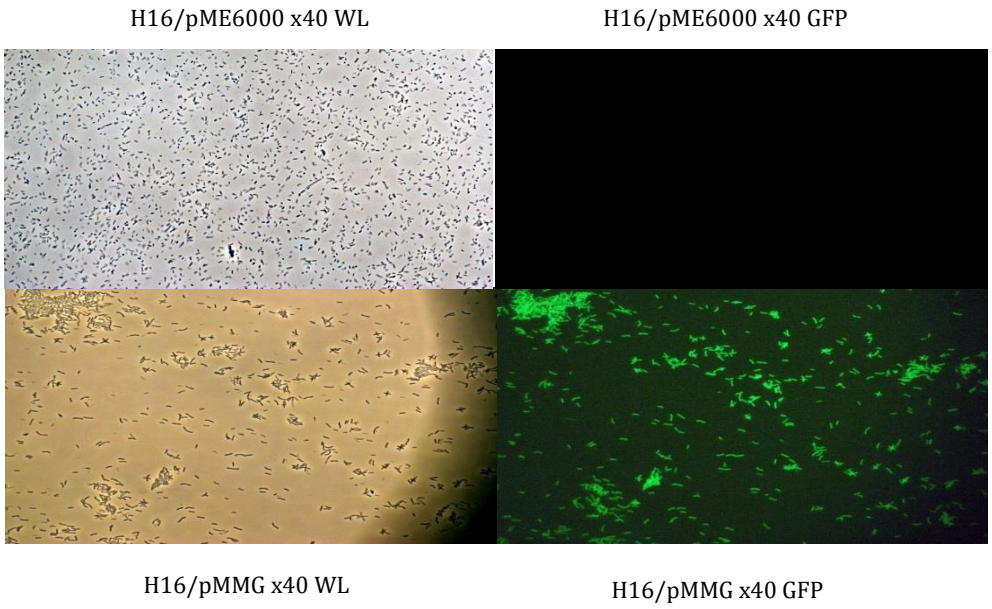


Figure 0.5 – Microscopy pictures of C. necator carrying the empty plasmid pME6000 or the GFP-expressing plasmid pMMG

Pictures were taken with the objective x40 under white light (WL) then with a 400/30 excitation filter and a 508/20 emission filter to detect GFP fluorescence.

Plasmid stability was verified by comparing the p15A-pVS1 shuttle plasmid, pMMG and the pBBR1-derived plasmid pME6000 in *C. necator* H16. Cells were grown at 30°C in nutrient LB medium without antibiotic selection for 4 days, as detailed in **Chapter 2**.

The stability assay confirmed that the pVS1 replicon is very stable in *C. necator* compared to the control plasmid pME6000 (**Figure 3.6**). Both vectors were maintained for 96 h (ca. 100 generations) in 96-100% of the population tested. Consequently, the pVS1 replicon can be utilized as an alternative to the pBBR1 replicon, the main gram-negative origin of replication used in previous studies with *C. necator* (Gruber *et al.*, 2014). In addition, pME6000 and pMMG could be co-transformed in *C. necator*, which confirmed the compatibility of the two replicons.

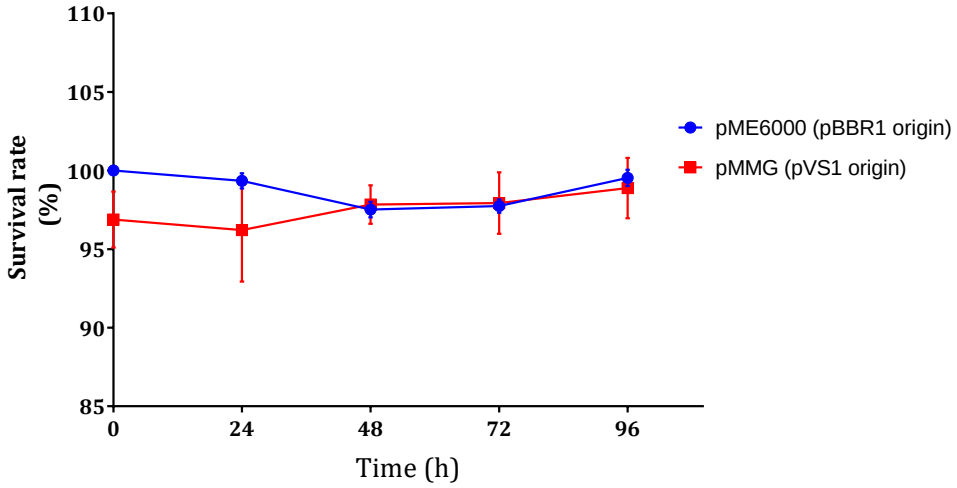


Figure 0.6 – Assessment of the stability of the pVS1 replicon in C. necator

The stability of the plasmids pME6000 (pBBR1 origin - blue) and pMMG (p15A-pVS1 shuttle double origins - red) were assessed in H16 over 96 h.

3.2.2 Targeting the restriction-modification system of C. necator

C. necator has been found to have a low transformation efficiency, compared to *E. coli*, of between 10^1 to 10^3 colony forming unit (CFU)/µg DNA for a low-copy number plasmid and 10^5 CFU/µg DNA for a medium-copy number plasmid (Sato, Fujiki and Matsumoto, 2013; Tee *et al.*, 2017). *C. necator* cells were transformed with varying amounts of plasmid DNA, utilising the low copy number replicon vector, pMMG (pVS1-p15A origins) and the medium copy-number plasmid replicon pME6000 (pBBR1 origin), (**Figure 3.7**).

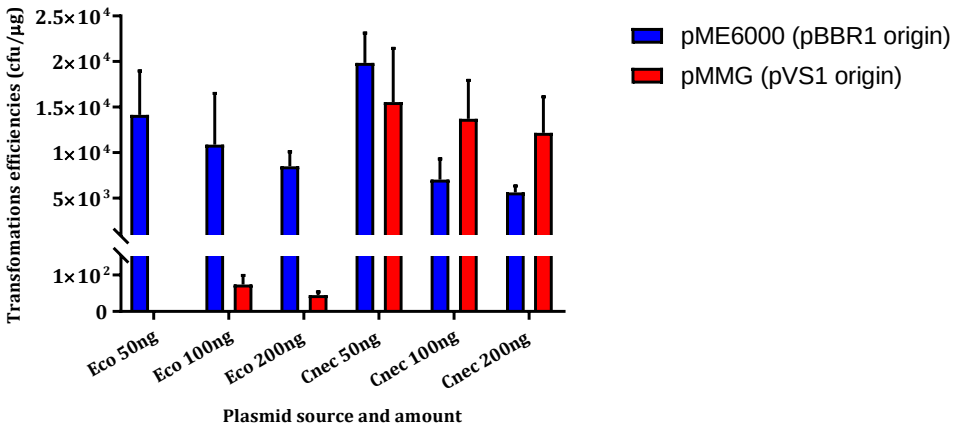


Figure 0.7 – Effect of DNA quantity, plasmid copy number and plasmid source on transformation efficiency

Transformation efficiencies associated with pME6000 (pBBR1 origin) are in blue while those obtained after transformation with pMMG (p15A-pVS1 double origin) are shown in red. Each plasmid was isolated from *E. coli* and transformed into *C. necator* (Eco 50, 100 and 200 ng) then they were isolated from *C. necator* and re-introduced in *C. necator* wild-type (Cnec 50, 100 and 200 ng).

As a general trend, there was an inverse correlation between the quantity of DNA used and the transformation efficiency. The plasmid pME6000 yielded a 10^2-fold higher transformation efficiency than the low copy number vector pMMG, when the plasmids were isolated from *E. coli*. However, when the plasmids were isolated from *C. necator*, pMMG achieved similar transformation efficiencies to pME6000 (Cnec 50 ng) outperforming the pBBR1 vector (Cnec 100 ng and Cnec 200 ng). Thus, transformations with the pMMG plasmid isolated from *C. necator* resulted in a higher efficiency (1-2x10^4-fold increase) compared to the same plasmid isolated from *E. coli*. However, transformation efficiencies with pME6000 were either similar for both *C. necator* and *E. coli* (10^4 for Eco 50 ng and Cnec 50 ng), or decreased slightly, when the plasmid was isolated from *C. necator* (~5x10^3 CFU/µg). In summary, plasmid isolation from *C. necator* is beneficial for low-copy number vectors, allowing the acquisition of a strain-specific methylation profile that reduces DNA restriction and enhances transformation efficiency.

It has been postulated that the host restriction/modification (RM) system contributes to poor transformation efficiencies (Chen *et al.*, 2008). Therefore, targeting this system in *C. necator* could improve transformation efficiencies and provide an ideal gene deletion target for recombineering. The location of genes involved in DNA RM systems encoded by *C. necator* H16 can be found on ReBASE Genomes (http://tools.neb.com/genomes/index.php?search) (**Figure 3.8**).

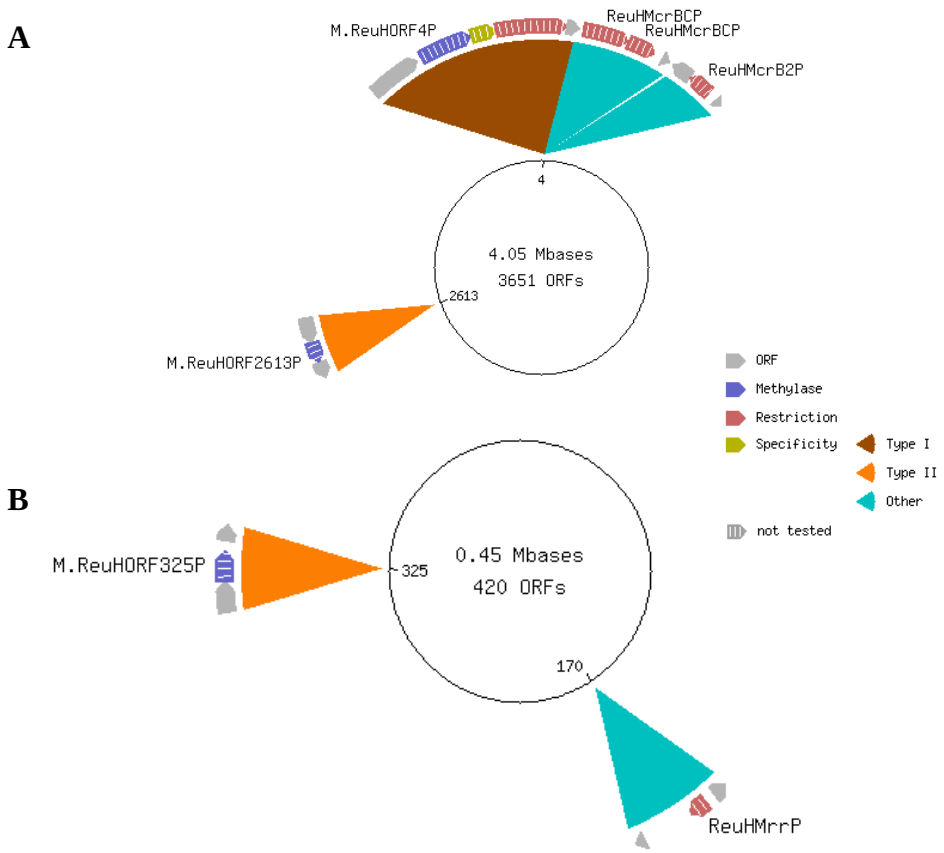


Figure 0.8 – Location of restriction-modification genes of C. necator H16 (Ralstonia eutropha H16 on ReBASE Genomes)

A. Map of chromosome 1; **B**. Map of the megaplasmid pHG1. Chromosome 2 has no RM system.

RM systems regulate horizontal gene transfer, cleaving unmethylated or differently methylated foreign DNA (restriction activity) and methylating the host DNA (modification activity) (Tock and Dryden,

2005). They are present in the vast majority of bacteria and archaea and are organised in four types (Roberts *et al.*, 2003). Type I systems form a single enzyme complex, which includes a restriction endonuclease and a modification methyltransferase. These proteins share a conserved sequence and structure and cleave at various locations, located between 100 and 50,000 bp from the recognition site (Kennaway *et al.*, 2012). Type II systems have enzymes specialised in either restriction or modification activity. The proteins show great variability in both sequence and structure, but still cleave at a fixed recognition site (Pingoud and Jeltsch, 2001). This feature makes them extremely useful molecular tools, and most of the restriction enzymes originally come from type II RM systems (Roberts *et al.*, 2003). Type III RM systems, like type I systems, combine the restriction and methylation activities within one complex, but the cleavage site is located at a short distance from the recognition site (Roberts, 1990). Finally, the type IV RM systems are different from the other RM systems, since they only target methylated DNA. For example, Mrr, a type IV restriction endonuclease, restricts adenine- and cytosine-methylated exogenous plasmids in the cell (Waite-Rees *et al.*, 1991).

Chromosome 1 has four RM complexes, a type I RM system, a type II RM system and two type IV RM systems (Schwartz and Friedrich, 2001; Pohlmann *et al.*, 2006). Chromosome 2 appears not to encode genes with RM activity (Pohlmann *et al.*, 2006), the megaplasmid pHG1 has one type II RM system, and one type IV RM system (Schwartz and Friedrich, 2001). The *hsdR* (H16_A0006) gene encodes the restriction subunit of the Type I MReuHORF4P restriction cluster (pink in **Figure 3.8**). It could be an interesting target for recombineering, potentially yielding an increase in transformation efficiency, as this system severely limits the introduction of unmethylated or differently methylated plasmids (Hobson *et al.*, 2008).

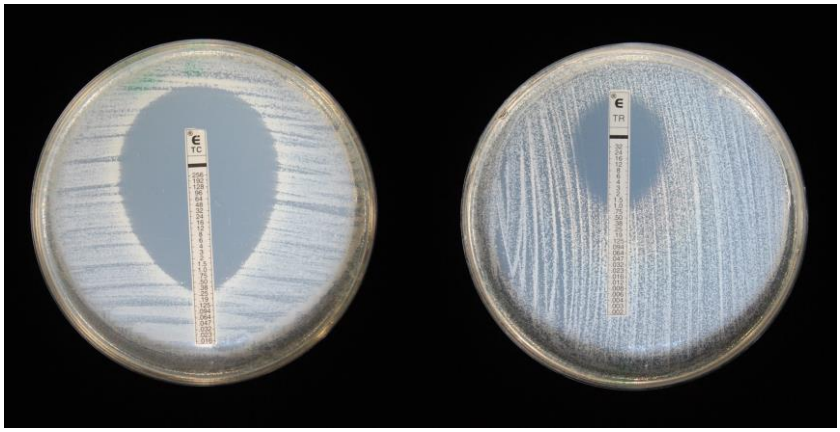
3.2.3 Optimisation of the antibiotic selection marker

Kim *et al.* demonstrated the need for a suitable selection marker, to select for the first recombination event (Kim *et al.*, 2014). Several antibiotic resistance cassettes were therefore tested to determine a suitable selection marker for use in *C. necator*. The trimethoprim-resistance gene, *dhfR,* is 237 bp-long and could be a good candidate for selection in *C. necator*.

Natural susceptibility of *C. necator* to trimethoprim was assessed utilising the MIC (Minimal Inhibitory Concentration) with an Etest (bioMérieux©) on both LB and 0.4% (w/v) sodium gluconate minimum medium (0.4% SG-MM), as both these media are routinely used for growth of *C. necator* in the laboratory. Previous studies have shown that *C. necator* is sensitive to tetracycline (Christian Arenas, personal communication). Tetracycline was therefore utilised as a sensitivity control in the MIC.

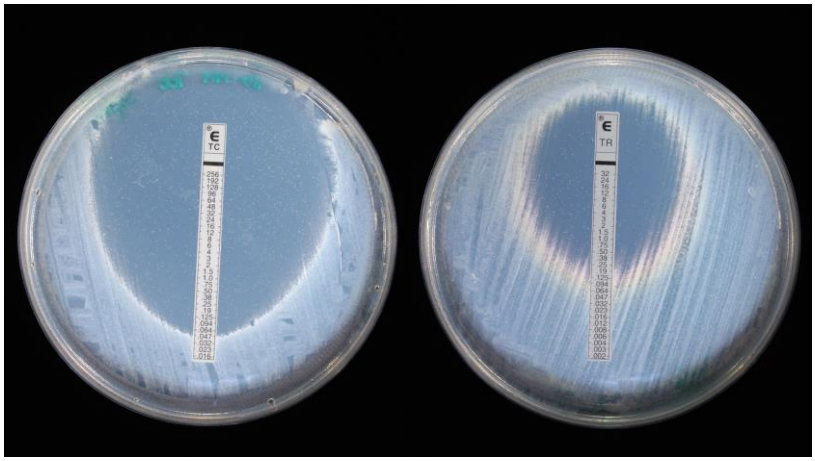
Tetracycline MICs were four to six times lower than trimethoprim MICs reflecting the higher sensitivity of *C. necator* to tetracycline (**Figure 3.9**). Although both antibiotics have a broad spectrum, MICs were also lower on 0.4% SG-MM than on LB. *C. necator* was shown to be sensitive to trimethoprim at 1.5 µg/mL on LB and 0.38 µg/mL on 0.4% SG-MM. Increasing concentrations of trimethoprim were tested in LB agar and 60 µg/mL trimethoprim was determined to be a good working concentration for a clean selection of trimethoprim-resistant colonies.

A



Tet MIC : 0.38 µg/mL Tm MIC : 1.5 µg/mL

B



Tet MIC : 0.064 µg/mL Tm MIC : 0.38 µg/mL

Figure 0.9 – Determination of the tetracycline and trimethoprim MICs of C. necator

MICs were determined in (**A**) LB and (**B**) 0.4% (w/v) SG-MM.
Tet tetracycline; Tm trimethoprim; MIC minimal inhibitory concentration
Tetracycline MICs on LB and 0.4% SG-MM have previously been determined in our laboratory. It was used here as a control. After incubation, a symmetrical inhibition ellipse appeared along the strip. The MIC value is read from the scale in µg/mL where the pointed end of the ellipse meets the strip.

3.2.4 λ-Red-assisted homologous recombination in C. necator

Two plasmids for λ-Red recombineering were subsequently designed and built: pSLVS1 and pSLBBR1, which included either the low copy-number replicon p15A-pVS1 or the medium copy-number replicon pBBR1, both expressing the λ-Red system under an arabinose-inducible promoter. A suicide plasmid was constructed, pSV200, with a ColE1 origin of replication (that is not functional in *C. necator*), an I-SceI cutting site, two

transcriptional terminators and the *dhfR* gene, conferring resistance to trimethoprim (**Figure 3.10**). These were flanked by the upstream and downstream *hsdR* homology arms for allelic exchange.

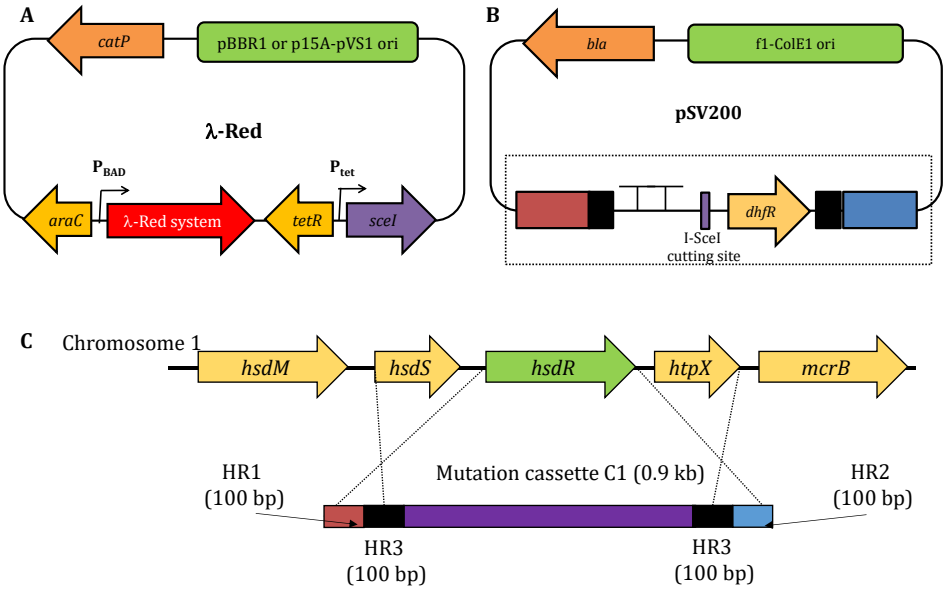


Figure 0.10 – Schematic representation of λ-Red assisted knockout of hsdR

A. Details of the λ-Red plasmid used to express the λ proteins (Gam, Beta and Exo) downstream the arabinose-inducible promoter P_{BAD} and the endonuclease I-SceI under anhydrotetracycline control (P_{tet}). Two versions of the plasmid are available: pSLBBR1 (medium-copy pBBR1) and pSLVS1 (low-copy p15A-pVS1).
B Details of the plasmid pSV200 bearing the mutation cassette C1 (spotted-frame box). The homology regions HR1 (red), HR2 (blue) and HR3 (black), as referred in **Figure 3.2,** frame the double transcriptional terminators (double black 'T' s), the I-SceI cutting site (purple) and the *dhfR* gene, conferring resistance to trimethoprim.
catP: chloramphenicol acetyl-transferase; *araC*: arabinose operon regulatory protein; ; *tetR*: tetracycline repressor protein; P_{tet}: anhydrotetracycline-inducible tetA promoter ; *sceI*: endonuclease I-SceI *; bla*: beta-lactamase; *dhfR*: dihydrofolate reductase, *tetR/A:* confers resistance to tetracycline.
C. Schematic representation of the first recombineering event driven by the recombineering helper system in chromosome 1 *of C. necator* H16.
hsdM: Type I RM system methylation subunit, *hsdS*: Type I RM system specificity subunit, *hsdR*: Type I RM system restriction subunit, *htpX*: M48 family peptidase, *mcrB*: 5-methylcytosine-specific restriction enzyme B, HR: homology region. HR1 and H3 on one side and HR3 and HR2 on the other side overlap over 50 bp. The homology regions are complimentary to the sequences immediately upstream or downstream *hsdR*.

C. necator H16 cells were transformed with both the λ-Red pSLVS1 and pSLBBR1 plasmids. Herein these strains will be referred to as $H16_{pVS1}$ and $H16_{pBBR1}$. The λ-Red system was subsequently induced with 10 mM

arabinose. Both $H16_{pVS1}$ and $H16_{pBBR1}$ were transformed with either the suicide plasmid pSV200, containing the mutation cassette or the 0.9 kb-long linear cassette only, named C1, generated by restriction enzyme digestion or by PCR amplification, using pSV200 as a template. Though linear DNA has been successfully utilised in *E. coli* to generate gene knockouts (Thomason, Costantino and Court, 2016), this may not be replicable in H16. Hence, both circular and linear DNA allelic replacement cassettes were tested in *C. necator*.

Once transformed, cells were recovered for 3 h in SOC media supplemented with 10 mM arabinose to maximise the expression of the λ proteins. Cells were spread onto LB agar supplemented with chloramphenicol (to maintain the λ-Red vector) and trimethoprim (to identify recombinants).

The number of recombinants obtained was low (on average between 0 and 15 CFUs) except for $H16_{pVS1}$ transformed with digested C1 (up to 85 CFUs) (**Figure 3.11**). Since C1 was excised from the plasmid pSV200, isolated directly from *C. necator*, methylation of the donor DNA has been conserved and inherently favours better processing of the exogenous DNA in this strain.

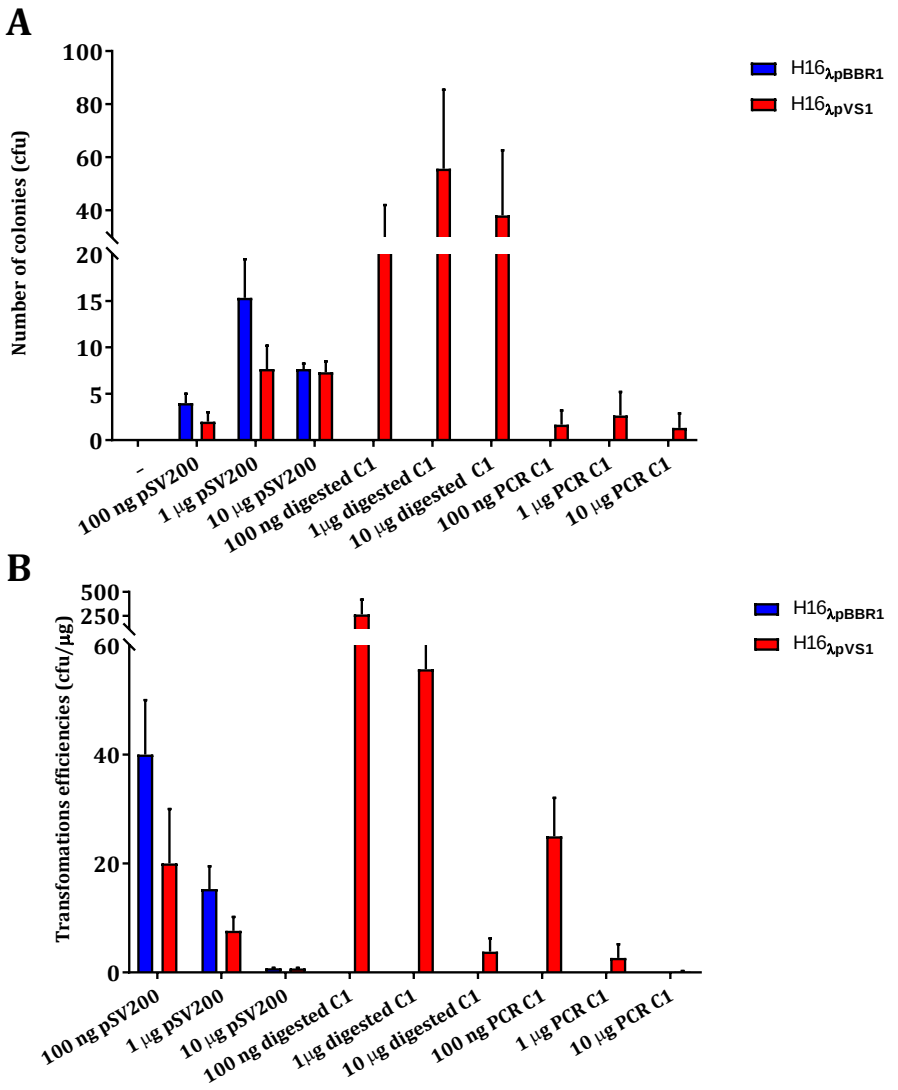


Figure 0.11 – Transformation of C. necator H16$_{pVS1}$ and H16$_{pBBR1}$ strains with increasing amounts of donor DNA targeting hsdR

A. Growth of *C. necator* H16$_{pBBR1}$ (in blue) and H16$_{pVS1}$ (in red) cells transformed with increased amounts of donor DNA and incubated for 48 h at 30°C on LB agar plates supplemented with 50 µg/mL chloramphenicol and 60 µg/mL trimethoprim (LBCTm).
B. Transformation efficiencies of *C. necator* H16$_{pVS1}$ and H16$_{pBBR1}$ cells transformed with increased amounts of donor DNA.

Transformants were re-streaked onto LBCTm agar plates. No colonies grew on the second set of LBCTm plates, suggesting that the first colonies were escapers and their DNA had not recombined with the mutation cassette. Indeed, recombineering relies heavily on a good transformation efficiency. To improve the transformation efficiency in *C. necator*, several parameters were optimised further:

- extended recovery times up to 8 h,
- overnight incubation with arabinose prior to cell treatment to enhance the concentration of λ proteins,
- amplification and purification of the cassette prior to transformation to minimise DNA degradation due to thawing.

However, none of the above led to a significant improvement in recombineering efficiency in *C. necator*, suggesting that the λ-Red system may never work in this organism without further adaptations. The λ-Red system has been very successfully used in *Enterobacteriaceae,* particularly *E. coli* and close relatives (Murphy, Campellone and Poteete, 2000; Murphy and Campellone, 2003; Poteete, 2008). However, the use of this bacteriophage-derived system outside of this family has been limited, perhaps due to host-specific limitations (Derbise *et al.*, 2003; Lesic and Rahme, 2008). Therefore, alternative strategies were investigated.

3.3 RecET-assisted homologous recombination in *C. necator*

3.3.1 Design of a recET system in C. necator

The *recET* system, derived from the Rac prophage, is another commonly used recombineering system for gene editing (Zhang *et al.*, 1998). It has been successfully applied to multiple host systems such as *E. coli*, *Pseudomonas syringae* and *Zymomonas mobilis* (Lesic and Rahme, 2008; Wu *et al.*, 2017). The assessment of the system in *C. necator* is described here.

Given that the chromosomal context of the *hsdR* gene may make a gene deletion difficult, we selected another candidate, the *phaC* gene. The *phaC* deletion has been well characterised in the PHB-negative strain, *R. eutropha* PHB-4, derived from the wild strain H16 (Raberg *et al.*, 2014). A PHB mutant can be easily identified using the dye Nile red on minimum medium agar (Gorenflo *et al.*, 1999). The lipophilic Nile red stains PHB granules, which then appear fluorescent under blue light. Hence, the detection of *phaC* mutants would be greatly facilitated. In addition, both the Δ*phaC* and Δ*phaCAB* mutant strains have already been generated in our laboratory and confirmed by Sanger sequencing, providing a good control reference for comparison (Dr Christian Arenas, personal publication).

The mutation cassette was subsequently redesigned: the *dhfR* gene was replaced with the spectinomycin resistance gene *aadA*, and the flanking homology regions were extended to 950 bp (**Figure 3.12**). Homologous recombination (HR) efficiency can be significantly improved if the length of the homology arms is increased (Li *et al.*, 2014). Short fragments are more prone to degradation during uptake or processing and recombination is generally but not systematically dependent on the length of the homology arms (Shen and Huang, 1986; Khasanov *et al.*, 1992).

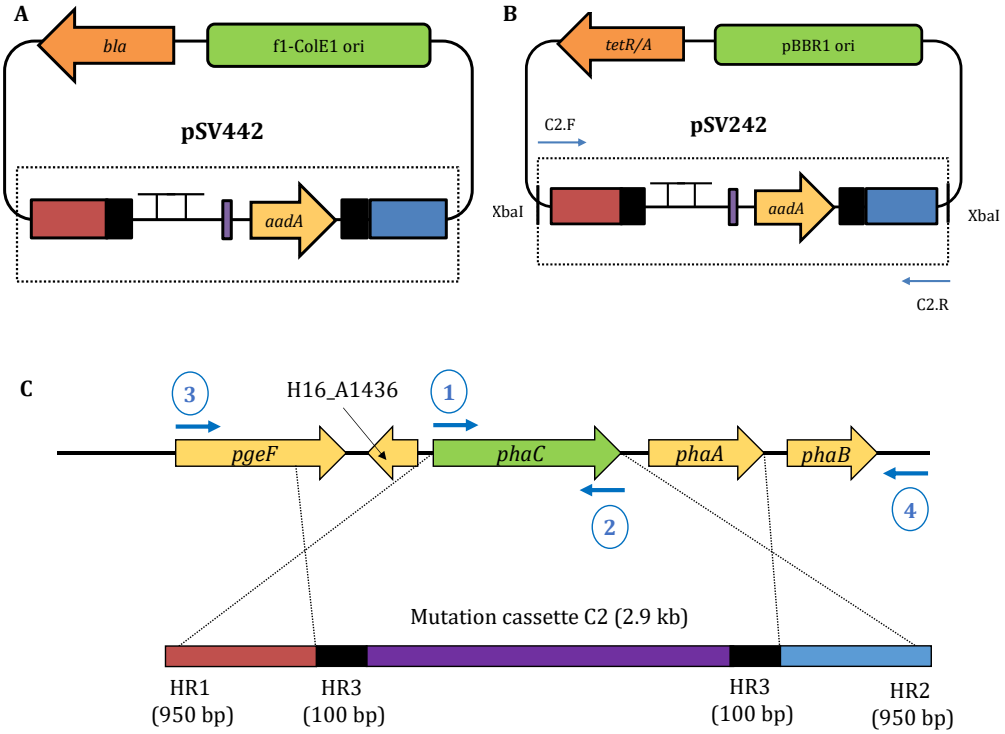


Figure 0.12 – Deletion of phaC by homologous recombination in chromosome 1 of C. necator H16

A. pSV442 is a high-copy suicide plasmid carrying the mutation cassette C2 (spotted-frame box), composed of the 950 bp homology regions HR1 (red) and HR2 (blue), the 100 bp HR3 sequence (black), the double transcriptional terminators (double black 'T's), the I-SceI cutting site (purple) and the *aadA* gene, conferring resistance to spectinomycin.
B. pSV242 is a *C. necator*-compatible version of pSV442 with a medium-copy origin. C2 is extracted from pSV242 by restriction digestion using XbaI or by PCR with the primer pair C2.F/R, amplifying a 2.9 kb product.
bla: beta-lactamase; *aadA*: streptomycin 3''-adenylyltransferase, *tetR/A:* confers resistance to tetracyclin.
C. Location of the *phaC* gene on chromosome 1.
pgeF: peptidoglycan editing factor, *phaC*: poly(3-hydroxybutyrate) polymerase, *phaA*: acetyl-CoA acetyltransferase, *phaB*: acetoacetyl-CoA reductase; HR: homology region. The product of H16_A1436 is annotated as a hypothetical protein in GenBank. The homology regions are complimentary to the sequences immediately upstream or downstream *phaC.* 1: primer *PhaC*.IN.F, 2: *PhaC*.IN.R, 3: *PhaC*.OUT.F, 4: *PhaC*.OUT.R. Primer pair 1+2 produces a PCR band of 500 bp whilst Primers 3 and 4 amplify a product of 3.8 kb from the wild type chromosome, 3.2 kb when the mutation cassette has been inserted.

3.3.2 Cupriavidus-derived recET recombination system (recET-C)

A BLAST-N analysis was performed on the genomes of *Cupriavidus* spp. to identify Beta or RecT-like proteins, to identify a more suitable recombineering system for H16. The *recE* and *recT* genes of *Cupriavidus gilardii CR3* were selected, synthesised (Eurofins™) and asseled into a low-

copy-number vector called p*recET-C* (**Figure 3.13**), which was transformed into *C. necator* H16 to generate H16$_{recET-C.}$

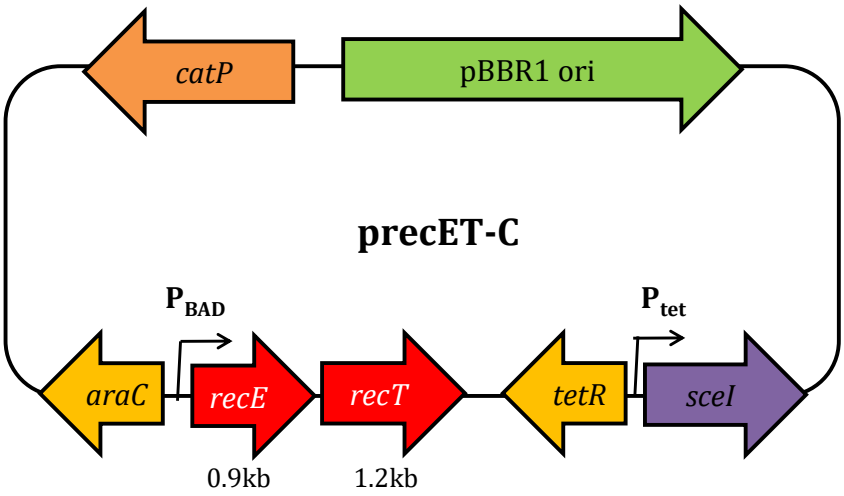


Figure 0.13 – Schematic representation of the medium-copy helper plasmid precET-C

The expression of the *C. gilardii recE* and *recT* genes is controlled by the arabinose-inducible promoter P_{BAD} whilst the endonuclease I-SceI is downstream the anhydrotetracycline-induced P_{tet}.
catP: chloramphenicol acetyl-transferase; *araC*: arabinose operon regulatory protein; *recE*: exodeoxyribonuclease VIII; *recT*: recombinase; *tetR*: tetracycline repressor protein; *sceI*: endonuclease I-SceI.

H16$_{recET-C}$ was transformed with the suicide plasmid pSV442 or the cassette C2, generated via PCR or enzymatic digestion. To further improve the recombination efficiency, the H16$_{recET-C}$ cells were grown at 30°C, the optimal growth temperature for *C. necator*, and 37°C. Previous studies have shown that an elevated temperature can modify protein synthesis, modulating the RM system, thus preventing the degradation of non-methylated exogenous DNA (Wards and Collins, 1996).

The number of transformants remained low, similar to the results obtained for the λ-Red system (≤10 CFUs) (**Figure 3.14**).

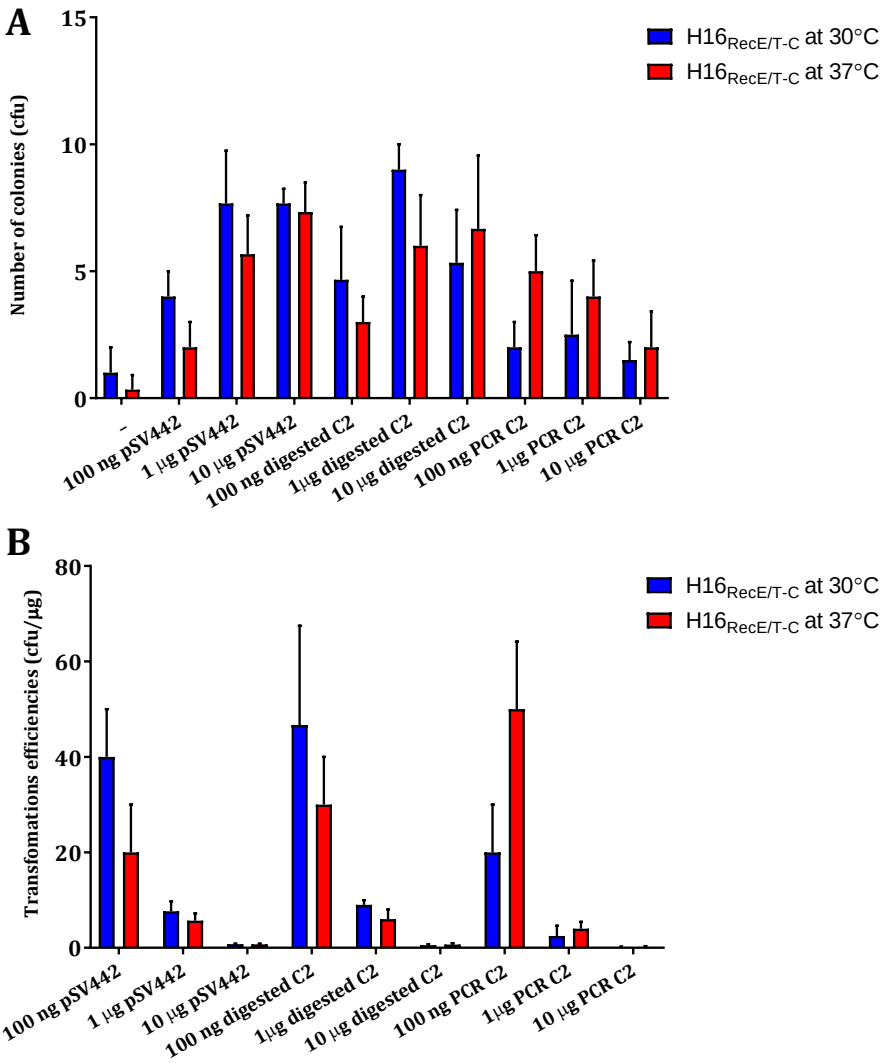


Figure 0.14 – Transformation of H16$_{recET\text{-}C}$ with increasing amounts of donor DNA targeting phaC

Cells were prepared for electrocompetence at 30°C (blue) or 37°C (red).
A. Growth of H16$_{recET\text{-}C}$ cells transformed with donor DNA and incubated for 48 h at 30°C on LB agar plates supplemented with 50 µg/mL chloramphenicol and 300 µg/mL spectinomycin (LBCS).
B. Transformation efficiencies of H16$_{recET\text{-}C}$ cells transformed with donor DNA.

Although *recET*-driven recombination in *E. coli* resulted in up to 80 x 10^6 cells (Zhang *et al.*, 1998), no more than six clones could be obtained in *Z. mobilis* (Wu *et al.*, 2017) and similar results were seen in *C. glutamicum* (Huang *et al.*, 2017), with a maximum number of 60 colonies being generated, before further optimisation of the system (Huang *et al.*, 2017; Wu *et al.*, 2017). A strong RM system could account for the low

transformation efficiencies *in C. necator* (Schwartz and Friedrich, 2001; Pohlmann *et al.*, 2006), and growth at 37°C did not enhance the transformation efficiency of the strain. The small number of colonies obtained were re-streaked onto fresh LBCS plates, reducing the possibility of false positives being carried over. On average, 50% of the colonies selected grew on the second set of plates (**Figure 3.15**).

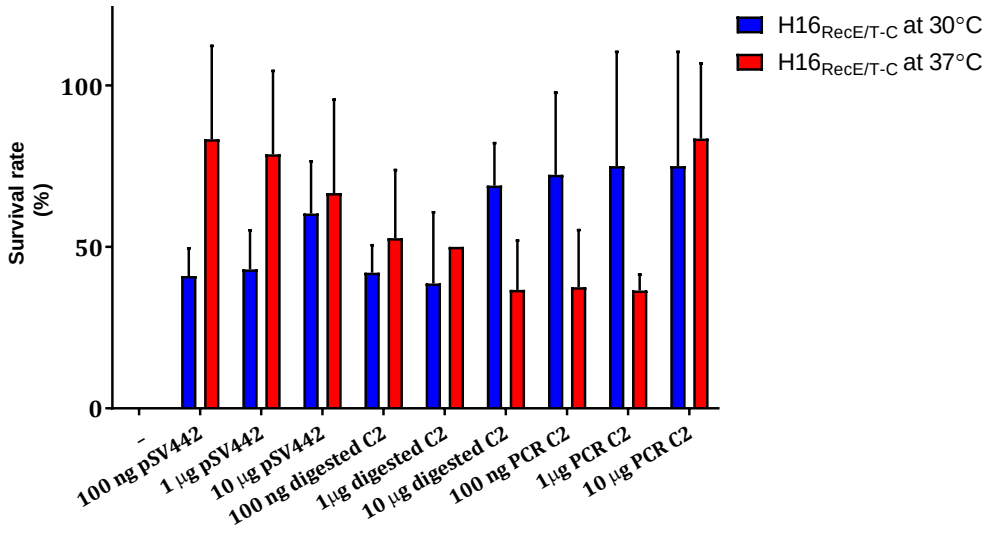


Figure 0.15 – Survival rate of C. necator H16$_{recET\text{-}C}$ colonies after replating onto fresh LBCS agar plates.

Cells were prepared for electrocompetence at 30°C (blue) or 37°C (red).
The graph shows the average data from three biological replicates with their standard deviation.

Selected colonies were screened for integration of the mutation cassette into the chromosome, utilising colony PCR (cPCR), targeting the *aadA* gene. For 6 out of the 9 colonies screened, a band of 800 bp could be detected, the other six colonies had a similar PCR profile to the negative control in Lane 10 (**Figure 3.16**).

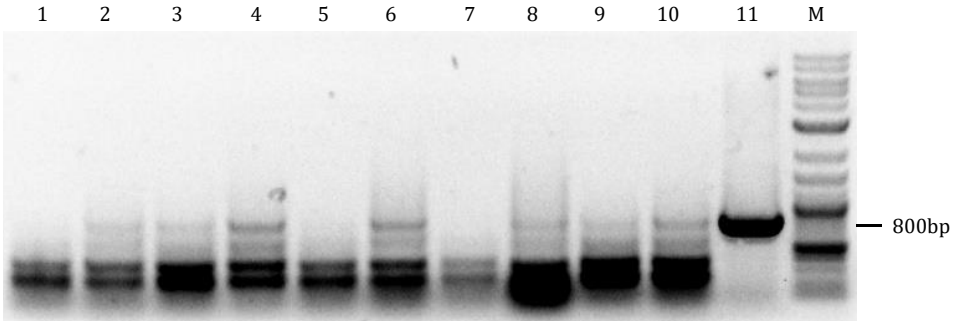


Figure 0.16 – Screening of H16$_{recET\text{-}C}$ transformants re-streaked twice on LBCS via cPCR utilising SpecR.F/R primer pair.

1-8: H16$_{recET\text{-}C}$ cells transformed with 1 µg PCR cassette C2; 9: H16$_{recET\text{-}C}$ cells transformed with 1 µg pSV442; 10: H16$_{recET\text{-}C}$ cells, used as negative control; 11: PCR control on 100 ng plasmid pSV242; M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb).

Further colonies were screened from both the PCR and plasmid transformations, however, only two of the colonies in lanes 12 and 13 had an 800 bp band, corresponding to the *aadA* gene (**Figure 3.17A**). PCR was utilised to detect deletion of the *phaC* gene in these strains. However, all the colonies still had a band of 500 bp, similar to the wild type, lane C (**Figure 3.17B**), suggesting that the cassette may have integrated at an alternative site in the chromosome. Nonetheless, further PCR amplification with the C2. F/R primers, designed to detect the full cassette, yielded no positive colonies. The *Cupriavidus recET* system proved unsuccessful, under the conditions tested.

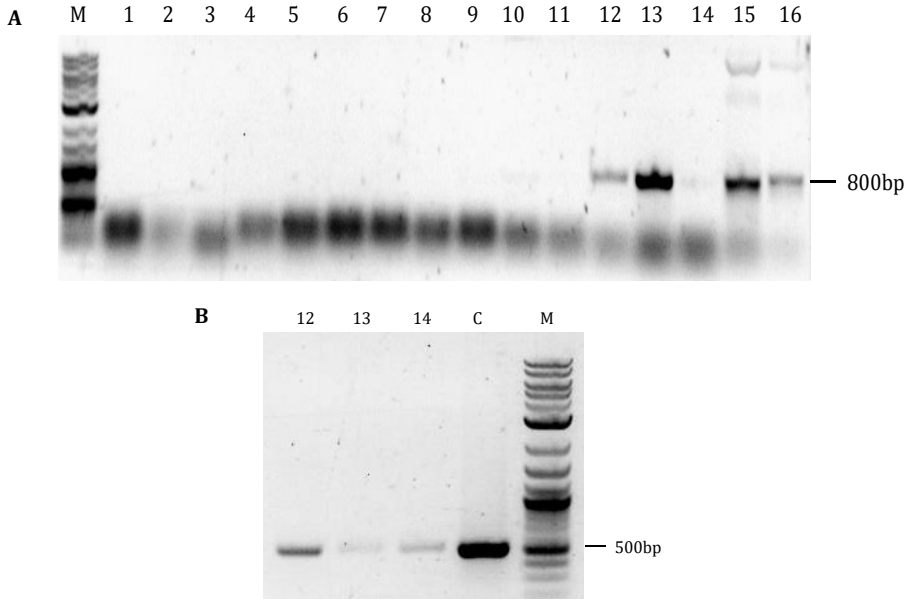


Figure 0.17– Screening of H16$_{recET-C}$ transformants re-streaked twice on LBCS via cPCR

A. Screening of the cassette integration utilising SpecR.F/R.
B. Screening of the *phaC* deletion with *PhaC*_IN.F/R. *PhaC*_IN primers amplify the 500 bp N-terminal constitutive domain of *phaC*, *phaC_N*.
1-6: H16$_{recET-C}$ cells; 7-11: H16$_{recET-C}$ cells transformed with 10 µg PCR cassette C2; 12-14: H16$_{recET-C}$ cells transformed with 1 µg pSV442; 15-16: PCR control on 100 ng plasmid pSV242; C: PCR control on 100 ng H16 total DNA; M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb).

3.3.3 A Pseudomonas-derived recET recombination system (recET-P)

The *recET* genes from *Pseudomonas syringae* pv. *syringae* have already been utilised to successfully generate multiple mutants in *Pseudomonas syringae* pv. tomato DC3000 (Swingle *et al.*, 2010). Gene replacement of a 1.3 kb TonB-dependent siderophore receptor, utilising the *recET* system, resulted in 100% efficiency and a frequency of 45 recombinants per 10^8 viable cells, using 1 µg of a PCR-generated dsDNA tetracycline-resistant cassette, flanked by 700 bp homology arms.

The synthesised *recET-P* genes (Psyr_2820- Psyr_2821) were cloned into the p*recET-P* vector, under an arabinose inducible promoter (**Figure 3.18**).

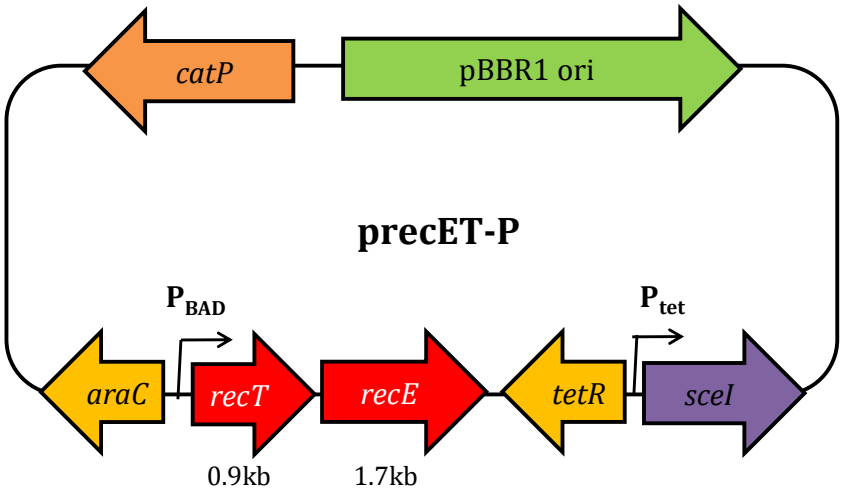


Figure 0.18 – Schematic representation of the medium-copy helper plasmid precET-P

The expression of the *recE* and *recT* genes from *P. syringae* is controlled by the arabinose-inducible promoter P_{BAD} whilst the endonuclease I-SceI is downstream of the anhydrotetracycline-induced P_{tet}.
catP: chloramphenicol acetyl-transferase; *araC*: arabinose operon regulatory protein; *recE*: exodeoxyribonuclease VIII; *recT*: recombinase; *tetR*: tetracycline repressor protein; *sceI*: endonuclease I-SceI.

C. necator was transformed with p*recET-P* and the resulting strain was called H16$_{recET\text{-}P}$. The H16$_{recET\text{-}P}$ strain was grown at 30°C and prepared for electroporation as detailed in **Chapter 2**. It was then transformed with either the suicide plasmid pSV442 or the linear cassette C2.

The number of transformants obtained from the H16$_{recET\text{-}P}$ strain was also low and comparative to the ones obtained from previous attempts to utilise both the *recET-C* and the λ-Red systems (**Figure 3.19**). Interestingly, the suicide plasmid pSV442 and the PCR-amplified cassette C2, were both efficient DNA donors when compared to the digested cassette. In contrast to the previous transformations, there is a direct correlation between the quantity of DNA transformed and number of CFUs. However, transformation efficiencies remained very low. In all the homologous recombination (HR) experiments conducted, high error bars reveal a variation in transformation from one replicate to another. The use of different batches of cells and the different conditions utilised to optimise

transformation efficiency may explain this phenomenon. Nonetheless, the number of colonies varies within no more than one order of magnitude.

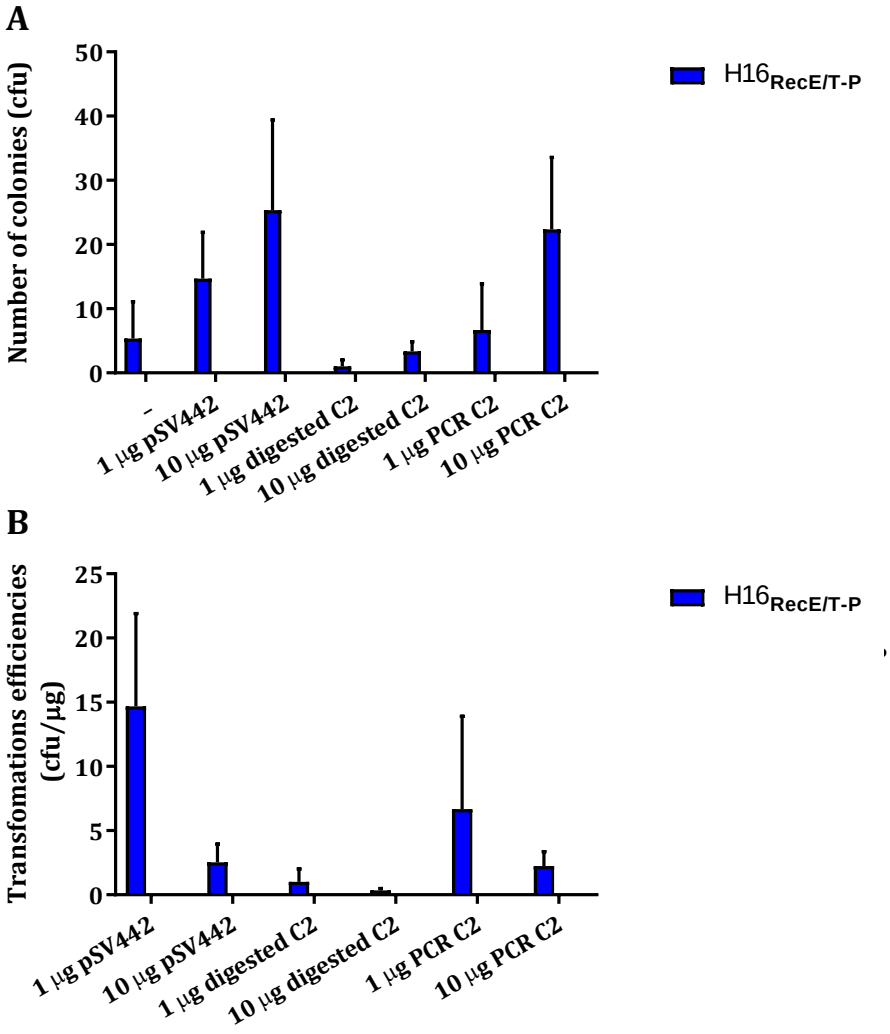


Figure 0.19 – Transformation of H16$_{recET-P}$ with increasing amounts of donor DNA targeting phaC

Growth (**A**) and transformation efficiencies (**B**) of H16$_{recET-P}$ cells transformed with increasing amounts of donor DNA.

Selected colonies were re-streaked onto LBCS agar, however, only a small number of colonies were able to regrow following re-streaking (**Figure 3.20**). Highlighting the high level of spontaneous antibiotic resistance mutants isolated from *C. necator* during the transformation.

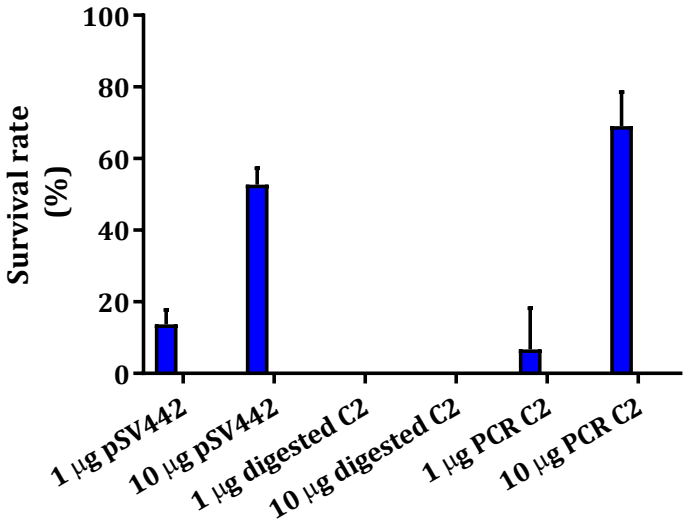


Figure 0.20 – Percentage of survival of H16$_{recET-P}$ after re-streaking onto fresh LBCS agar plates

Selected colonies from the plates with the highest number of transformants (10 µg donor DNA, either plasmid or linear cassette) were screened by PCR for integration of the cassette at the *phaC* locus (**Figure 3.21**). The *aadA* gene was amplified from all colonies tested, from cells transformed with both the suicide vector pSV442 and the linear cassette C2.

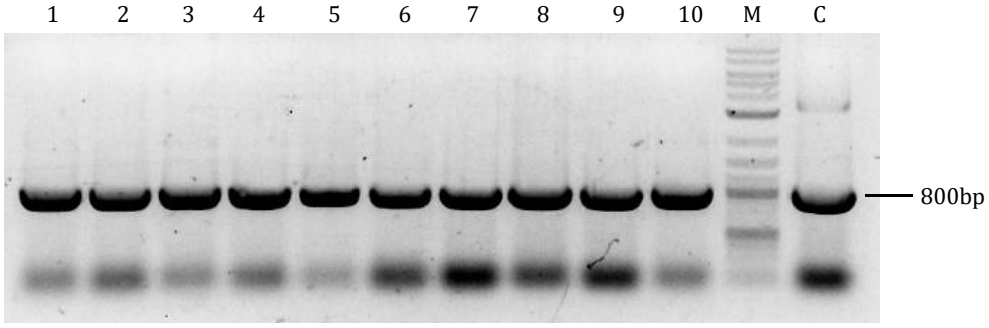


Figure 0.21 – Screening of H16$_{recET-P}$ transformants re-streaked twice on LBCS via cPCR utilising SpecR.F/R primer pair.

1-5: H16$_{recET-P}$ cells transformed with 10 µg PCR cassette C2; 6-10: H16$_{recET-C}$ cells transformed with 10 µg pSV442; C: PCR control on 100 ng plasmid pSV242; M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb).

The colonies were then screened with the *phaC_N* primers to detect the target gene *phaC*. The selected colonies generated a different profile to the wild type, utilising H16 total DNA as a control. A band of 500 bp could not be detected in the 10 colonies screened, although the PCR profile was difficult to interpret (**Figure 3.22A**). Therefore, a second set of external *phaC* primers were utilised, to determine whether the full CDS of the *phaC* gene had been successfully replaced with the mutation cassette. Amplification from wild type H16 genomic DNA should yield a band of 3.8 kb, including the *phaC* gene (1.7 kb). The introduction of the cassette should generate a slightly smaller band of 3.2 kb. Subsequent screening of the Spec-positive transformants with the external primers *PhaC*_OUT.F/R produced a unique PCR profile, the main band being approximately 2.5 kb (**Figure 3.22B**). Given the band was smaller than expected, bands 3, 5, 9 and 10 were gel-extracted and analysed by Sanger sequencing, however the sequences did not align with the *phaC* operon.

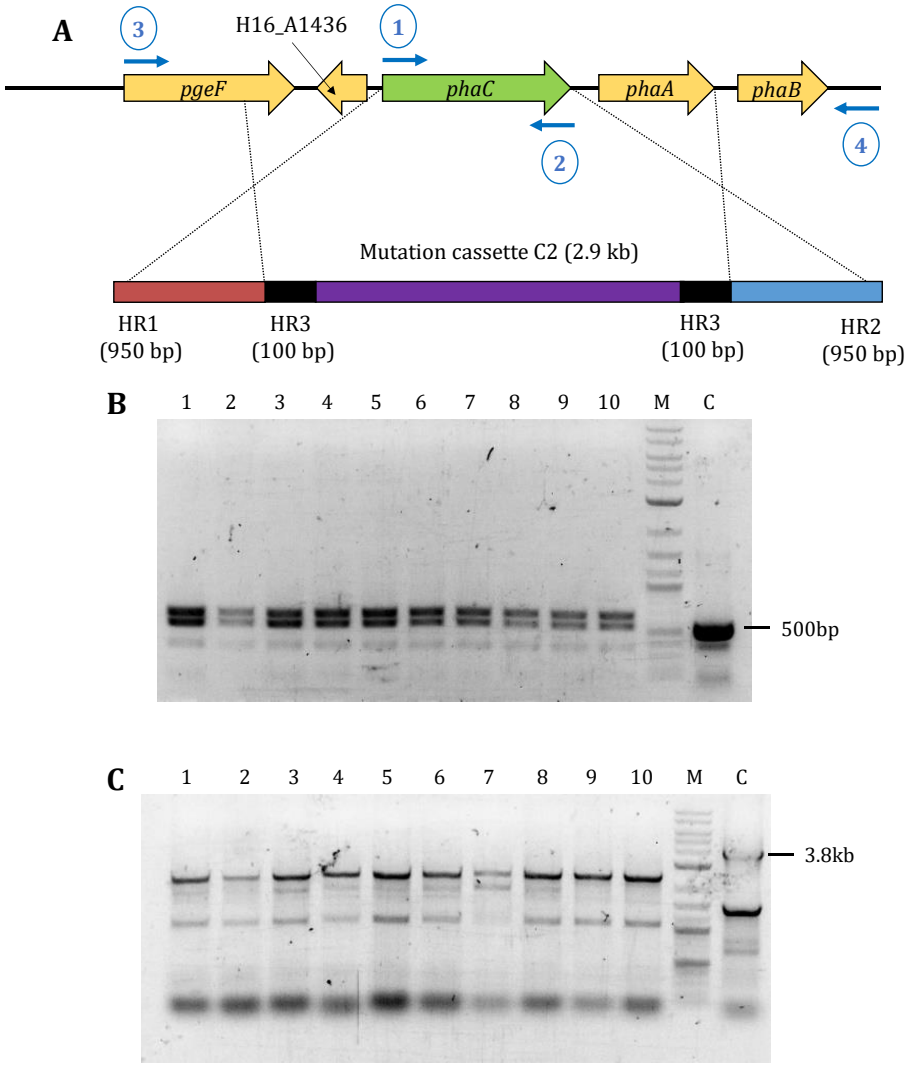


Figure 0.22 – Screening of H16$_{recET\text{-}P}$ transformants restreaked twice on LBCS via cPCR

A. Schematic representation of the recombination of the mutation cassette C2 at the *phaC* locus in chromosome 1 *of C. necator* H16.
1: primer *PhaC*.IN.F, 2: *PhaC*.IN.R, 3: *PhaC*.OUT.F, 4: *PhaC*.OUT.R. Primer pair 1+2 produces a PCR band of 500 bp whilst Primers 3 and 4 amplify a product of 3.8 kb from the wild type chromosome, 3.2 kb when the mutation cassette has been inserted.
Screening with *PhaC*_IN primers (**B**) should yield a band at 500 bp overlapping the sequence of the N-terminal constitutive domain *phaC_N* whilst the *PhaC*_OUT pair (**C**) should generate a 3.8 kb product covering *phaC* and the left and right homology regions flanking the gene.
1-5: H16$_{recET\text{-}P}$ cells transformed with 10 µg PCR cassette C2; 6-10: H16$_{recET\text{-}C}$ cells transformed with 10 µg pSV442; C: PCR control on 100 ng H16 total DNA; M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb).

Total DNA was extracted from strains 3, 5, 9 and 10 and the PCR was repeated utilising both internal and external *phaC* primers. Two controls were utilised, H16 and *C. metallidurans* CH34, the latter being a closely

related strain included to pre-empt contamination. Total DNA was extracted from both controls and 100 ng of H16 or CH34 DNA was utilised in each reaction. Surprisingly, the *phaC_N* internal primers amplified a 500 bp band across all the strains tested. However, the external *phaC* primers still amplified a 2.5 kb product (Lanes 1-4, **Figure 3.23**) The PCR profile for the selected strains does not match the one obtained for CH34, so contamination with this bacterium was excluded.

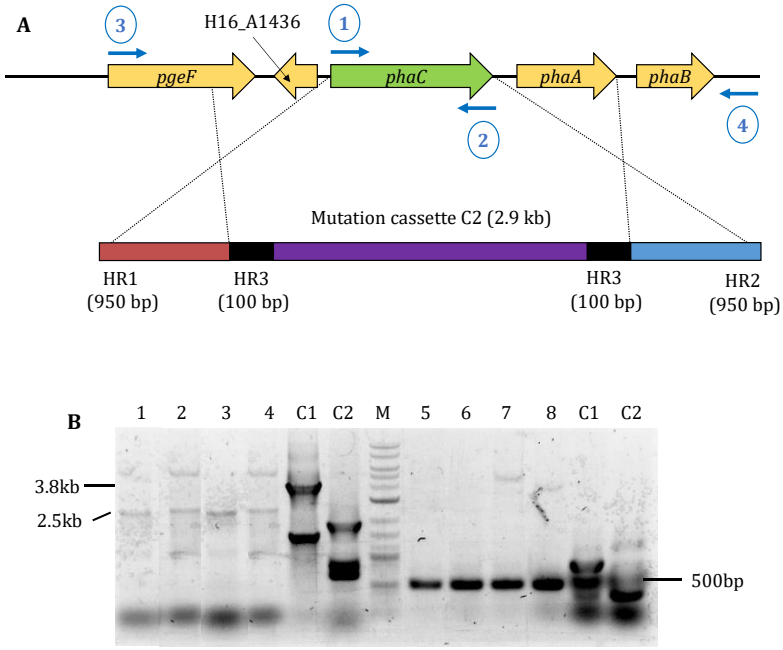


Figure 0.23 – Screening of DNA isolated from H16$_{recET\text{-}P}$ transformants via cPCR

A. Schematic representation of the recombination of the mutation cassette C2 at the *phaC* locus in chromosome 1 *of C. necator* H16.
1: primer *PhaC*.IN.F, 2: *PhaC*.IN.R, 3: *PhaC*.OUT.F, 4: *PhaC*.OUT.R. Primer pair 1+2 produces a PCR band of 500 bp whilst Primers 3 and 4 amplify a product of 3.8 kb from the wild type chromosome, 3.2 kb when the mutation cassette has been inserted.
B.1-2 and 5-6: 100 ng of total DNA extracted from H16$_{recET\text{-}P}$ cells transformed with 10 µg PCR cassette C2; 3-4 and 7-8: 100 ng of total DNA extracted from H16$_{recET\text{-}P}$ cells transformed with 10 µg pSV442; C1: PCR control on 100 ng H16 total DNA; C2: PCR control on 100 ng CH34 total DNA; M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb).

Despite successive attempts a *phaC gene* knockout could not be generated. Due to time constraints these mutants were not pursued further,

however, further investigation could shed some light on *recET* mediated recombination in this system.

3.4 The effect of gene knockouts on homologous recombination in *C. necator*

3.4.1 Homologous recombination in the Δmmr mutant strain

The *mrr* (RS32315) gene belongs to the ReuHMrrP RM system, located on the megaplasmid pHG1 (**Figure 3.24**). It codes for a putative type II endonuclease (Schwartz and Friedrich, 2001), although Mrr has been marked as a type IV restriction endonuclease (Roberts *et al.* 2003).

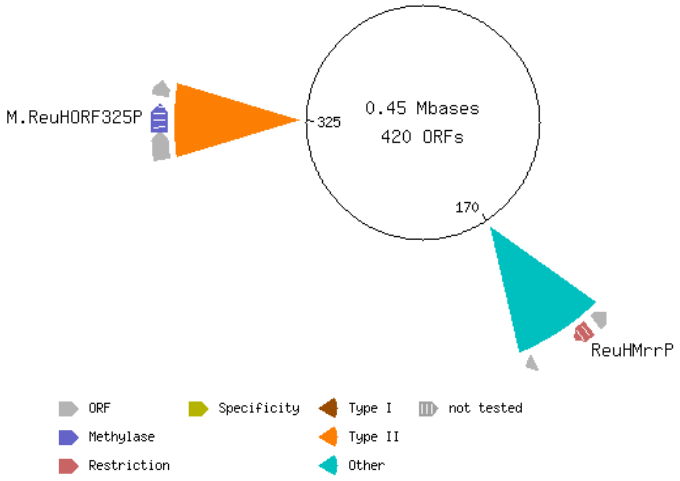


Figure 0.24 – Restriction map of the megaplasmid pHG1 (Ralstonia eutropha H16 on ReBASE Genomes)

pHG1 hosts two RM systems: one type II and one type IV located at diametrical opposites of the plasmid.

If the *mrr* gene does belong to the type IV endonuclease RM system, it would mean that adenine- and cytosine-methylated DNA, generated through cloning in *E. coli*, would not be readily accepted by *C. necator*. The endonuclease could be responsible for foreign DNA rejection, so inhibiting its activity may improve recombination in *C. necator.*

Type II restriction endonucleases and type IV RM systems possess a similar cleavage site, which can explain the predictive association of *mmr* to these RM types (Bujnicki and Rychlewski, 2001). If *mmr* was acting as a type II endonuclease, it would cleave any foreign DNA targeting

electroporated plasmid DNA and linear DNA, at sequence-specific sites. The deletion of the *mmr* gene could certainly impair the RM system in *C. necator*, potentially allowing for improved transformation efficiencies.

Dr Muhammad Ehsaan generated an *mmr* mutant in our laboratory. p*recET-C* and p*recET-P* were transformed into the Δ*mmr* strain to produce Δ$mmr_{recET\text{-}C}$ and Δ$mmr_{\text{recET-P,}}$ respectively. HR targeting the *phaC* gene was carried out in these strains using the suicide vector pSV442 and the linear cassette C2. No transformants were observed after 72 h incubation at 30°C despite many attempts.

3.4.2 Homologous recombination in a ΔrecJ mutant strain
3.4.2.1 Generation of a ΔrecJ mutant strain

16 proteins from the genome of *C. necator* have currently been annotated as either an exonuclease or as having exodeoxynuclease activity (**Table 3.1**). Some of these exonucleases may inhibit homologous cross-overs and impair homologous recombination (Đermić *et al.*, 2005). In *E. coli,* subsequent knockouts of the *xonA* and *recJ* genes led to an important increase in recombination frequencies, implying that these ssDNA exonucleases contribute to the degradation of ssDNA when they are not protected by single-stranded binding proteins (Lovett, Kolodner and Lovett, 1989). Kufryk *et al.* demonstrated that a Δ*recJ* strain of *Synechocystis* spp. PCC 6803 had a twofold increase in transformation efficiency compared to the wild-type strain (Kufryk *et al.*, 2002).

Table 0.1 – List of exonucleases in C. necator H16

Locus	Name	Protein	Function
	addB	Two domain protein: Inactivated superfamily I helicase and RecB family exonuclease	
H16_A3163		Transcriptional regulator, Exoribonuclease II-family	
H16_A2734	*xseB*	Exodeoxyribonuclease 7 small subunit	Bidirectionally degrades single-stranded DNA into large acid-insoluble oligonucleotides, which are then degraded further into small acid-soluble oligonucleotides.
H16_A0609	*xseA*	Exodeoxyribonuclease 7 large subunit	
H16_A2741	*polA*	DNA polymerase I	In addition to polymerase activity, this DNA polymerase exhibits 5'-3' exonuclease activity.
H16_A2372		Predicted 3'-5' exonuclease related to the exonuclease domain of PolB	
H16_A0900	*dnaE1*	DNA-directed DNA polymerase	3'-5' exonuclease activity
H16_B1062	*dnaE2*	Error-prone DNA polymerase	DNA polymerase involved in damage-induced mutagenesis and translation synthesis (TLS). It is not the major replicative DNA polymerase. 3'-5' exonuclease activity
H16_A2389	*dnaX*	DNA polymerase III subunit gamma/tau	DNA polymerase III is a complex, multichain enzyme responsible for most of the replicative synthesis in bacteria. This DNA polymerase also exhibits 3' to 5' exonuclease activity.
H16_A2467	*dnaQ*	DNA polymerase III subunit epsilon	DNA polymerase III is a complex, multichain enzyme responsible for most of the replicative synthesis in bacteria. The epsilon subunit contain the editing function and is a proofreading 3'-5' exonuclease.
H16_A1570	*holB*	DNA polymerase III delta prime subunit	3'-5' exonuclease activity
H16_B0196	*sbcC*	DNA repair exonuclease, SbcC	
H16_B0195	*sbcD*	Nuclease SbcCD subunit D	SbcCD cleaves DNA hairpin structures. These structures can inhibit DNA replication and are intermediates in certain DNA recombination reactions. The complex acts as a 3'->5' double strand exonuclease that can open hairpins. It also has a 5' single-strand endonuclease activity.
H16_A2391	*cB*	ATP-dependent exoDNAse (Exonuclease V) beta subunit	exonuclease activity ATP-dependent DNA helicase activity
H16_A1171	*cJ*	Single-stranded-DNA-specific exonuclease *RecJ*	5'-3' exonuclease activity

H16_A1368	Exonuclease III	endonuclease activity exodeoxyribonuclease III activity
H16_A0223	Exodeoxyribonuclease III	
H16_A2430	Hypothetical membrane associated protein, contains exonuclease domain	
H16_B2389	Putative exonuclease	
H16_B0768	Exonuclease	DNA-directed DNA polymerase activity

Bioinformatic analysis was utilised to identify the gene encoding the single-strand-DNA-specific exonuclease, RecJ, in *C. necator*. The *recJ* gene was amplified from *C. necator* and cloned into the pLO3 vector, to generate the pLO3_Δ*recJ* plasmid (**Figure 3.25**).

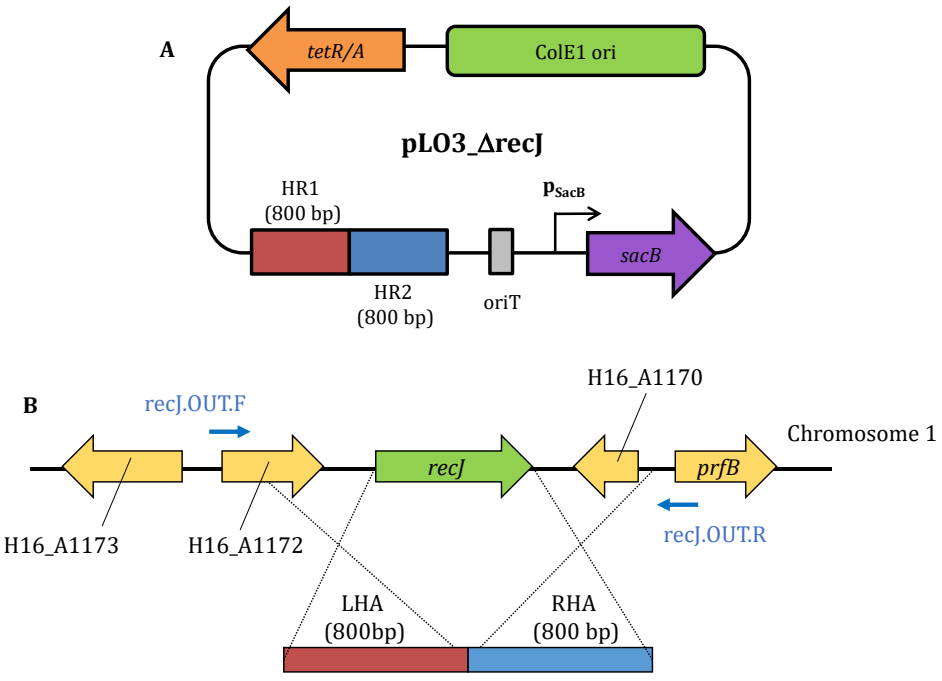


Figure 0.25 – Deletion of recJ by double crossover and sucrose counter-selection in chromosome 1 of C. necator H16

A. Schematic representations of the suicide plasmid used for deletion of *recJ* in *C. necator.* The pLO3_Δ*recJ* vector was electroporated into *E. coli* S17-1 and subsequently conjugated with *C. necator* H16. Positive transformants were selected on tetracycline. The excision of the cassette was triggered in the presence of sucrose, a counter-selective marker, utilizing the *sacB* gene, coding for levansucrase. Cells can keep the mutational insertion or revert to wild-type.

tetR/A: confers resistance to tetracycline; LHA: left homology arm, RHA: right homology arm, *sacB*: levansucrase from *B.subtilis*.
B. Diagrammatic representation of the upstream and downstream region surrounding the *recJ* locus in *C. necator*.
H16_A1173: ABC-type transporter, permease component: LPT family, H16-A1172: Hypothetical membrane associated protein, *recJ*: Single-stranded-DNA-specific exonuclease; H116_A1170: Putative peptidase. The homology regions are symmetrically complimentary to the sequences immediately upstream or downstream *recJ.*

Generation of a Δ*recJ* mutant strain was carried out as described in **Chapter 2**. Tetracycline-sensitive clones were screened for deletion of the *recJ* gene using *recJ*.OUT.F/R. A PCR band of 1.9 kb was amplified from five of the 23 colonies tested and two of these colonies (clones 4 and 8, **Figure 3.26**) were selected for Sanger sequencing, confirming the intended deletions in these clones.

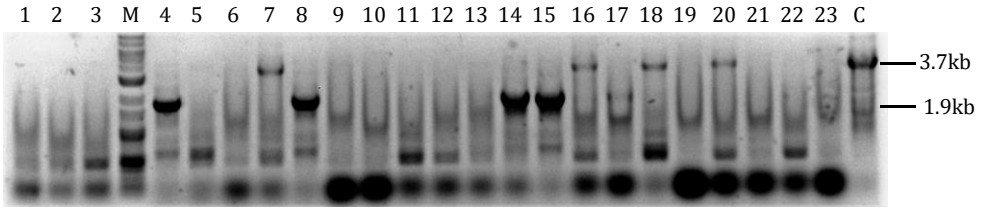


Figure.0.26 – Screening of H16 tetracycline-sensitive transformants, after passage on LSLB enriched with sucrose, via cPCR utilising recJ_OUT.F/R primer pair.

RecJ_OUT primers amplified a 3.7 kb product covering the coding sequence of *recJ* (RS05815) and the 800 bp-long left and right homology regions flanking the gene. A band of 1.9 kb indicate a complete deletion of *recJ*.
1-23: Potential *RecJ*$^{-}$ H16 mutant cells; C: PCR control on 100 ng H16 total DNA; M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb).
Colonies 4, 8, 14, 15 and 17 showed a band at 1.9 kb.

3.5.2.2 Homologous recombination in the ΔrecJ mutant strain

The p*recET-P* plasmid was transformed into the Δ*recJ* strain to produce Δ*recJ*$_{recET\text{-}P}$. H16$_{recET\text{-}P}$ and Δ*recJ*$_{recET\text{-}P}$ were grown at 30°C and prepared for electroporation, as described in **Chapter 2**, with both the linearised cassette, C2 and the suicide plasmid pSV442, targeting the *phaC* gene.

Deletion of the *recJ* gene did increase transformation and recombination efficiencies in *C. necator* (**Figure 3.27A**). A 9-fold increase in efficiency was seen with 1 μg digested cassette (**Figure 3.27B**). The *ΔrecJ* strain was also less susceptible to spontaneous antibiotic-resistance, as no colonies were detected when the transformation was performed with water. However, transformation with 10 μg of the PCR cassette resulted in lower CFUs and transformation efficiencies in $\Delta recJ_{recET\text{-P}}$, than in the $H16_{recET\text{-P.}}$ However the error bars obtained for the $H16_{recET\text{-}P}$ strain are substantial.

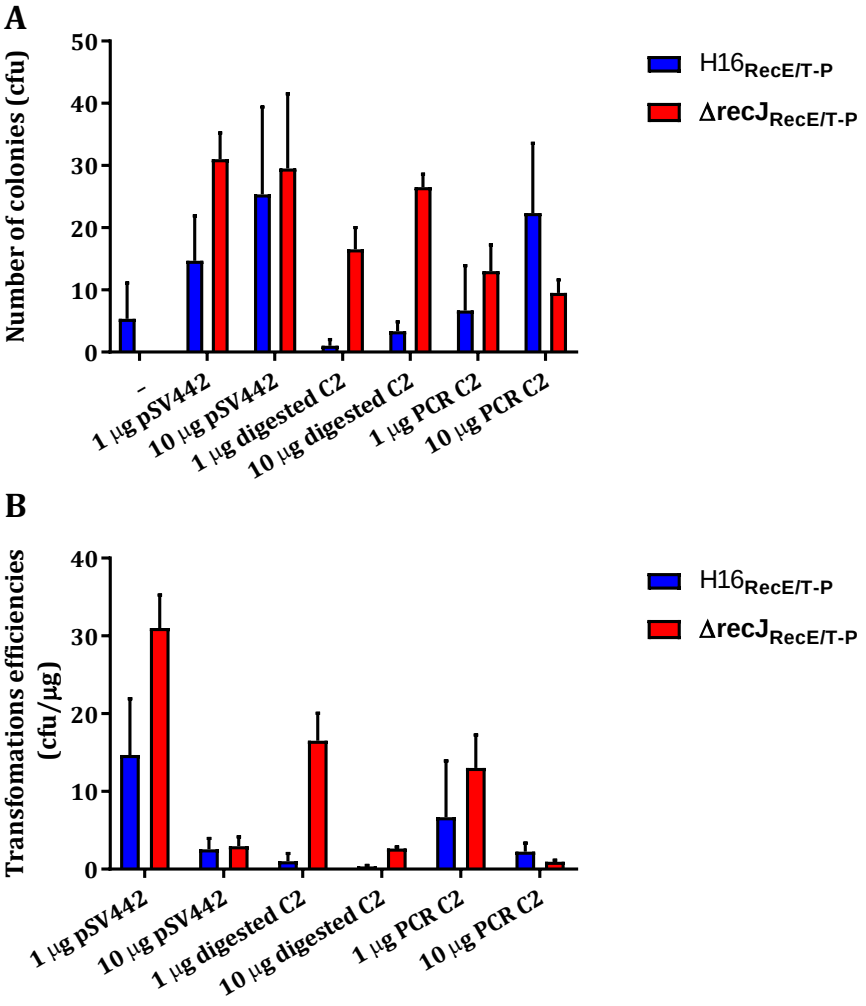


Figure 0.27 – Transformation of H16$_{recET-P}$ and ΔrecJ$_{recET-P}$ with increasing amounts of donor DNA targeting phaC

A. Growth of H16*recET-P* (blue) and Δ*recJ*$_{recET-P}$ (red) cells transformed with increasing amounts of donor DNA (circular plasmid pSV442 or linear cassette C2) and incubated for 48 h at 30°C on LBCS.
B. Transformation efficiencies of *C. necator* H16$_{recET-P}$ and Δ*recJ*$_{recET-P}$ cells transformed with increasing amounts of donor DNA targeting *phaC*.

Selected clones were re-streaked onto fresh LBCS plates and the survival rate was estimated (**Figure 3.28**). The survival rate of the Δ*recJ* strain was higher than the wild type strain across all of the DNA conditions tested (amount and type of DNA, plasmid vs linear DNA). The survival rate being between 20% (1 µg pSV442) and 55% (1 µg PCR C2). These higher

survival ratios may be linked to the higher transformation and recombination efficiencies associated with this strain.

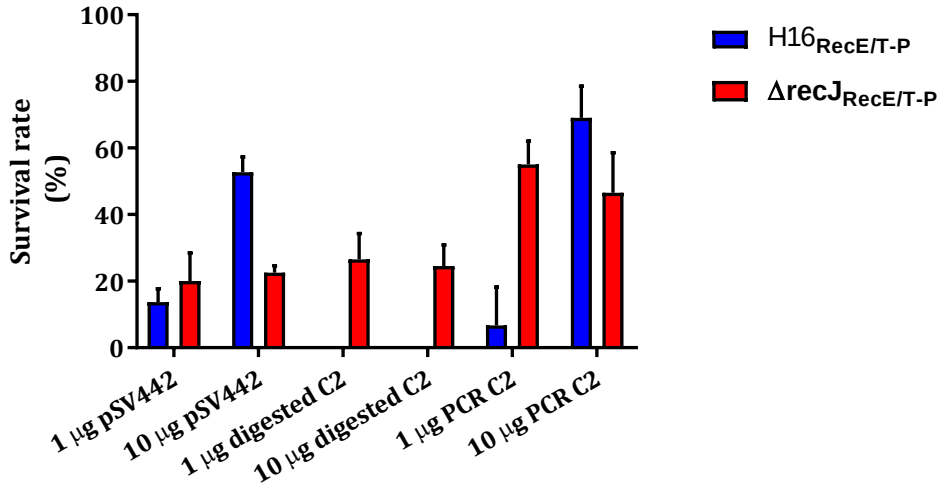


Figure 0.28 – Percentage of survival of C. necator H16$_{recET-P}$ (blue) ΔrecJ$_{recET-P}$ (red) after colony purification onto fresh LBCS agar plates.

cPCR was performed on 60 colonies from the re-streaked LBCS agar plates, utilising the PCR primers *PhaC*_OUT F/R (**Figure 3.29**). An expected product size of 3.2 kb was not detected, meaning integration of the cassette in the chromosome had not been successful. This was consistent with the other approaches already utilised in this chapter.

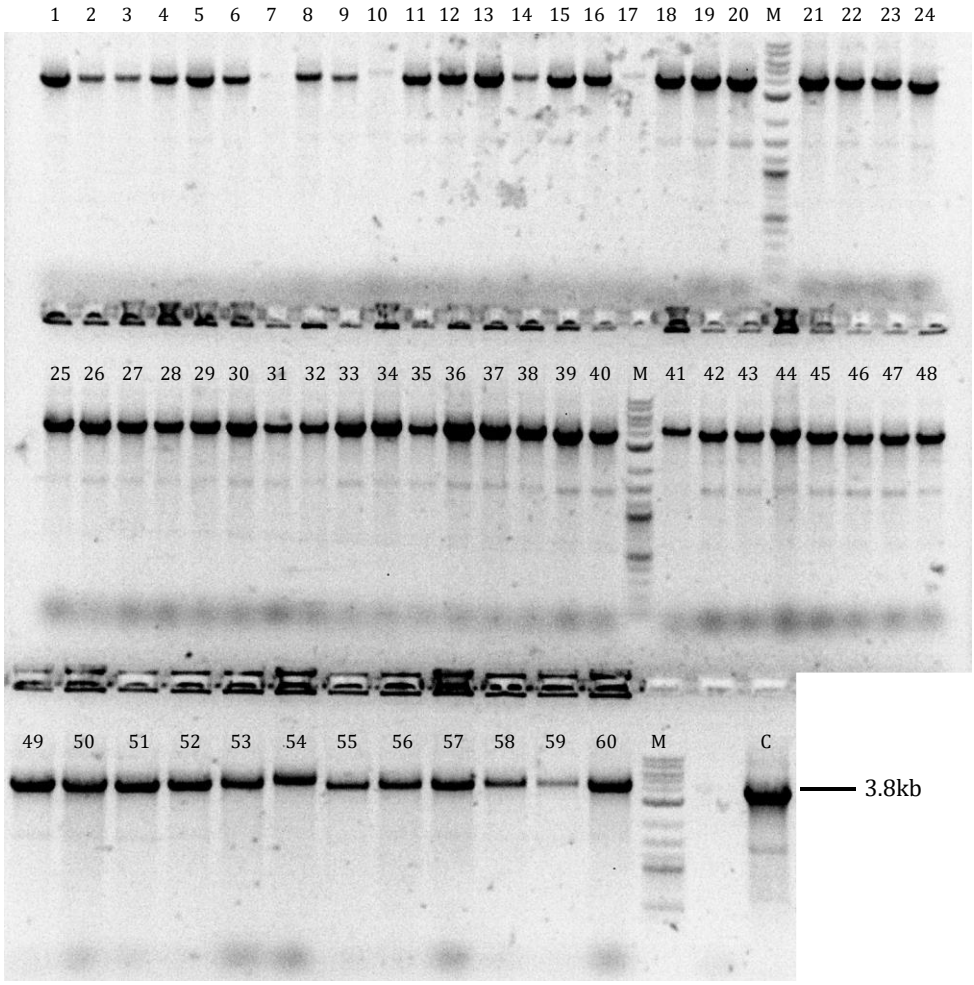


Figure 0.29 – Screening of Δ$recJ_{recET\text{-}P}$ *purified transformants via cPCR utilising PhaC_OUT.F/R primer pair.*

Lane 1: H16$_{recET\text{-}P}$ cells transformed with 1 µg pSV442; 2-8: H16$_{recET\text{-}P}$ cells transformed with 10 µg pSV442; 9: H16$_{recET\text{-}P}$ cells transformed with 1 µg PCR cassette C2; 10-16: H16$_{recET\text{-}P}$ cells transformed with 10 µg PCR cassette C2; 17-24: Δ*recJ*$_{recET\text{-}P}$ cells transformed with 1 µg pSV442; 25-32: Δ*recJ*$_{recET\text{-}P}$ cells transformed with 10 µg pSV442; 33-38: Δ*recJ*$_{recET\text{-}P}$ cells transformed with 1 µg digested cassette C2; 39-46: Δ*recJ*$_{recET\text{-}P}$ cells transformed with 10 µg digested cassette C2; 47-54: Δ*recJ*$_{recET\text{-}P}$ cells transformed with 1 µg PCR cassette C2; 55-60: Δ*recJ*$_{recET\text{-}P}$ cells transformed with 10 µg PCR cassette C2; C: PCR control on 100 ng H16 total DNA; M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb).

3.6 Discussion

In the *Cupriavidus* genus, gene knockouts are generated by allelic replacement utilising conjugation via spot mating and counterselection with the *sacB* gene on sucrose plates (Quandt and Hynes, 1993; Lenz *et al.*, 1994) or using the Cre/Lox system (Grosse, Friedrich and Nies, 2007). These techniques have been used to generate multiple mutant strains (Julian *et al.*, 2009; Volodina, Raberg and Steinbüchel, 2016) and optimised over time (Lenz and Friedrich, 1998; Mifune, Nakamura and Fukui, 2010). However, they remain fastidious, time-consuming and in some cases, inefficient (Chen *et al.*, 2015).

It was therefore essential to establish a HTP technique capable of generating rapid gene mutations. The λ-Red system has revolutionised genetic editing in *E. coli* (Datsenko and Wanner, 2000; Murphy, Campellone and Poteete, 2000; Murphy, 2016) and the establishment of the MAGE technique (Wang *et al.*, 2009) represented the climax of HTP tool development for the re-programming of genetic networks in *E. coli*.

The λ-Red system was assessed in *C. necator* but proved inefficient to promote gene replacement in the strain. Despite several design optimisations (quantity of donor DNA electroporated, length and position of homologous arms, target gene), the recombineering system could not generate mutants. This highlights the limited success of λ-Red outside of *E. coli*.

Indeed, λ-Red recombination in *P. aeruginosa* requires 8 µg of DNA and 400-600 bp homology arms (Lesic and Rahme, 2008) while *E. coli* recombinants are easily generated from 10-100 ng linear fragments harbouring as little as 35 bp flanking regions (Datsenko, 2000). The λ-Red method has been applied in diverse microorganisms such as *Yersinia*, *Shigella* or *Agrobacterium* species (Derbise *et al.*, 2003; Hu *et al.*, 2003,

2014). However, recombination rates were on average 10-120 recombinants/µg of DNA whilst recombination in *E. coli* reached 10^8 recombinants/µg of DNA (Yu *et al.*, 2000). Successful gene replacement was largely target-dependent (Derbise *et al.*, 2003) and large amounts of DNA were required to obtain high-efficiency (Lesic and Rahme, 2008). Consequently, the λ-Red technique was not systematically established as a recombineering standard in chassis outside *E. coli* or closely related enterobacterial species like *Salmonella* (Chakravortty, Hansen-Wester and Hensel, 2002; Lu *et al.*, 2002) .

Nonetheless, extensive studies of λ-Red adaptation in *E. coli* and non-related bacteria provided a range of improvements that could have been conducted in *C. necator*. Variations in homology length, quantity of donor DNA and gene target are the most common parameters known to improve recombineering efficiencies (Derbise *et al.*, 2003; Lesic and Rahme, 2008) and they all have been evaluated in the present study. The linear invasive DNA, a key element in recombination, could be further optimised via addition of protective 5' phosphorothioate bonds (Hossain *et al.*, 2015) when generated by PCR. In addition, the deletion target could have been shortened, as small chromosomal rearrangements are more efficient than large rearrangements (Wang *et al.*, 2009). Finally, it would have been interesting to transform single-stranded oligonucleotides. Indeed, the λ-Red system would be limited to the Beta protein minimising the size of the helper plasmid. The process could be faster and easily scaled-up to a high-throughput level, like in the MAGE technique (Wang *et al.*, 2009).

The strong RM protective system of *C. necator* might also contribute to the poor recombination efficiencies obtained with the λ system. The deletion of *recJ* did improve transformation efficiencies so the deletion of multiple genes coding restrictive enzymes (*xseAB*, *sceCD*, *addAB*) could have contributed to a λ-Red-generated mutant strain (Mosberg, Lajoie and Church, 2010). Nonetheless, the knock-out of four exonucleases (*RecJ*,

ExoVII, ExoI and ExoX) activated the mismatch repair system in *E. coli*, which generated a low mutability in deficient strains (Burdett *et al.*, 2001).

This supports the assumption that the genetic manipulation of environmental or pathogenic wild type isolates cannot be treated in the same manner as domesticated laboratory species. λ-Red methodological adaptations are often necessary in these strains to obtain satisfying recombination efficiencies. Alternatively, genome manipulations could be tackled with methods different than Lambda-Red, starting with the universal CRISPR/Cas9 technique.

Ultimately, a strain-specific phage-derived recombineering system would be required to develop a high-throughput recombineering system replacement in non - *E. coli* hosts. On this topic, nine putative recombinase-exonuclease pairs were compared for oligonucleotide recombination in *E. coli* (Datta *et al.*, 2008). Those enzymes were selected for their sequence analogy to Beta/Exo and RecET proteins and isolated from phages and prophages of Gram-positive and –negative strains such as *Bacillus subtilis*, *Staphylococcus aureus* or *Vibrio cholera*. Unsurprisingly the pairs coming from *E. coli* related species provided higher recombination rates than those of distant hosts.

Similarly, Swingle *et al* identified enzymes encoded by the genome of *Pseudomonas syringae* pv. *syringae* B728a, with functions similar to the λ Beta/Exo proteins and showed efficient dsDNA recombination in the related *P. syringae* pv. tomato DC3000 (Swingle *et al.*, 2010). An identical approach was conducted in this study and a recombinase/exonuclease pair was identified among the proteins encoded by the genome of *C. gilardii*. These genes were cloned, expressed on a low-copy backbone and tested for dsDNA recombination. Unfortunately, it did not lead to the generation of the desired deletion mutant.

Consecutively, the *recET* system isolated from *Pseudomonas syringae* pv. *syringae* B728a was tested for deletion of the *phaC* gene in H16. All transformants were Spec-positive after selection on agar plates and screening of the cassette via cPCR. Four of these clones showed a 2.5 kb PCR product with external primers instead of the 3.8 kb wild type product. However, internal primers still amplified by PCR the wild type 500 bp band, suggesting the production of a mixed population of wild-type and mutant strains. A further analysis of these potential mutants could have allowed the elucidation of mechanisms and regulation underlying homologous recombination in *C. necator*.

Deletion of the mismatch repair system (Δ*mmr*) did not allow the generation of recombinants while in the *RecJ*-deficient strain Δ*recJ* an increase in transformation-recombination efficiency was observed. This supports the assumption that native exonucleases participate in an intricate recombination-regulating network which has not been elucidated yet. At this stage, a better understanding of molecular mechanisms is necessary to develop a rational and efficient HR system in *C. necator*. However, the purpose of the present study was to develop high-throughput tools, which could be both readily transferable between closely related strains and could be utilised to engineer improved ethylene production in *Cupriavidus*. Therefore, more generic and reliable methods were evaluated, starting with the highly efficient CRISPR/Cas9 system.

Development of a CRISPR/Cas9-based genome editing tool in *C. necator*

4.1 Introduction

As reported in Chapter 3, the bottleneck for any editing method to be efficient in *C. necator*, is the activation of native recombination pathways. A single, unrepaired double strand break (DSB) may be lethal for the cell (Kuzminov, 1999). To negate this , the cell must deploy a strategy to re-ligate both DNA ends usually via homologous recombination (Hiom, 2009).

CRISPR/Cas9 is a versatile, efficient and simple technique, which relies on the endonuclease Cas9, generating DSBs in the genome through directed cleavage (Jinek *et al.*, 2012). Therefore, the CRISPR/Cas9 system was implemented in *C. necator* to generate localised DSBs and promote homologous recombination in the presence of undamaged homologous DNA. Once established in *C. necator*, it could enable the rapid HTP genomic editing of the strain.

The CRISPR/Cas9 system relies on two components: the dsDNA endonuclease, Cas9, and a single guide RNA (sgRNA), of approximately 90 nucleotides, which combines the programmable guide RNA (gRNA) and an RNA scaffold allowing for the RNA molecule to complex with the Cas9 protein (Hsu, Lander and Zhang, 2014). Bioinformatic analysis on the sequence of the target gene allows the identification of the protospacer adjacent motif (PAM) sequence (5'-NGG-3', the archetypal sequence being from the *Streptococcus pyogenes* Cas9) and the 20 bp sequence directly upstream of PAM, called protospacer, which guides the Cas9:sgRNA complex (Anders *et al.*, 2014). The gRNA binds complementary to the strand opposite the protospacer, and the two-nuclease domains of Cas9, RuvC and HNH, cleave the dsDNA. This usually occurs 3 bp upstream of the PAM region in the seed sequence (Sternberg *et al.*, 2015) (8bp at the 3' end of the guide (Gorski, Vogel and Doudna, 2017)), generating a DSB (**Figure 4.1**). Then, either the Non-Homologous End Joining (NHEJ) or the Homology-Directed Repair (HDR) pathways fix the genomic lesion. The NHEJ pathway

catalyses the resection and ligation of both ends (Della *et al.*, 2004) via the combined action of the recombinase Ku (H16_B2355) and the multi-functional nuclease/polymerase/ligase LigD (H16_B2352) (Pohlmann *et al.*, 2006). This error-prone repair pathway leaves a scar in the form of indel (insertion/deletion) mutations (Della *et al.*, 2004; Ran *et al.*, 2013). The HDR pathway allows for high-fidelity and precise editing, utilising a repair template, either in the form of a plasmid or a linear dsDNA cassette (Ran *et al.*, 2013).

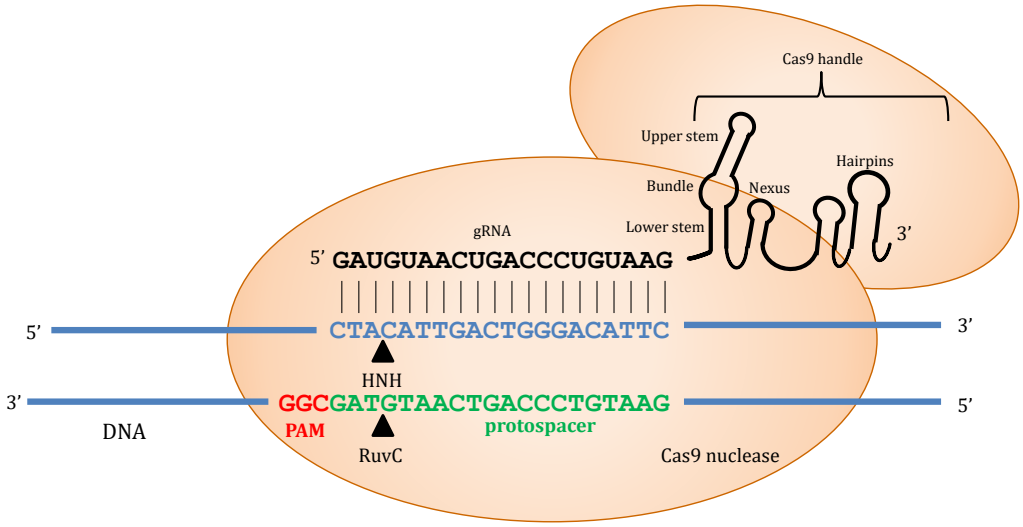


Figure 0.1 – Molecular mechanism of Cas9 double-stranded DNA cleavage

The Cas9 nuclease creates a DSB 3 bp upstream of the PAM sequence upon recognition of the PAM motif and complementary binding of the gRNA to the sequence opposite to the protospacer. Cas9 contains the RuvC and the HNH nuclease domains (arrowheads). Adapted from Stelle *et al*, 2015 and Yao *et al*, 2018.

Chapter 4 presents the assessment of the *S. pyogenes* CRISPR/Cas9 system for plasmid and chromosome editing in *C. necator*. The system proved successful when cleaving plasmids, but it was not successful in editing chromosomal genes. The CRISPR/Cas9 system was then combined with homologous recombination as detailed in **Chapter 3,** to generate the CRISPR-MAGE (CRMAGE) tool. Despite many improvements, the use of CRISPR/Cas9 in *C. necator* initially remained unsuccessful. Finally, the

chapter ends with some preliminary results demonstrating that an improved version of the CRISPR/Cas9 system can be utilised in *C. necator*.

4.2 Establishment of a CRISPR/Cas9-based genome editing system in *C. necator*

4.2.1 Construction of a functional CRISPR/Cas9 vector for C. necator

The pressing need to generate effective HTP genome editing tools for *C. necator* led to the development of a two-plasmid system. The first plasmid expressed the Cas9 nuclease and the second plasmid would transcribe the sgRNA, comprising the gRNA and the RNA scaffold. The second plasmid could be re-utilised to introduce different targeting gRNA sequences. To minimize the potential toxicity of the CRISPR system, a low-copy number shuttle vector (built with the p15A-pVS1 origins of replication) was selected (Heeb *et al.*, 2000) and the native *S. pyogenes cas9 (Spycas9)* gene was placed under the control of the arabinose-inducible promoter pBAD, to generate the plasmid pSV500 (**Figure 4.2A**). The sgRNA was constitutively expressed on two different vectors: pTarget and pTarget_HA (**Figure 4.2B/C**). The first plasmid expresses only the sgRNA, whilst the second plasmid also harbours homologous arms, which are complimentary to the flanking sequence of the target gene. The pTarget vectors could be redesigned further to allow plasmid curing. For example, a second gRNA targeting the gRNA plasmid could be transiently transcribed under an inducible promoter (Cobb, Wang and Zhao, 2015). Alternatively, an I-SceI cutting site (rare and not usually present in bacterial genomes (Jasin, 1996)) could be introduced onto the plasmid backbone, the *cas9* vector pSV500 already having a copy of the *sceI* gene, which could be expressed to cure the pTarget plasmid once the genome editing has been completed.

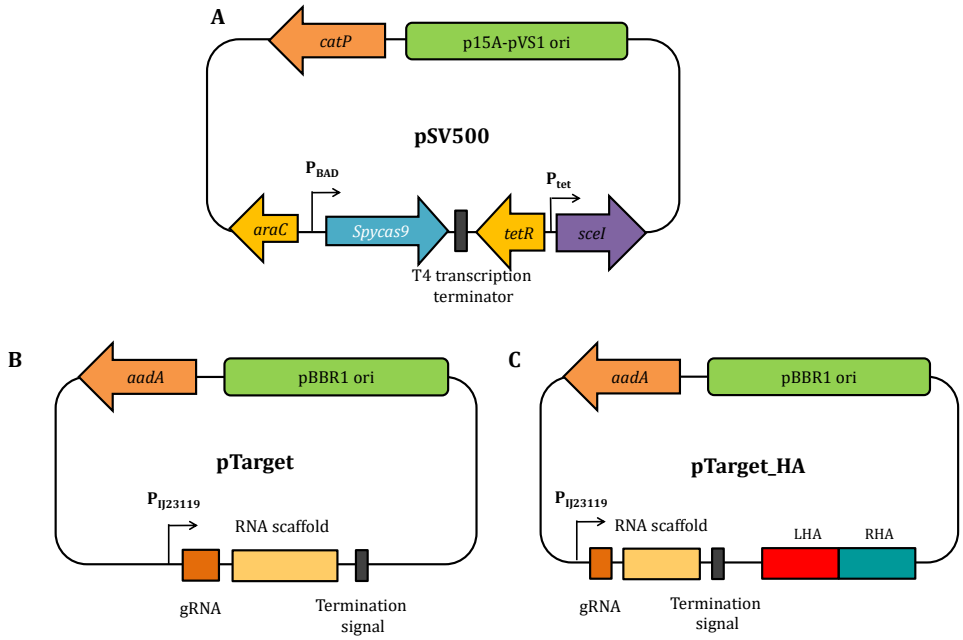


Figure 0.2 – Schematic representation of the three different plasmids used for CRISPR/Cas9 editing in C. necator

A. The low-copy number pSV500 plasmid expressing the arabinose inducible Cas9 nuclease from *S. pyogenes* and the anhydrotetracycline-inducible meganuclease, I-SceI.
B. The medium-copy number plasmid, pTarget. This plasmid series contains the sgRNA, under the control of a constitutive promoter. Each sgRNA was designed to target either an exogenous plasmid or a chromosomal gene. Plasmid design was based on the pTarget series used in *Streptomyces* (Y. Jiang *et al.* 2015).
C. The sgRNA plasmid series, pTarget, was modified by the addition of homology arms, these were utilised as donor DNA for the HDR-mediated precise genome editing.
catP: chloramphenicol acetyl-transferase; *araC*: arabinose operon regulatory protein; P_{BAD}: arabinose promoter in *E. coli* (Schleif, 2010); tetR: tetracycline repressor;P_{tet}: promoter for the bacterial *tetR gene*; *sceI*: endonuclease I-SceI; *aadA*: streptomycin 3''-adenylyltransferase; $P_{ij23119}$: strong synthetic promoter used in *E. coli*; LHA: left homology arm; RHA right homology arm.

Based on the work of Jiang *et al,* the gRNA was transcribed utilising the constitutive promoter pIJ23119 in both the pTarget and pTarget_HA plasmids (Jiang *et al.*, 2015). The promoter pIJ23119 was designed and constructed in 2006, by the iGEM team at Berkeley, as part of a mini synthetic promoter library (http://parts.igem.org/Part:BBa_J23119). The promoter pIJ23119 is currently the strongest reported sigma 70 consensus promoter utilised in *E. coli* (Yan and Fong, 2017). However, the activity of this promoter needed to be evaluated in *C. necator*.

The plasmid pBBRR33, kindly provided by Dr Samantha Bryan, expressed the enhanced Yellow Fluorescent Protein (eYFP), under the control of a rhamnose-inducible promoter, Indeed, pRhaB activity and eYFP fluorescence have recently been demonstrated in *C. necator* (Pfeiffer and Jendrossek, 2012; Sydow *et al.*, 2017). The vector pIJ23119_EYFP was built using the plasmid pBBRR33 as a template; the *rhaR-rhaS*-P_{RhaB} fragment was excised from the plasmid and replaced with the constitutive pIJ23119 promoter pRhaB (**Figure 4.3A-B**). The pBBRR33 plasmid without the rhamnose promoter was also utilised as a control to assess background fluorescence (p_EYFP) (**Figure 4.3C**).

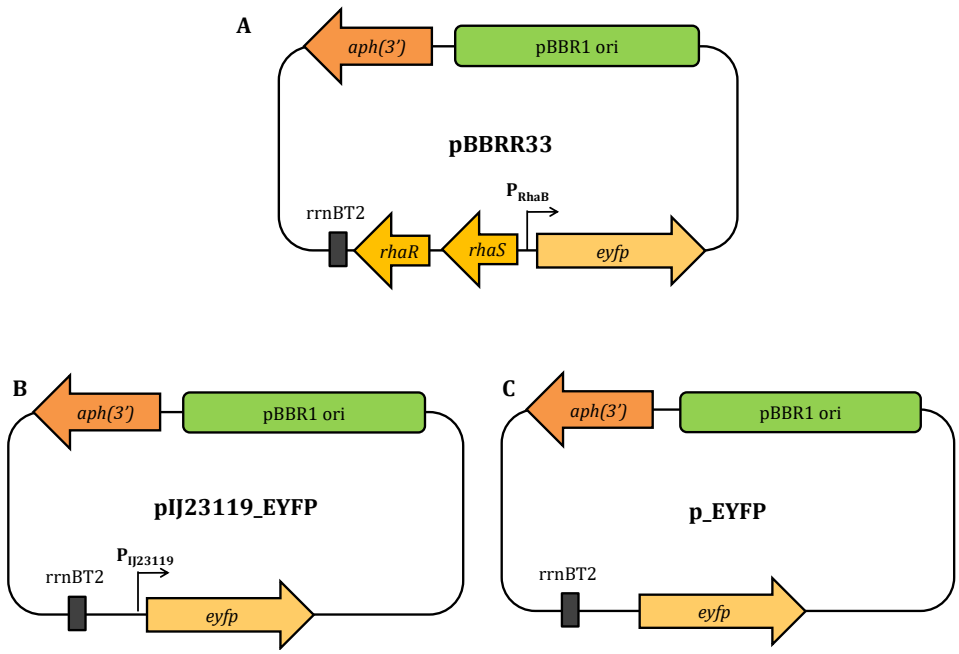


Figure 0.3 – Schematic representation of eYFP-expressing vectors for measurement of pIJ23119 activity

A. pBBRR33 is a derivative of the pBBR1 vector, expressing *eyfp* under the tight control of the rhamnose promoter P_{rhaB} from the *E. coli rhaBAD* operon (Giacalone *et al.*, 2006).
B. The *rhaR-rhaS*-P_{RhaB} inducible system was excised from the plasmid pBBRR33 and replaced with the synthetic constitutive promoter $P_{IJ23119}$ to generate the plasmid pIJ23119_EYFP. EYFP was used as a reporter to assess the activity of the of $P_{IJ23119}$ promoter.
C. The promoter sequence upstream of the *eyfp* gene in pBBRR33 was excised from the plasmid allowing transcriptional activity from the backbone of pBBRR33 to be assessed.

aph(3'): aminoglycoside phosphotransferase from Tn*5* conferring resistance to kanamycin; *rrnBT2*: transcription terminator T2 from the gene *rrnB* in *E. coli* ; *rhaR*: transcriptional activator of *rhaR* and *rhaS*; *rhaS*: positive regulator of the *rhaB* promoter; *eyfp*: enhanced yellow fluorescent protein; P_{rhaB}: rhamnose-inducible promoter.

The rhamnose–inducible promoter tightly regulated the expression of *eyfp* in *E. coli,* as no fluorescence signal was detected in the non-induced sample (**Figure 4.4A**). After an initial lag in expression of approximately 50 min (half the experimental time), cells induced with rhamnose reached $5x10^3$ arbitrary units (A.U.), while fluorescence in DH5α/pIJ23119_EYFP fluctuated around 10^4 A.U., confirming the strength of the promoter. A similar result was observed in *C. necator* H16 **(Figure 4.4B).** However, a higher fluorescence output of up to $2x10^4$ A.U was noted with pBBRR33 in the presence of rhamnose. Negligible *eyfp* expression was reported in the p_EYFP control (data not shown). The reporter gene expression in H16/pIJ23119_EYFP culminated in a fluorescence level of $1x10^5$ A.U. Thus, the promoter pIJ23119 is functional in *C. necator* and mediates strong constitutive expression of the *eyfp* gene. The promoter pIJ23119 was therefore used to drive the transcription of the sgRNA in the subsequent CRISPR experiments.

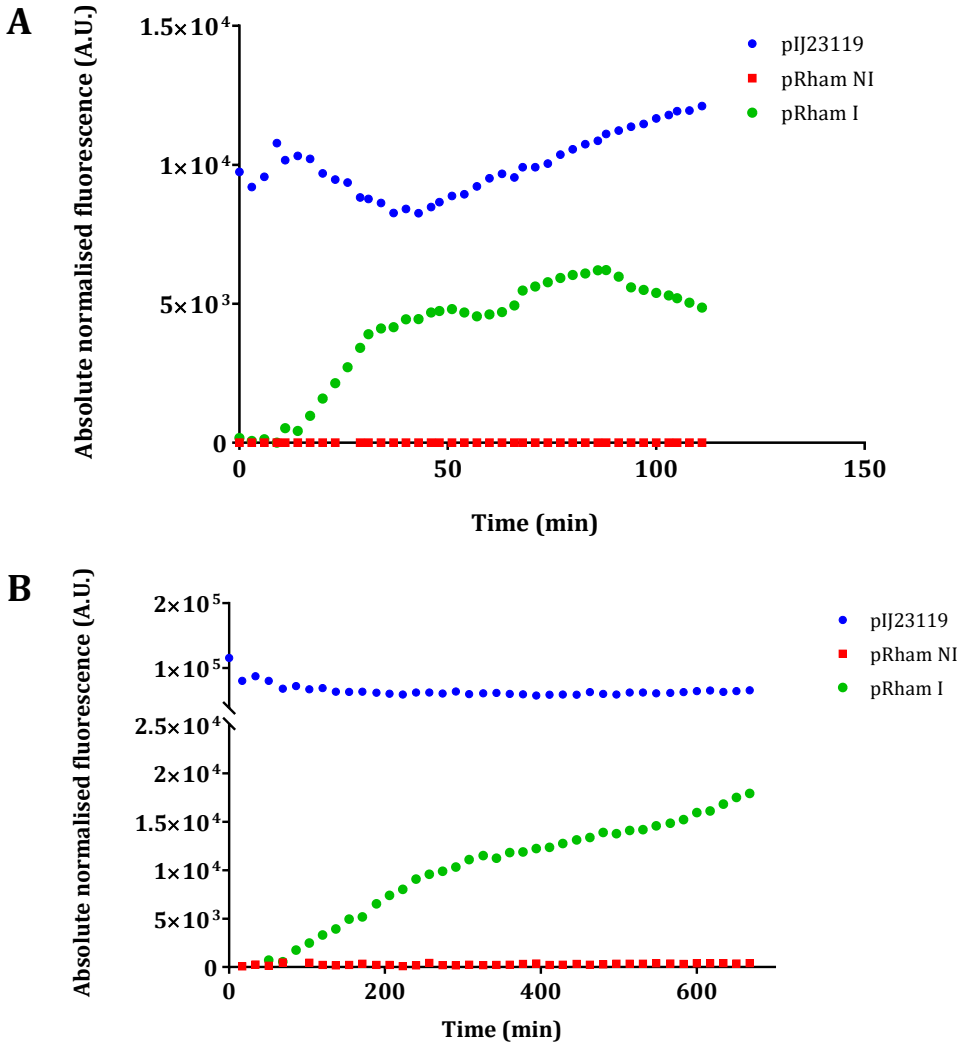


Figure 0.4 – Activity of the promoter pIJ23119 in E. coli and C. necator

Arbitrary normalised fluorescence in (**A**) *E. coli* DH5α and (**B**) *C. necator* H16 harbouring either the plasmid pIJ23119_EYFP (blue) or pBBRR33, the pBBRR33 plasmid was induced (green) with 0.2% rhamnose, non-induced cells were used a control (red). Time t=0 corresponds to rhamnose induction. Error bars are too small to be visible. The fluorescence activity from p_EYFP is not shown, as it was found to be negligible and the curves merged with the X-axis.

4.2.2 CRISPR/Cas9-assisted plasmid cleavage in C. necator

The CRISPR ribonucleoprotein complex was assessed for functionality in *C. necator,* i.e. whether (i) the sgRNA assembled correctly with the nuclease Cas9, (ii) the complex recognised the region complementary to the gRNA and (iii) Cas9 cleaved the target DNA to

generate a DSB. The two-plasmid system described above was utilised with a gRNA directed against the *cas9*-expressing plasmid pSV500. A gRNA was designed to target the gene coding for the stability protein StaA, in the pVS1 replicon. The gRNA was amplified and cloned into the plasmid pTarget to create pSV650 (**Figure 4.5**). Efficient cleavage by the CRISPR complex should destabilise the replication of pSV500, thus curing the plasmid, leading to loss of chloramphenicol resistance and cell death on selective plates.

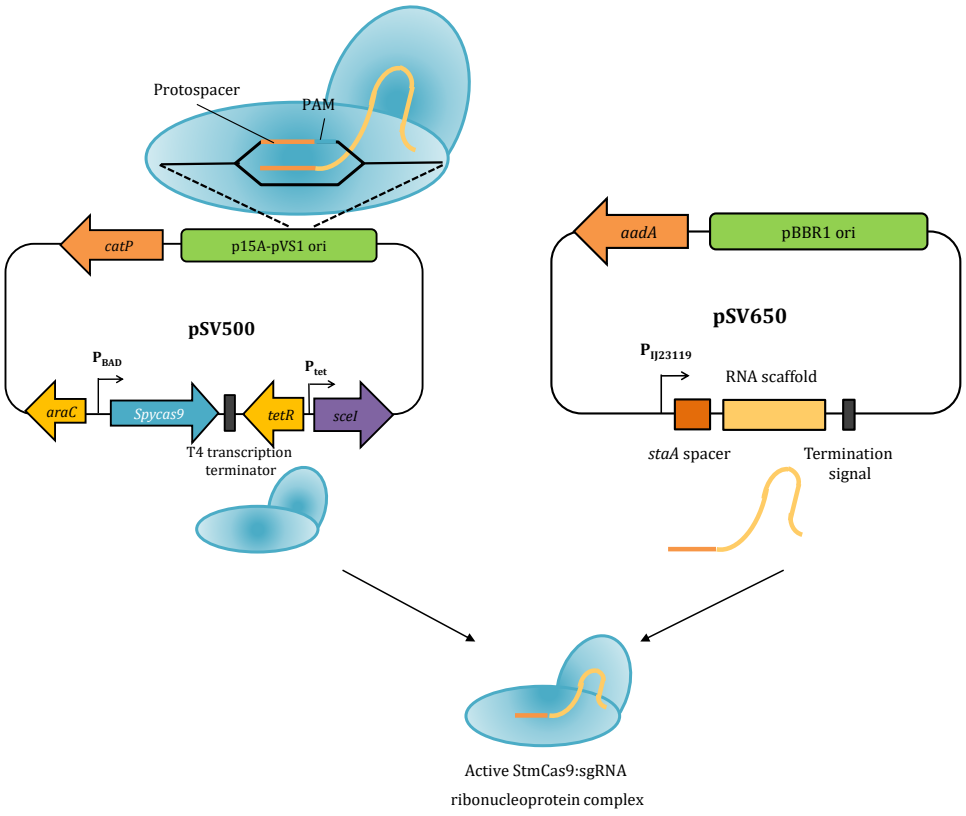


Figure 0.5 – Overview of the CRISPR/Cas9 two-plasmid testing system

The CRISPR/Cas9 ribonucleoprotein complex can only be assembled if both plasmids are maintained, pSV500 with an arabinose inducible Cas9 and pSV650 containing the sgRNA. Once the complex is formed, it scans pSV500 and detects the protospacer region of *staA,* which matches the gRNA sequence. If the protospacer is followed at its 3'end by a PAM motif, Cas9 cleaves the DNA, introducing a lethal DSB.

aadA: streptomycin 3''-adenylyltransferase; *catP*: chloramphenicol acetyltransferase; *araC*: L-arabinose regulatory protein; P_{BAD}: promoter of the L-arabinose operon of *E. coli* (Guzman *et al.*, 1995); *Spycas9*; Cas9 endonuclease (*S. pyogenes*); *tetR*: tetracycline repressor; P*tet*: promoter for bacterial *tetR; sceI*: I-SceI endonuclease *(S. cerevisiae);* $P_{ij23119}$: strong synthetic promoter for *E. coli*.

C. necator cells carrying the Cas9 vector pSV500 ($H16_{Cas9}$) were made competent in both the presence and absence of arabinose, to induce *cas9* expression. Both induced and non-induced cells were transformed with water (negative control) or 100 ng of the plasmid pSV650, incubated for recovery with or without arabinose and spread onto selective plates containing: 50 μg/mL chloramphenicol (Cm 50), 300 μg/mL spectinomycin (Spec 300) and the combination of both antibiotics (Cm + Spec). The experiment was repeated in triplicate.

The *S. pyogenes* CRISPR/Cas9 system functioned as expected in *C. necator* H16. Increased mortality was observed on the plates from the cells expressing Cas9 compared to the non-induced control (**Figure 4.6A**), demonstrating that the CRISPR system can be utilised effectively in *C. necator* to cleave plasmids.

The transformation frequency (TF) was calculated for each strain containing both plasmids, by dividing the CFUs obtained on the Cm + Spec plates by the CFUs on the Cm plates, this figure was then normalised to 10^9 cells. The TF of $H16_{Cas9}$/pSV650 following arabinose induction (purple bar in **Figure 4.6B**) was $8.4x10^7$ arbitrary units (A.U.), an order of magnitude lower than in the non-induced transformed cells ($7.3x10^8$ A.U.). In addition, the survival rate was 88% lower in induced cells than in non-induced cells, demonstrating the efficiency of the CRISPR complex in plasmid curing in *C. necator*. However, successful cleavage of the plasmid only occurred in 88% of the cells (**Figure 4.6C**), as some colonies could still grow on Cm + Spec plates.

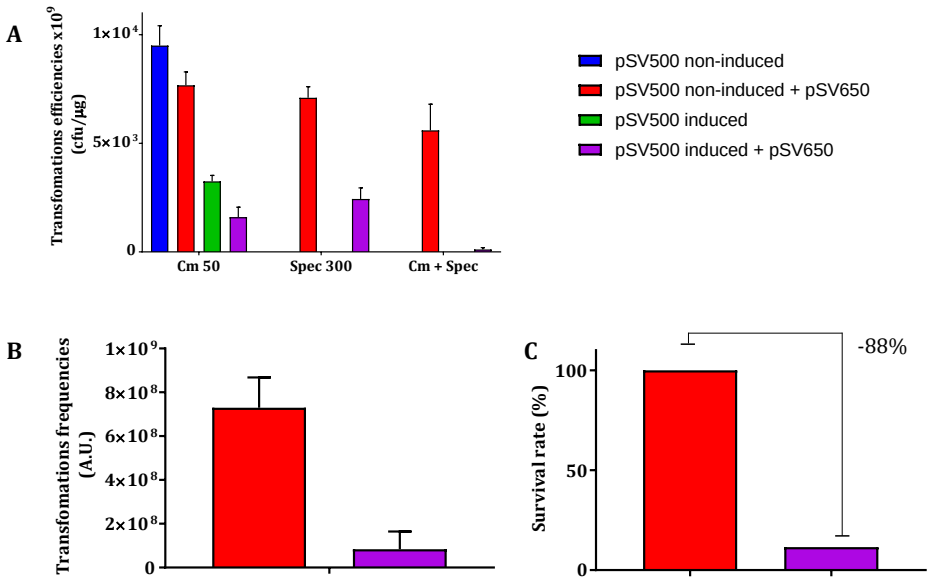


Figure 0.6 – Cleavage of the Cas9-expressing plasmid pSV500 by the CRISPR/Cas9 two-plasmid system in C. necator

A. Transformation efficiencies (cfu/µg) of H16$_{Cas9}$ with pSV650.
H16$_{Cas9}$ cells were transformed with either water (blue and green) or 100 ng of the plasmid pSV650, with the sgRNA targeting *staA* on pSV500 (red and purple). Two subgroups of cultures (green and purple) were constantly induced with 10 mM arabinose, from induction through to recovery. However, no arabinose was added to the selective agar plates, which were only supplemented with antibiotics: 50 µg/mL chloramphenicol (Cm 50), 300 µg/mL spectinomycin (Spec 300) or the combination of both (Cm + Spec).
B. Transformation frequencies (TF) of each strain transformed with pSV650. TF were calculated by dividing the CFUs obtained in Cm + Spec by the CFUs on the Cm plates, the figure was then normalised to 10^9 cells: TF = $(Cm^R{+}Spec^R\ colonies)/\ (Cm^R\ colonies)$ x10^9.
C. The transformation frequency of the non-induced H16$_{Cas9}$/pSV650 strain was set to 100% survival and survival rates of the induced population were calculated accordingly.

To investigate the fate of these CRISPR escapers, sixteen colonies were selected from the induced H16$_{Cas9}$/pSV650 plates and re-streaked onto fresh LB Cm + Spec plates and LB Cm 50 agar plates. Eight colonies were tested from the LB Cm + Spec plates and five colonies were tested from the LB Cm 50 plates. Selected colonies were tested via PCR for the pSV500 plasmid using the primers I-SceI.F/R, which amplify a 1.4 kb region of the plasmid (**Figure 4.7A**). To confirm that plasmid curing was associated with the activity of a fully assembled CRISPR system, i.e. that the sgRNA plasmid had been conserved in the selected recombinants, a second PCR screening

was performed with the primer pair pBBR1.F/R, which amplify the replicon of pSV650.

The plasmid pSV650 was detected in all the colonies re-streaked on Cm + Spec, demonstrating that it had been conserved through the editing process. The pSV500 plasmid was successfully cured in these strains by a fully assembled CRISPR-Cas9 system. The escapees were therefore likely to be spontaneous chloramphenicol resistant strains. To verify this hypothesis, a region of the Cas9 plasmid pSV500 was also amplified via PCR from colonies re-streaked onto LB Cm 50 plates only. All the selected colonies had a band of 1.4 kb, specific to the plasmid (**Figure 4.7B**).

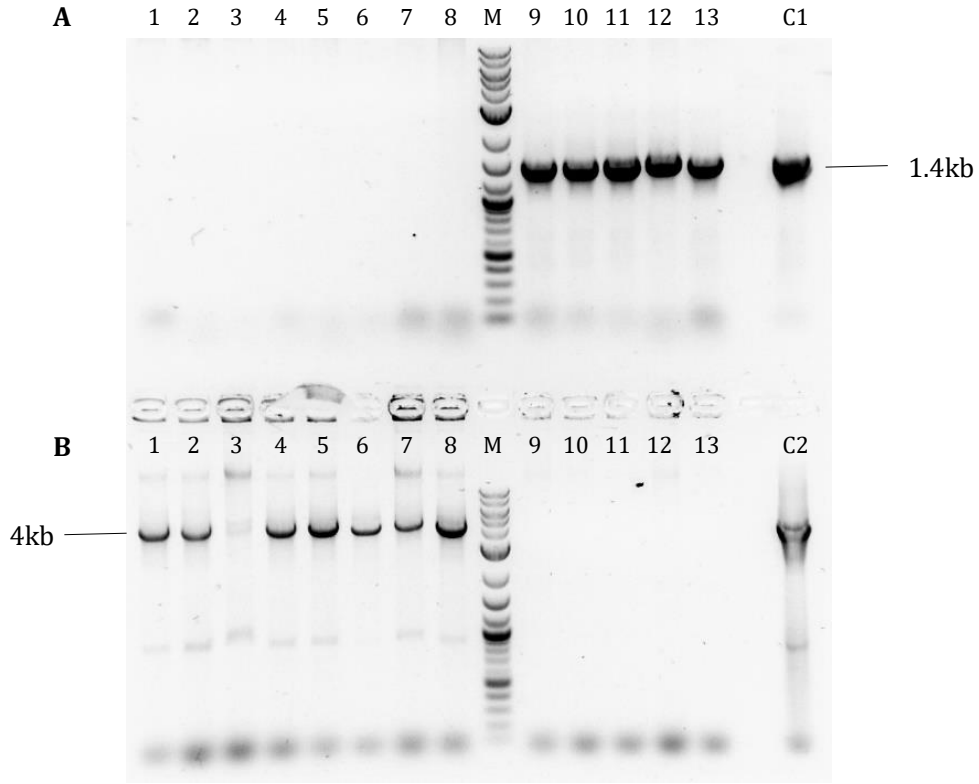


Figure 0.7 – Screening via cPCR of H16$_{Cas9}$ transformants re-streaked on LB Cm50 and LB Cm + Spec agar plates.

A. Amplification of the I-SceI-expressing region of pSV500 with primers I-SceI.F/R.

B. Amplification of the pBBR1 replicon of pSV650 with the primers pBBR1.F/R. 1-8: Clones from re-streaks on LB Cm + Spec; C1: PCR control on 100 ng plasmid pSV500; 9-13: Clones from re-streaks on LB Cm, C2: PCR control on 100 ng plasmid pSV650; M: 2- NEB® 2-Log DNA ladder (0.1 – 10.0 kb).

It can therefore be concluded that the vector pSV500 was still in the cells after the first passage on Cm + Spec plates and was maintained in the cells re-streaked onto the Cm only plates. The same colonies had lost the plasmid pSV650, probably because of the absence of selection pressure i.e. Spec was not maintained on the Cm plates. These colonies therefore escaped through loss of the plasmid pSV650. However, the vector pSV500 was lost in all of the cells re-streaked onto Cm + Spec. Repeated patching of the colonies onto Cm + Spec plates significantly reduced the number of false positives. Therefore, the CRISPR/Cas9 system works in *C. necator* for plasmid cleavage, with an efficiency of 88% and re-streaking the mutants minimises cPCR screening.

4.2.3 CRISPR/Cas9-assisted chromosome editing in C. necator

To test the efficiency of the CRISPR/Cas9 system on one of the chromosomes of *C. necator,* three gRNAs were identified for both the sense and antisense DNA strands in the conserved *hsdR_N* domain of the *hsdR* gene (H16_A0006), already defined as a desirable target for assisted homologous recombination in **Chapter 3 (Figure 4.8A-B)**. The gRNAs were cloned in two versions of the gRNA-transcribing vectors: pTarget_*hsdR* and pTarget_HA_*hsdR* (**Figure 4.8C**). The first backbone contained the sgRNA only (gRNA and RNA scaffold) and was used to investigate whether NHEJ-based CRISPR editing could function in *C. necator*. The second plasmid contained two homology arms of 1 kb each, flanking the target gene, to promote HDR.

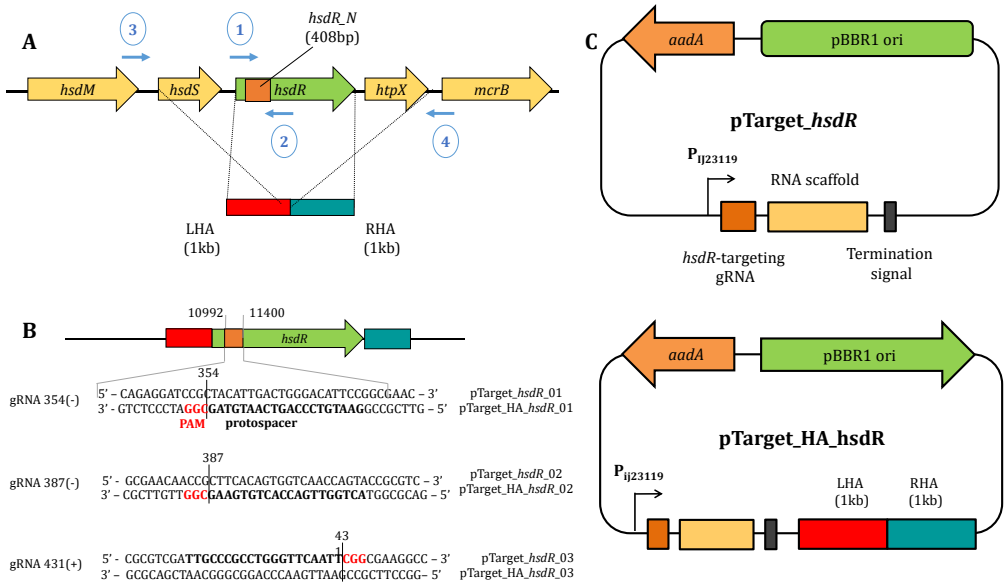


Figure 0.8 – Testing the efficiency of the CRISPR/Cas9 system in the genome of C. necator H16

A. Diagrammatic representation of the *hsdR_N* operon showing the region targeted by the homologous arms of the pTarget_HA_*hsdR* series.
B. The gRNA sequences, identical to the protospacer region of the chromosome (in bold) selected for knockouts of *hsdR_N* using CRISPR/Cas9 from *S. pyogenes*. Three guides were identified using the Benchling CRISPR feature (https://benchling.com/editor). Based on algorithms (Hsu, Lander and Zhang, 2014; Doench *et al.*, 2016), the software of the open-source platform identified and ranked every 20 bp sequence followed by a PAM sequence within the target gene. Two scores were provided for each gRNA candidate: the on-target score that reflected the cleavage efficiency related to the sequence itself and the off-target score which predicts mis-annealing in the genome. The higher the scores, the more efficient the cleavage by Cas9. In addition gRNAs located close to the 5' end of the coding domain sequence (CDS) were more likely to result in an early truncated protein, (Doench *et al.*, 2016). gRNAs satisfying this criterion were selected. The number and sign attributed to each gRNA identifies the nucleotide and strand of the CDS where Cas9 will cleave the DNA with nucleotide 1 being the first of the coding sequence.
C. gRNAs 354(-), 387(-) and 431(+) were cloned into the two gRNA-transcribing backbones, pTarget_*hsdR* and pTarget_HA_*hsdR* and validated by sequencing.
hsdM: Type I restriction-modification system methylation subunit, *hsdS*: Type I restriction-modification system specificity subunit, *hsdR*: Type I restriction-modification system restriction subunit, *htpX*: M48 family peptidase, *mcrB*: 5-methylcytosine-specific restriction enzyme B, LHA: left homology arm, RHA: right homology arm. Primer 1: *HsdR*_IN.F; primer 2: *HsdR*_IN.R; primer 3: S.ReuHI.F; primer 4: MDH.R. S.ReuHI.F and MDH.R primers amplify a product of 4 kb in wild type strains and 3 kb in mutant cells whereas the couple *HsdR*_IN.F/R would amplify a 200 bp fragment from wild type cells only.

H16$_{Cas9}$ cells were transformed with 100 ng of each plasmid pTarget_*hsdR, pTarget*_HA_*hsdR* and pSV650, which targets the Cas9 plasmid, pSV500. It was expected that Cas9 cleavage of the genome would stimulate homologous recombination. The 'repair or die 'situation proved

difficult and most of the cells did not recover from the lethal DSB. That is why reduced efficiencies were observed after transformation (**Figure 4.9**).

The number of cells that escaped plasmid cleavage was comparable to that obtained in the plasmid cleavage experiment (93x10^9 cfu/µg versus 113x10^9 cfu/µg) detailed in section *4.2.2*. This validates the robustness of the cleavage control, targeting the vector pSV500.

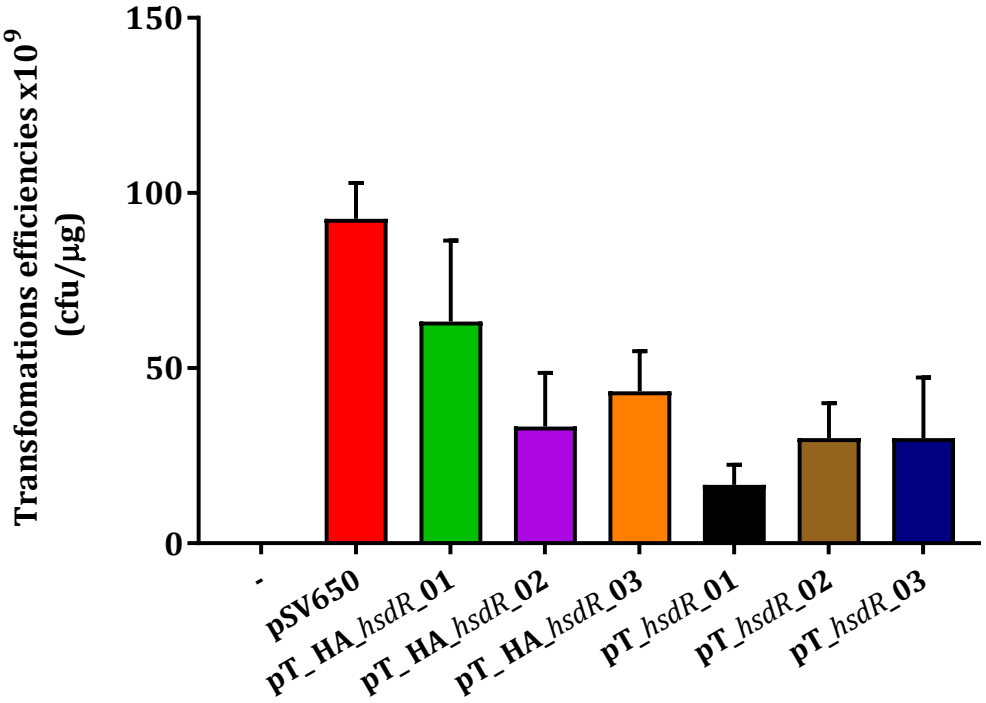


Figure 0.9 - Testing the efficiency of the CRISPR/Cas9 system in chromosomal deletion of hsdR_N

$H16_{Cas9}$ cells were transformed with either water (_), 100 ng of pSV650 targeting the *staA* gene on the pSV500 plasmid or one of the pTarget vectors targeting *hsdR_N*, pTarget (pT) and pTarget_HA (pT_HA).

Cells transformed with homologous DNA on the gRNA plasmid (pTarget_HA) recovered slightly better than those transformed with the pTarget vectors, which relied solely on the native NHEJ system. However, transformation efficiencies did not exceed 6.3x10^{10} cfu/µg, and a maximum of nine colonies were counted on the plates. These very small yields drastically limited the pool of potential Δ*hsdR* mutants. Transformants were re-streaked onto fresh Cm + Spec plates and subjected to cPCR.

The primers *HsdR*_IN.F/R were used to amplify a 500 bp band, including the targeted domain *hsdR_N* (408 bp) in all of the colonies tested (**Figure 4.10**). Therefore, no deletion mutants were detected.

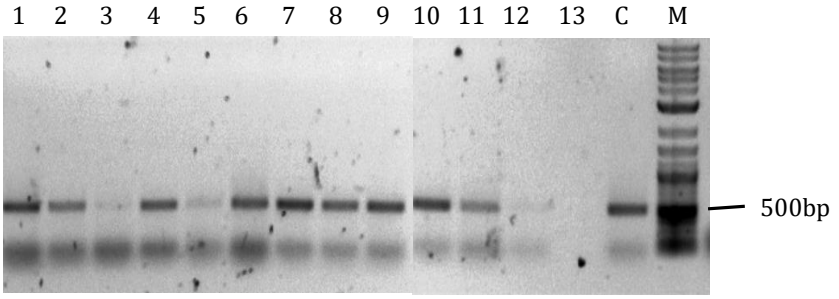


Figure 0.10 – PCR screening of the H16$_{Cas9}$ hsdR transformants re-streaked on Cm + Spec agar plates via cPCR utilising HsdR_IN.F/R primer pair.

Amplification of the wild-type sequence of the *hsdR_N* conserved domain with the primers *HsdR*_IN generated a band of 500 bp, while amplification from a mutant clone should generate a band of 100 bp. No mutants were detected.
1-9: H16$_{Cas9}$ cells transformed with 100n g pTarget_HA_*hsdR*_01; 10-12: H16$_{Cas9}$ cells transformed with 100 ng pTarget_HA_*hsdR*_02; 13: H16$_{Cas9}$ cells transformed with 100 ng pTarget_HA_*hsdR*_03; C: PCR control on 100 ng H16 genomic DNA; M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb).

The CRISPR/Cas9-based deletion of *hsdR* was attempted in parallel with the knock-out of *odhA* (H16_A2325), also known as *sucA in E. coli* (Hein and Steinbüchel, 1996), which codes for the E1 component of the 2-oxoglutarate dehydrogenase complex, along with the dihydrolipoamide succinyltransferase (E2) and the lipoamide dehydrogenase (E3). The complex catalyses the conversion of 2-oxoglutarate to succinyl-CoA and CO_2 and is a key step in the tricarboxylic acid (TCA) cycle.

The objective was to utilise CRISPR to improve 2-oxoglutarate availability in the cell allowing more flux towards the ethylene-forming enzyme, EFE, discussed in **Chapter 5**.

As with the deletion of the *hsdR_N* gene, H16$_{Cas9}$ was transformed with the plasmid series pTarget_*odhA* and pTarget_HA_*odhA*, targeting the 5' end of the 2-oxoglutarate dehydrogenase gene (**Figure 4.11**). However,

the homologous arms on the pTarget_HA_*odhA* plasmid series were designed to ensure a scarless in-frame deletion of *odhA*.

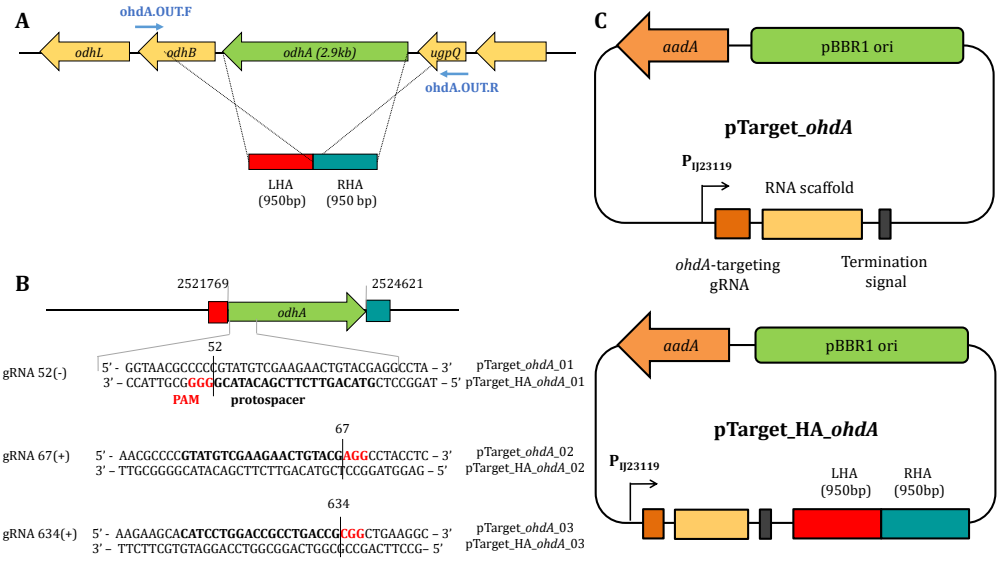


Figure 0.11 – Overview of the CRISPR/Cas9-based gene knockout of odhA in the genome of C. necator H16

A. Representation of the *odhA* gene and the genomic region targeted by the homologous arms of the pTarget_HA_*odhA* series;
B. Details of the gRNA sequences, regions identical to the protospacer region of the chromosome are shown in bold, gRNA sequences were designed *in silico* for knockout of the *odhA* gene, utilizing the CRISPR/Cas9 from *S. pyogenes*.
C. gRNAs 52 (-), 67(+) and 634(+) were inserted into the two gRNA-expressing backbones, pTarget_*odhA* and pTarget_HA_*odhA* and validated by sequencing.
odhL: dihydrolipoamide dehydrogenase (E3); *odhB*: dihydrolipoamide S-succinyltransferase (E2); *odhA*: 2-Oxoglutarate dehydrogenase (E1); *ugpQ*: glycerophosphoryl diester phosphodiesterase; H16_A2327 produces an ABC-type transporter. *odhA*.OUT.F/R primers amplify a product of 4 kb in wild type strains and 1.3 kb in mutant cells.

Like the transformation efficiencies obtained for *hsdR*, the transformation efficiencies obtained from the *odhA* knockouts ranged from between 30 and 150 x10^9 cfu/µg. However, a reduced number of cells were obtained following transformation with the pTarget_HA series than with the pTarget series (**Figure 4.12**), the opposite trend was observed with the *hsdR* transformants.

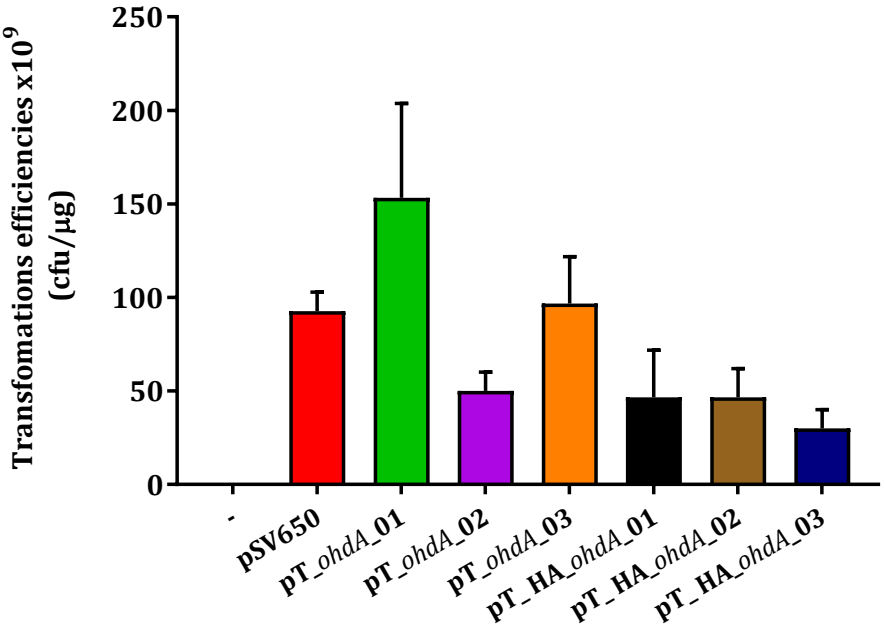


Figure 0.12 – Testing the efficiency of the CRISPR/Cas9 system in the deletion of the odhA gene (H16_A2325)

$H16_{Cas9}$ cells were transformed with either water (_), 100 ng of pSV650 or one of the pTarget vector series targeting *odhA* (H16_A2325). The pTarget (pT) series utilized the host NHEJ repair machinery while the pTarget_HA (pT_HA) vectors utilized sequences homologous to the flanking regions of the *odhA* gene, enabling precise HDR. Cells were left to recover for 3 h following transformation.

Colonies were screened via cPCR, utilising internal *odhA* primers, amplifying a small 800 bp fragment. In a small number of the colonies no band could be detected, suggesting that the gene had been successfully deleted (**Figure 4.13A**). A second PCR screen was performed on the same clones with external primers covering the full coding sequence of the gene, however the PCR profile was the same as the wild type, suggesting that the gene had not been deleted (**Figure 4.13B).** The initial PCR result could have been caused by insufficient cellular material or the presence of PCR inhibitors.

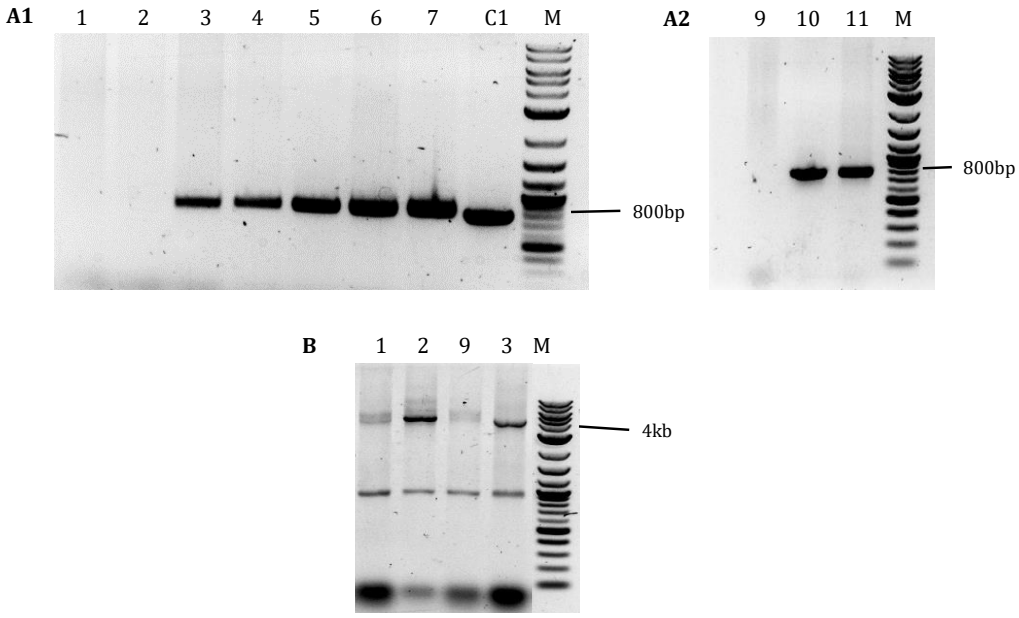


Figure 0.13 – Screening of H16$_{Cas9}$ odhA transformants re-streaked on Cm + Spec agar plates via cPCR utilising odhA_IN.F/R and odhA.OUT.F/R primer pairs.

A1 and A2. Screening of the *odhA* transformants with the *odhA*_IN.F/R primer pairs. 1-3: H16$_{Cas9}$ cells transformed with 100 ng pTarget_HA_*odhA*_01; 4-7: H16$_{Cas9}$ cells transformed with 100 ng pTarget_HA_*odhA*_02; 9-11: H16$_{Cas9}$ cells transformed with 100 ng pTarget_HA_*odhA*_03; C1: PCR control on 100 ng H16 genomic DNA; M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb). A band of 800 bp could be amplified from the wild type cells, C1. No band could be detected in 1, 2 and 9, the other colonies had a band of a similar size to the wild type.
B. Colonies 1, 2 and 9 were screened with the *odhA*.OUT.F/R primers. A band of 4 kb was amplified from Clone 3, which was used as a control (wt) for the colony PCR. A band of 3 kb was expected in the mutants 1, 2 and 9 but could not be detected.
1-3: H16$_{Cas9}$ cells transformed with 100 ng pTarget_HA_*odhA*_01; 4-7: H16$_{Cas9}$ cells transformed with 100 ng pTarget_HA_*odhA*_02; 9-11: H16$_{Cas9}$ cells transformed with 100 ng pTarget_HA_*odhA*_03; C1: PCR control on 100 ng H16 genomic DNA ®

It is possible that the hosts own NHEJ mediated repair system may have introduced indels into the gene, not detectable by gel electrophoresis. Three clones per guide RNA were analysed via sequencing, but the sequences showed no gene disruption via indel addition (**Figure 4.14**). Thus, there was no evidence of NHEJ-repaired Cas9 cleavage in *C. necator*.

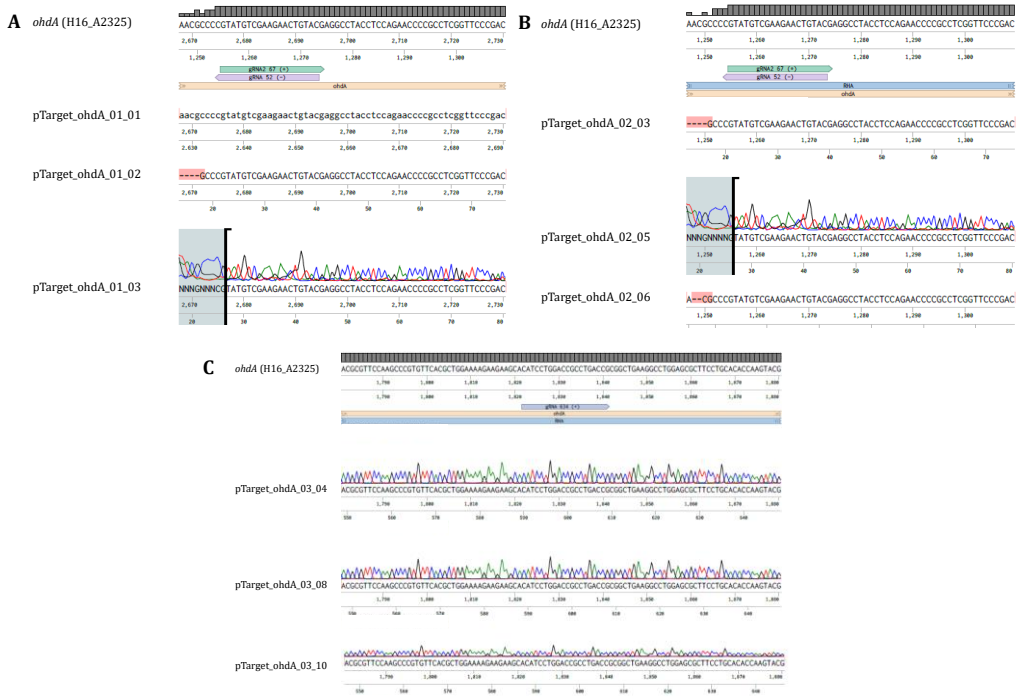


Figure 0.14 – Sequence alignment of H16$_{Cas9}$ cells transformed with pTarget_odhA vectors and re-streaked twice on Cm + Spec

H16$_{Cas9}$ cells transformed with pTarget_*odhA* were repatched on Cm + Spec and screened via cPCR with *odhA*_IN.F/R primers. The PCR products obtained were gel extracted and prepared for Sanger sequencing using *odhA*_IN.F. Three 'NHEJ' potential mutants per guide were screened for indels.
A. H16$_{Cas9}$ cells transformed with pTarget_*odhA*_01 [gRNA (52(-)].
B. H16$_{Cas9}$ cells transformed with pTarget_*odhA*_02 [gRNA (67(+)].
C. H16$_{Cas9}$ cells transformed with pTarget_*odhA*_03 [gRNA (364(+)].

Both attempts at CRISPR/Cas9-driven chromosomal gene deletion in *C. necator* were unsuccessful. Several parameters could be improved to achieve genome editing. For instance, the NHEJ pathway from *C. necator* has never been experimentally characterised. A recent publication reported that both the CRISPR system and the native NHEJ mechanism alone, were unable to generate mutants in *C. necator* (Xiong *et al.*, 2018). In future experiments, the HDR mechanism would be the preferred method and only pTarget vectors including homology arms would be utilised. The two-plasmid strategy was also reconsidered. It could be more efficient to have both the CRISPR elements and the homology arms on a single vector, reducing both the experimental time and energy cost to the cell in

maintaining two large plasmids. However, it was reported that the two-plasmid system allowed for more induction flexibility and helped to fine-tune transient protein expression, especially when other proteins like recombinases, were co-expressed with Cas9 (Jiang *et al.*, 2014; B. Li *et al.*, 2018; Sun *et al.*, 2018). In addition, electroporation of larger plasmids can be difficult, particularly in host backgrounds with a low transformation efficiency. Therefore, the two-plasmid approach was conserved in the next CRISPR experiments.

4.3 An alternative Cas9 for a better expression of the CRISPR/Cas9-based system in *C. necator*

The problems associated with chromosomal cleavage in *C. necator* utilising the Cas9 complex, may arise from the small number of Cas9: sgRNA complexes available in the cytoplasm. This could result from the low expression of Cas9, possibly inferred by codon bias. The *Spycas9* gene sequence has a GC content of 35% whereas *C. necator* has a GC-rich genome (approximately 65%) (Pohlmann *et al.*, 2006). Moreover, the codon preferences are different between both organisms. In H16, G-and C-ending codons are more abundant while ribosomes in *S. pyogenes* are more likely to use U or A as termination codons.

Cobb *et al* identified several rare codons in the sequence of the *Spycas9* gene (Cobb, Wang and Zhao, 2015) and designed a pCRISPomyces system with a refactored *cas9* gene, optimised for expression in three GC-rich Gram-positive *Streptomyces* strains (*S. lividans*, *S. viridochromogenes* and *S. albus*). According to the codon table database (https://www.kazusa.or.jp/codon/), *C. necator* H16 and *Streptomyces* spp. share a high-GC content (about 70% in *Streptomyces* species) and have a more similar codon bias.

The *cas9* gene (*Stmcas9*) was PCR amplified using the plasmid pCRISPomyces-2 (Addgene #61737) as a template and cloned into the low copy-number shuttle vector p15A-pVS1, to generate the plasmid pSV501 (**Figure 4.15A**). The vector was then transformed into *C. necator* ($H16_{StmCas9}$). The *hsdR_N* domain was utilised as a target for the new CRISPR/Cas9 system and the pTarget_HA_*hsdR* vector series was generated.

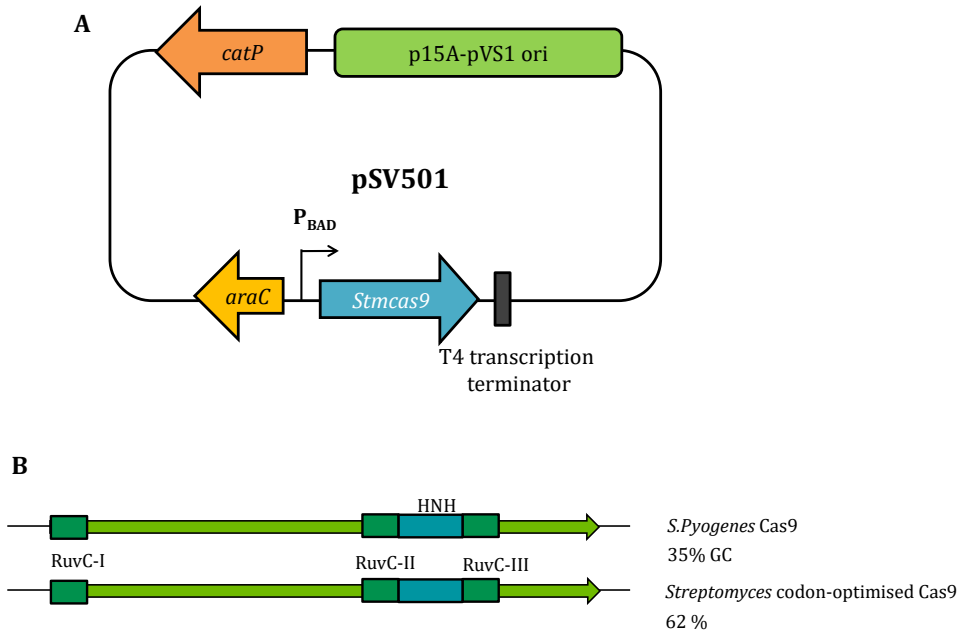


Figure 0.15 – Assembly of an alternative Cas9-expressing vector

A. The backbone of pSV500 was used to construct pSV501. The replicon, selection marker and inducible promoter were conserved; *Spycas9* was replaced with *Stmcas9* followed by the T4 transcription terminator. To streamline the structure and improve transformation efficiency of the vector, the P_{tet}:*sceI* complex was removed.
B. Both the SpyCas9 and StmCas9 have a similar architecture but differ the StmCas9 has a similar GC content to the host background.

Transformation efficiencies obtained with StmCas9 were similar to those obtained with SpyCas9 (**Figure 4.16**). In general, less CFUs were obtained with StmCas9, which may imply a more stringent selection, via better expression of Cas9 and an increased availability of the Cas9: sgRNA complex.

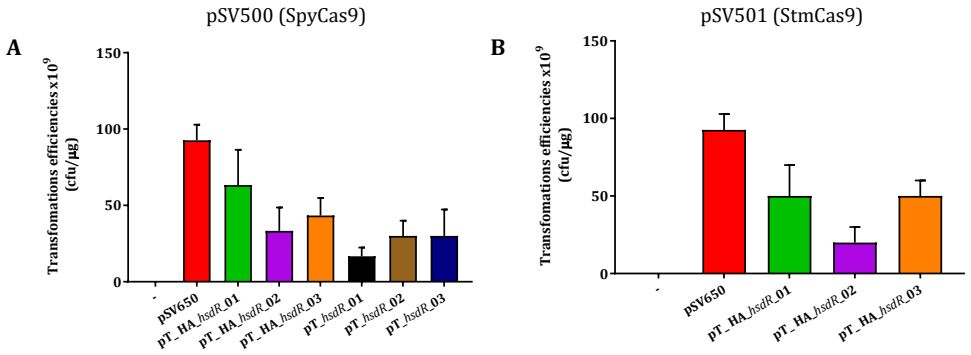


Figure 0.16 – Comparing the efficiency of SpyCas9 and StmCas9 in the deletion of hsdR_N

(A) H16$_{Cas9}$ cells and **(B)** H16$_{StmCas9}$ cells were transformed with either water (_), 100 ng of pSV650 targeting *staA* on the pSV501 plasmid or one of the pTarget_HA_*hsdR* vectors targeting *hsdR_N*. H16$_{Cas9}$ cells were also transformed with the pTarget_*hsdR* series for NHEJ-mediated indel formation.

Potential Δ*hsdR* mutants were screened for amplification of the *hsdR_N* gene with internal primers. All of the selected colonies showed a wild-type profile (**Figure 4.17**). Therefore, it was concluded that StmCas9 was not more efficient in genome editing than SpyCas9.

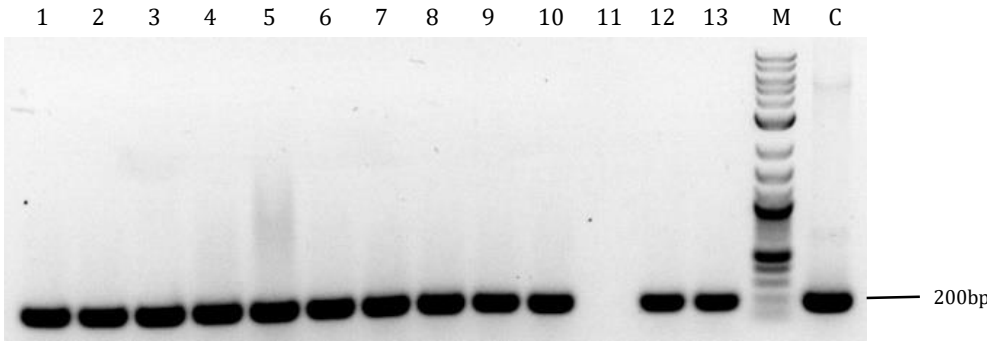


Figure 0.17 – Screening of H16$_{StmCas9}$ hsdR transformants re-streaked on Cm + Spec agar plates via cPCR utilising hsdR_IN.F/R primer pair.

1-4: H16$_{Cas9}$ cells transformed with 100 ng pTarget_HA_*hsdR*_01; 5-8: H16$_{Cas9}$ cells transformed with 100 ng pTarget_HA_*hsdR*_02; 9-10 and 12-13: H16$_{Cas9}$ cells transformed with 100 ng pTarget_HA_*hsdR*_03; C: PCR control on 100 ng H16 genomic DNA; M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb). No reaction was loaded in well 11.
A band of 200 bp was amplified from both the wild type and mutant cells, meaning *hsdR* had not been deleted.

The expression of Cas9 was evaluated in *C. necator,* $H16_{StmCas9.}$ The cells were cultured overnight in 10 mL of SOB medium both with and without 10 mM arabinose. Cells were harvested after 24 h for RNA extraction**.** Following reverse transcription, a diagnostic PCR was performed using 2 µL of cDNA and the primers StmCas9.IN.F/R to evaluate the expression of *cas9* under arabinose induction.

No band was detected in the (-) arabinose sample, whilst a band of ~200 bp was observed in the (+) arabinose sample. This confirmed the transcription of the *cas9* gene following arabinose induction (**Figure 4.18**).

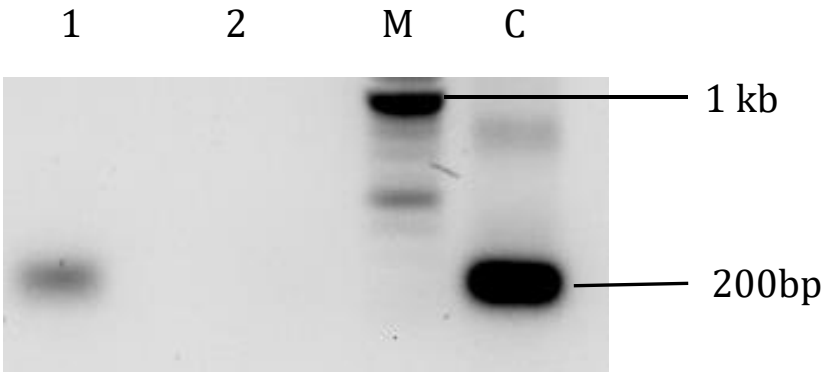


Figure 0.18 – Expression of Stmcas9 from pSV501 in C. necator

Lane 1: Amplification on cDNA from $H16_{Stmcas9}$ incubated in SOB Cm with 10 mM arabinose; lane 2: Amplification on cDNA from $H16_{Stmcas9}$ incubated in SOB Cm without arabinose; C: Control amplification on plasmid pSV501, M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb).

The expression of Cas9 was also evaluated via SDS-PAGE followed by Western Blotting using an anti-Cas9 antibody, but the colorimetric detection method used in the Western Blot protocol did not allow for adequate protein levels to be detected. Due to time constraints, it was not possible to develop a more sensitive detection method for the Cas9 protein expression in *C. necator*.

4.4 Towards a combinatorial system: CRISPR Optimized MAGE Recombineering (CRMAGE)

Multiplex Automated Genome Engineering (MAGE) was developed in 2009 (Wang *et al.*, 2009). The technique, supported by λ-Red recombineering, aims to introduce a pool of degenerate short single-stranded oligonucleotides to target multiple chromosomal loci. After iterations of growth, allelic replacement and recovery, the cell population can be assessed for genotypic/phenotypic changes. A subset of cells could then be processed for subsequent cycles of directed artificial evolution. Seven years later, Ronda and colleagues combined the CRISPR tool to the MAGE procedure to create CRISPR Optimised MAGE (CRMAGE), generating an even faster and more efficient genome engineering and evolution technology in *E. coli* (Ronda *et al.*, 2016). In this technique, CRISPR was implemented as a negative selection factor, to cleave the dsDNA still present in the wild-type protospacer and the adjacent PAM. Thus, the efficiency of CRMAGE was intrinsically improved compared to MAGE (up to 96.5%-99.7%). The team also developed a database for automation of gRNA design, fine-tuning this powerful editing process.

A simplified CRMAGE system was developed for use in *C. necator*. pCRMAGE was developed as a dual Cas9/*recET* expression plasmid (**Figure 4.19**). To ensure efficient production of the protein complexes, the medium-copy number pBBR1 replicon was used instead of the low-copy, shuttle vector p15A-pVS1, as in the previous CRISPR experiments. The kanamycin-resistance gene *aph3* was used as the selection marker in place of the *catP* gene, to circumvent spontaneous chloramphenicol resistance. In addition, the *C. gilardii* homologous *recET* system previously explored in **Chapter 3** and the DNA cytosine methyltransferase gene from *C. necator* JMP134 (Reut_A2987) were also expressed in a recombination/methylation operon, to reduce the restriction of the non-methylated plasmids transformed into the host strain. Two different

inducible systems were selected to finely control expression of each system. The *Stmcas9* gene was induced with arabinose and the recombineering genes were under the control of the endogenous *nahG*/P_{nahG} salicylate-inducible promoter (Hanko and Minton, 2017). This LysR family transcriptional regulator (H16_A1634) is located upstream of a salicylate hydroxylase (H16_A1633) on the minus strand of the chromosome (Pohlmann *et al.*, 2006). It has been characterised as a salicylate-inducible transcriptional repressor, which can be inactivated with low amounts of salicylate and offers a tight on/off regulatory system in *C. necator* H16 (Erik Hanko, personal communication). Further design optimisation was carried out on the sgRNA plasmid. The sgRNA was expressed on the low-copy number vector p15A/pVS1 and selected on tetracycline, a stringent antibiotic in *C. necator*.

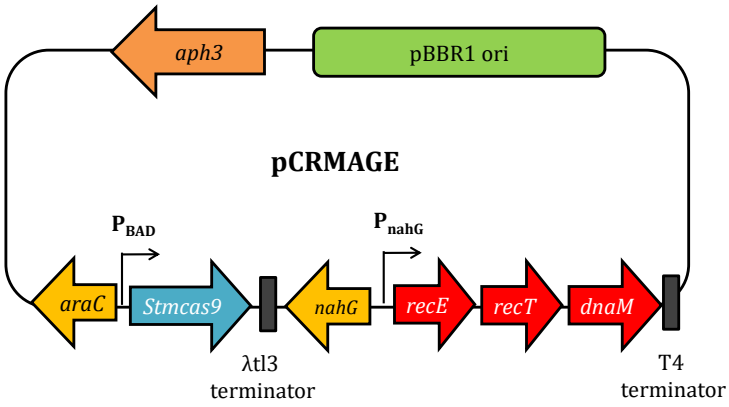


Figure 0.19 – Schematic representation of the CRMAGE plasmid

pCRMAGE expresses the *Stmcas9* gene under the arabinose inducible promoter (P_{BAD}) and the recombination/methylation operon consisting of the *recET* system, codon-optimised for *C. necator* and the DNA cytosine methyltransferase from *C. necator* JMP134 (Reut_A2987) under the salicylate-inducible *nahG*-P_{nahG} *system.*
aph3: aminoglycoside 3' phosphotransferase (confers resistance to kanamycin); *araC*: arabinose operon regulatory protein; *nahG*: salicylate hydroxylase; *recE*: exodeoxyribonuclease VIII; *recT*: recombinase; *dnaM:* DNA cytosine methyltransferase.

The pTarget series was redesigned for optimal CRMAGE-driven gene deletion with a new target, the poly (3-hydroxybutyrate) polymerase-encoding gene, *phaC*. The *phaC* knockout produces a visually distinguishable phenotype, in clones grown in the presence of Nile-Red.

Figure 4.20 summarises the modifications made to the pTarget series including (i) the promoter, (ii) length of the homology arms and (iii) the addition of a second terminator, following the termination signal of the single guide. The final pTarget plasmid included the constitutive transcription of the sgRNA under the control of pIJ23119 and rrnBT2, as well as 1 kb-long homology arms flanking the *phaC* gene.

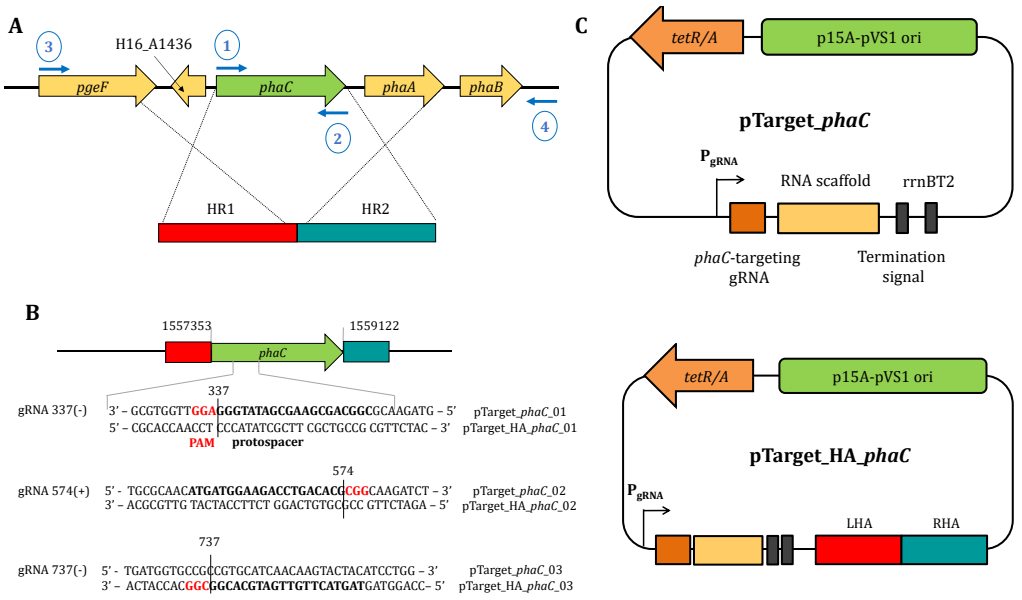


Figure 0.20 – Redesigning the pTarget vector series to achieve a knock-out of the phaC gene in C. necator H16

A. Representation of the *phaC* region targeted by the homologous arms of the pTarget_HA_*phaC* series.
pgeF: peptidoglycan editing factor, *phaC*: poly(3-hydroxybutyrate) polymerase, *phaA*: acetyl-CoA acetyltransferase, *phaB*: acetoacetyl-CoA reductase; HR: homology region. H16_A1436 is annotated as a hypothetical protein in GenBank. The homology regions are complimentary to the sequences immediately upstream or downstream *phaC.* Primer 1: *PhaC*.IN.F, 2: *PhaC*.IN.R, 3: *PhaC*.OUT.F, 4: *PhaC*.OUT.R. Primer pair 1+2 binds on both edges of the domain *phaC_N* and produces a PCR band of 500 bp whilst primers 3 and 4 amplify a product of 3.8kb from the wild type chromosome, 3.2kb after recombination.
B. The gRNA sequences, identical to the protospacer region of the chromosome (in bold) selected for knockout of the *phaC* gene using CRISPR/Cas9 from *S. pyogenes*. The number and sign attributed to each gRNA identifies the nucleotide and strand of the CDS where Cas9 will cleave the DNA, with nucleotide 1 being the first nucleotide in the coding sequence.
C. LHA and RHA design was revisited to include homologous arms of 0.5-1 kb. P_{gRNA} represents the rhamnose-inducible P_{Rham} or the constitutive $P_{IJ23119}$. The gRNAs 387(-), 574(+) and 737(-) were inserted into the two gRNA-transcribing backbones, pTarget_*phaC* and pTarget_HA_ *phaC* and validated by sequencing. Finally, rrnBT2, a strong terminator T2 of the ribosomal RNA gene *rrnB* in *E. coli* was added to the tail of the sgRNA to ensure transcription would not overrun and produce a longer RNA than desired.

The CRMAGE system was utilised to delete the *phaC_N* domain. An additional control was implemented to assess the performance of pSV600, a pBBR1-based plasmid similar in size to the pTarget_*phaC* plasmid, but with the additional inclusion of a non-targeting gRNA. The guide sequence of pTarget developed in the work of Jiang *et al* (Jiang *et al.*, 2015) targeted *cadA* (JW4092), coding for the inducible lysine decarboxylase responsible for cadaverine production, in *E. coli* MG1655. Blast searches and alignments with the 20 bp *cadA* spacer (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and the *C. necator* H16 genome showed no significant alignment. The sequence was therefore used as a gRNA0 control to evaluate gRNA specificity/off-target effects.

Transformation efficiencies were two to four times higher with CRMAGE than with the CRISPR/SpyCas9 and CRISPR/StmCas9 systems with equivalent quantities of DNA (**Figures 4.21 and 4.9, 4.12 and 4.16**). Transformation efficiencies with pSV600 reached 3.23×10^{11} cfu/µg, which was ~40% higher than with the other pTarget plasmids, thus proving that a strain-specific targeting guide was necessary to cleave efficiently in the chromosome.

Similar CFUs were observed with both the plasmid series with either the NHEJ- or HDR-editing vectors, except for pTarget_HA_*phaC*_02, which had a lower transformation efficiency. The similarity in CFUs obtained independently of the donor DNA was unexpected. Expression of the recombineering *recET* genes and the methylase *dnaM* gene was expected to promote homologous repair and consequently increase the number of transformants achieved, relative to the NHEJ vectors. All transformants were re-patched onto fresh LB plates supplemented with 300 µg/mL kanamycin and 15 µg/mL tetracycline and screened via cPCR.

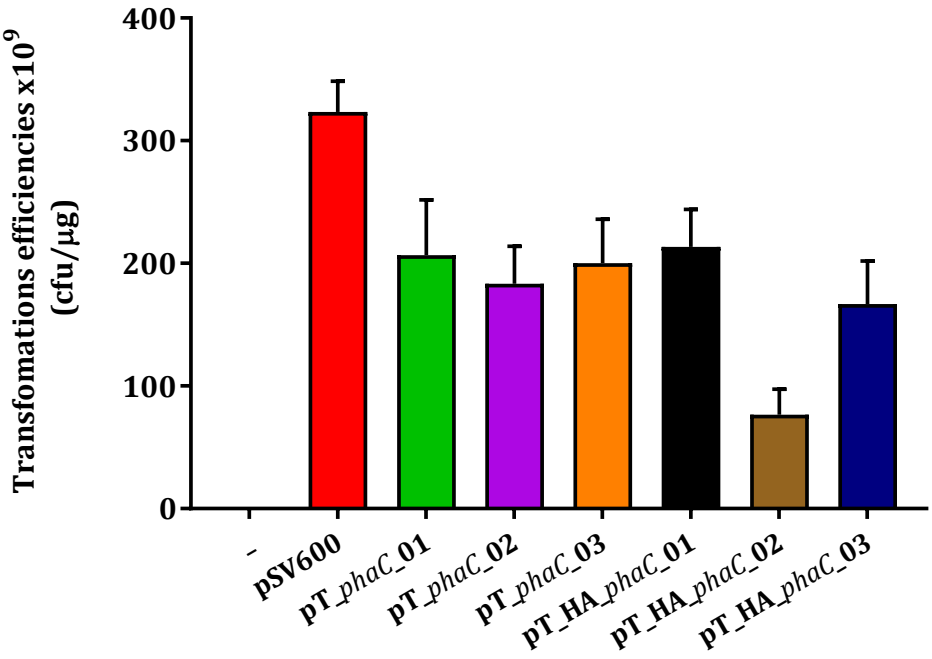


Figure 0.21 – Testing the efficiency of the CRMAGE system in the deletion of the phaC_N gene

H16$_{CRMAGE}$ cells were induced with arabinose and transformed with either water (_), 100 ng of pSV600 ('gRNA0 control') and one of the pTarget_*phaC* or pTarget_HA_*phaC* plasmids, targeting *phaC_N*, the conserved protein domain of *phaC* (H16_A1437), located at the N-terminal end of the protein.

PCR screening with *phaC* internal primers, generated a band of 500 bp in 5 out of the 30 colonies tested, these included all the different target plasmid variants (**Figure 22A**). However, all the cells transformed with the non-targeting pSV600 control vector had a wild-type profile i.e. a band of 500 bp (**Figure 22B**). Therefore, no spontaneous homologous recombination occurred when the Cas9:sgRNA0 and the donor DNA were utilised. The potential gene deletions observed in the pTarget_*phaC* and pTarget_HA_*phaC* transformants should therefore, only result from CRMAGE-driven editing. The colonies, which produced no internal band, were considered as CRMAGE-edited mutants. All the 30 colonies were screened further with primers flanking the *phaC* gene for confirmation.

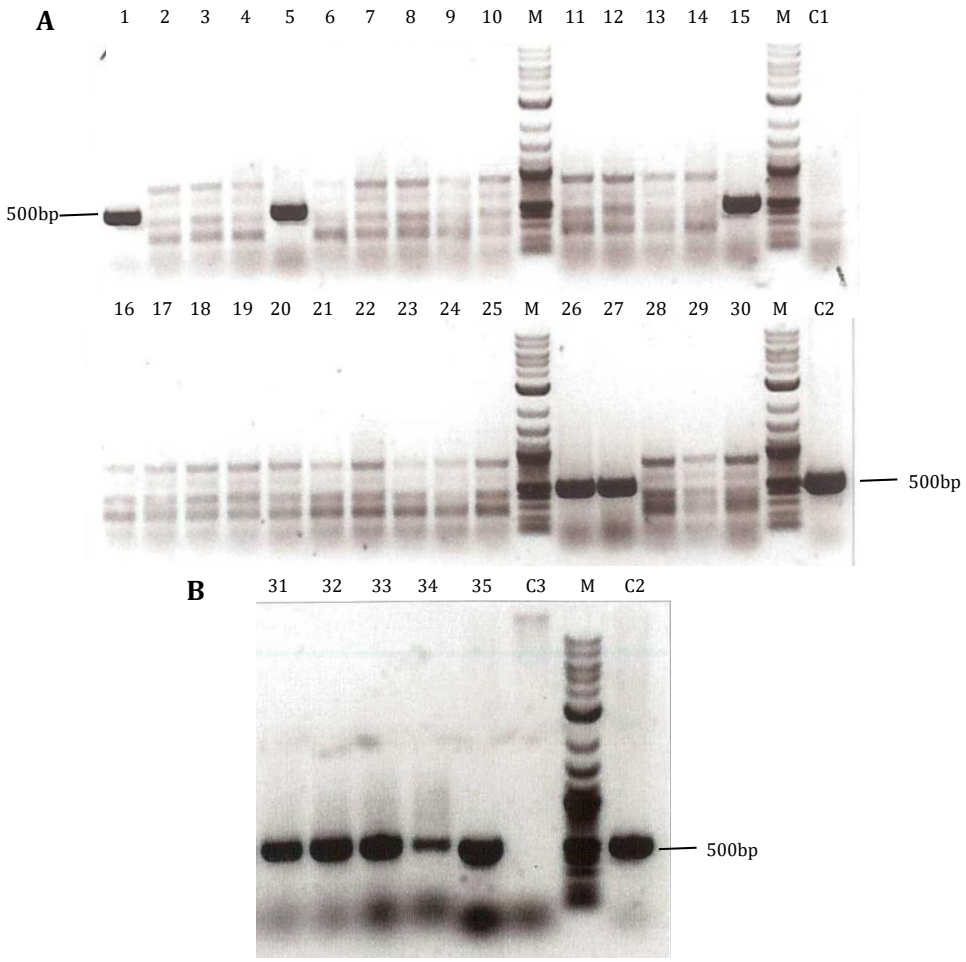


Figure 0.22 – Screening of phaC_N potential deletion mutants and control transformants re-streaked on LB Km + Tet agar plates via cPCR utilising phaC_IN.F/R primer pair

Colonies were screened via PCR utilizing *phaC*_IN.F/R. A band of 500 bp was amplified from the wild type cells while the mutant cells generated no band.
A. 1-5: H16$_{CRMAGE}$ cells transformed with 100 ng pTarget_*phaC*_01; 6-10: H16$_{CRMAGE}$ cells transformed with 100 ng pTarget_HA_*phaC*_01; 11-15: H16$_{CRMAGE}$ cells transformed with 100 ng pTarget_*phaC*_02; 16-20: H16$_{CRMAGE}$ cells transformed with 100 ng pTarget_HA_*phaC*_02; 21-25: H16$_{CRMAGE}$ H16 CRMAGE cells transformed with 100 ng pTarget_*phaC*_03; 26-30: H16$_{CRMAGE}$ cells transformed with 100 ng pTarget_HA_*phaC*_03; C1: primer dimers; C2: PCR control on 100 ng H16 genomic DNA; M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb).
B. 31-35: H16$_{CRMAGE}$ cells transformed with 100 ng pSV600; C3: PCR control on 100 ng CH34 genomic DNA.

All 30 colonies screened with external *phaC* primers displayed a wild-type 3.8 kb band, despite the previous absence of an internal band (**Figure 4.23**). Therefore, no homologous recombination-based deletion of the conserved domain was detected in the Km + Tet-resistant cells.

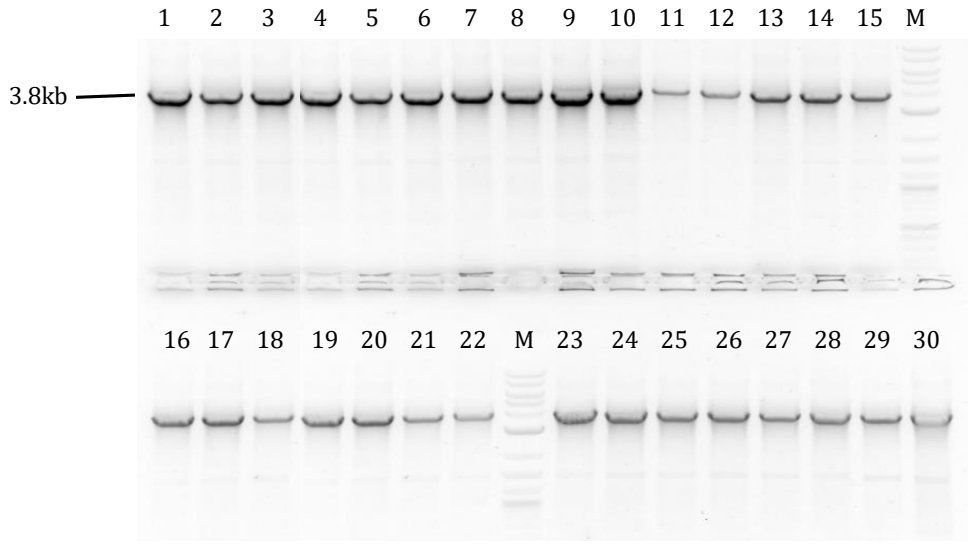


Figure 0.23 – Screening of phaC_N H16 deletion mutants re-streaked on LB Km + Tet agar plates via cPCR utilising PhaC_OUT.F/R primer pair

1-5: H16$_{CRMAGE}$ cells transformed with 100 ng pTarget_*phaC*_01; 6-10: H16$_{CRMAGE}$ cells transformed with 100 ng pTarget_HA_*phaC*_01; 11-15: H16$_{CRMAGE}$ cells transformed with 100 ng pTarget_*phaC*_02; 16-20: H16$_{CRMAGE}$ cells transformed with 100 ng pTarget_HA_*phaC*_02; 21-25: H16$_{CRMAGE}$ H16 CRMAGE cells transformed with 100 ng pTarget_*phaC*_03; 26-30: H16$_{CRMAGE}$ cells transformed with 100 ng pTarget_HA_*phaC*_03; M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb). A band of 3.8 kb was detected in the wild type and mutant colonies.

In parallel to the PCR and sequencing analysis, the CRMAGE colonies were screened on 0.4% Sodium Gluconate-Minimum Medium (0.4% SG-MM) agar plates supplemented with Nile-Red®. This lipophilic stain reacts with lipids and produces an intense fluorescence, when excited under blue light (Gorenflo *et al.*, 1999). Therefore, potential mutants were streaked onto Nile-Red® plates with both the wild type and Δ*phaC* mutant, allowing for comparison of their fluorescence profiles under blue light. All the colonies emitted a slightly reduced level of fluorescence, compared to the wild-type strain, while the *phaC* mutant had no detectable fluorescence (**Figure 4.24**). Both the controls and the CRMAGE colonies had a similar growth pattern on the selective medium, so the variation in intensity could not be attributed to a growth defect. It is interesting to note that Clone 1 (H16$_{CRMAGE}$/pTarget_*phaC*_01) produced an internal *phaC* band when screened by cPCR, so therefore should have displayed a wild-type

phenotype on the Nile red plates. However, it also had a reduction in fluorescence, similar to the other mutants, with no internal *phaC* band. Potential point mutations in the *phaC* gene causing a reduction of poly 3-hydroxybutyrate production could be the reason for this reduction in fluorescence. The inclusion of indels or other point mutations at the cleavage site targeted by Cas9 was investigated.

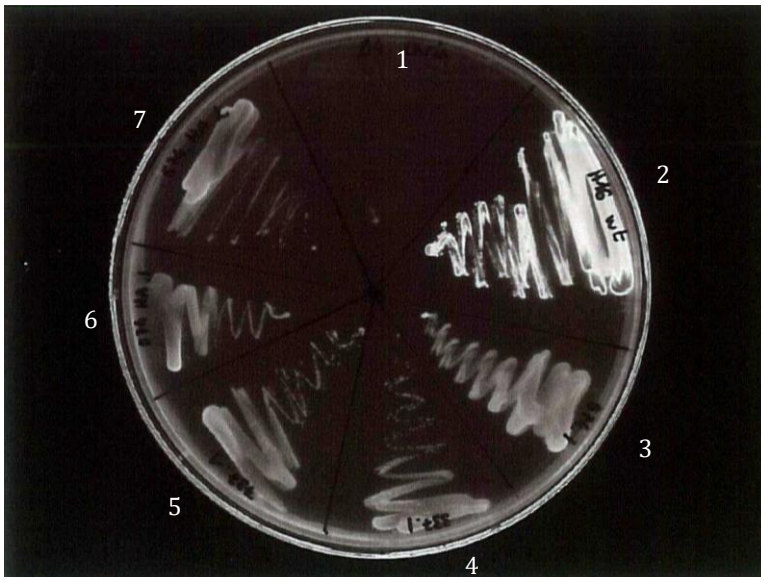


Figure 0.24 – Phenotypic screening of potential ΔphaC mutants on Nile-Red minimal medium agar plates

1: Δ*phaC* mutant strain; 2: *C. necator* H16 wild-type strain; 3: H16$_{CRMAGE}$/pTarget_*phaC*_02 (Clone 11 on **Figures 4.21 and 4.22**); 4: H16$_{CRMAGE}$/pTarget_*phaC*_01 (Clone 1); 5: H16$_{CRMAGE}$/pTarget_*phaC*_03 (Clone 21); 6: H16$_{CRMAGE}$/pTarget_*phaC*_HA_02 (Clone 16); 7: H16$_{CRMAGE}$/pTarget_*phaC*_HA_02 (Clone 17). After transformation and colony purification, CRMAGE colonies were re-streaked onto fresh 0.4% SG-MM agar plates supplemented with 100 ng/mL Nile-Red® and then incubated for 24 h at 30°C. To discriminate wild type from mutant colonies, the plate was photographed using a UV transilluminator.

Three colonies were selected for sequencing, which confirmed that none of the clones had been edited with the CRMAGE system, as they all had the conserved wild type sequence at the location of each targeted protospacer (**Figure 4.25**). Thus, the combination of StmCas9, *recET* and *dnaM* did not yield successful deletion mutants in *C. necator*.

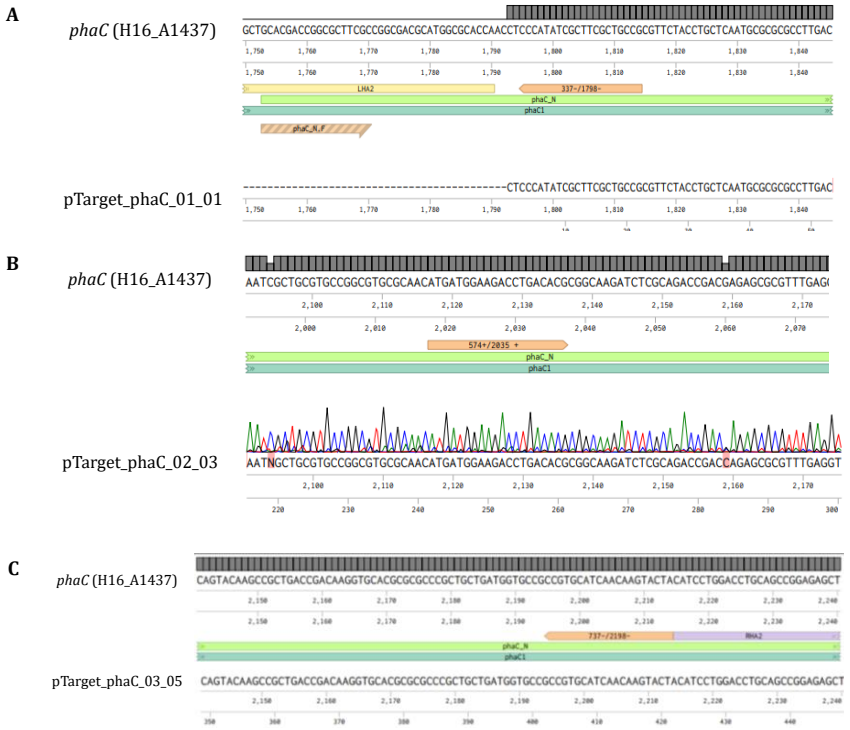


Figure 0.25 – Sequence alignment of H16$_{CRMAGE}$ cells transformed with the pTarget_phaC vectors and re-streaked twice on Km + Tet

H16$_{CRMAGE}$ cells transformed with pTarget_*phaC* were re-patched on Km + Tet and screened via cPCR with *phaC*_IN.F/R primers. The clones sequenced here correspond to the mutants screened via cPCR in **Figure 4.22**, lanes 1, 15 and 26. The PCR products obtained were gel extracted and prepared for Sanger sequencing using *phaC*_IN.F. Three 'NHEJ' potential mutants per guide were screened for indels in one of the experimental replicates but only one sequence is shown above, as they all had a wild-type sequence.
A. H16$_{CRMAGE\ cells}$ transformed with pTarget_*phaC*_01 [gRNA (337(-)].
B. H16$_{CRMAGE}$ cells transformed with pTarget_*phaC*_02 [gRNA (574(+)].
C. H16$_{CRMAGE}$ cells transformed with pTarget_*phaC*_03 [gRNA (737(-)].

Although the CRISPR system proved highly efficient at removing plasmids in *C. necator*, it was less efficient in generating chromosomal deletions. The low yield of recombinants selected after transformation of the guide plasmid into the Cas9-expressing cells strongly suggested that the endonuclease was able to cleave the DNA. Consequently, the host DNA damage repair system might be the bottleneck preventing effective utilisation of the CRISPR system in *C. necator*. This is further supported by the results presented in **Chapter 3,** where homologous recombineering

also proved difficult. Therefore, deciphering the molecular mechanics of recombination repair would be extremely useful in understanding how to resolve the problems associated with recombineering and CRISPR.

4.5 Establishing an efficient CRISPR/Cas9 system in *C. necator*

Ines Canadas (SBRC Nottingham) developed a CRISPR system for use in Clostridia, which allows for the tight control of *cas9* expression utilising a theophylline riboswitch (**Figure 4.26**).

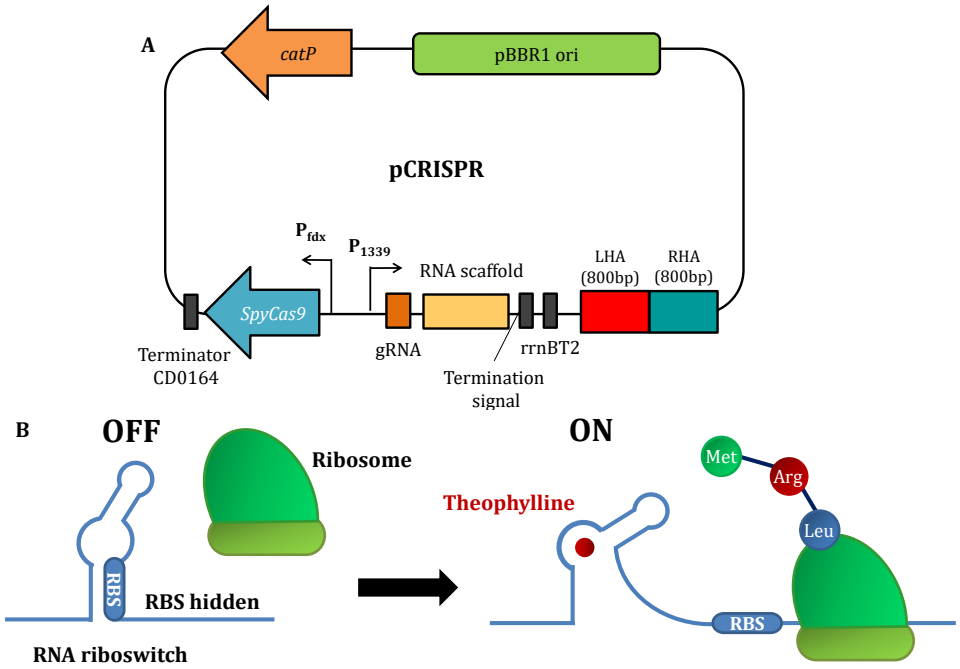


Figure 0.26 – Single-plasmid riboswitch-controlled CRISPR system in C. necator

A. pCRISPR is a chloramphenicol-resistant medium-copy plasmid expressing the native *S. pyogenes cas9* (*Spycas9*) downstream of the constitutive ferredoxin P_{fdx} promoter. The initiation of translation is controlled via a riboswitch, which overlaps the RBS sequence of the mRNA and prevents Cas9 production in the absence of theophylline. In the opposite direction is a constitutively transcribed p1339 (Cac *araE*) - driven sgRNA. The guide was programmed to bind either the *cadA* spacer of *E. coli* (pCRISPR_gRNA0) or the *phaC* locus in the genome of *C. necator* (pCRISPR_*phaC*). All pCRISPR plasmids displayed 800 bp-long left and right homology arms for HDR repair.
B. Riboswitch molecular mechanism. In the ligand-free form, the riboswitch conformation impairs ribosome binding to the RBS. Theophylline binding to the aptamer domain results in rearrangement of the translational initiation region and the RBS sequences becomes available.

This system offered two main differences compared to the CRISPR tools developed in this study. First, it was composed of a single plasmid combining the expression of SpyCas9 and the transcription of a sgRNA. Second, a theophylline riboswitch tightly controlled the expression of Cas9

at the translational level (Topp *et al.*, 2010). The riboswitch formed by the 5' untranslated, transcribed region from the ferredoxin promoter overlapped with the ribosome binding site (RBS) and formed a stem-loop RNA structure, that restricted translation initiation of the downstream *cas9* open reading frame. Upon addition of theophylline in the medium, the aptamer of the riboswitch acted as a receptor to the small molecule and prevented the masking of the RBS in the mRNA, allowing translation to begin (Van Vlack and Seeliger, 2015). This structure was actually the core of the system as it had shown great, fast and dynamic response to the inducer molecule and tight repression of translation of the downstream open reading frame, in the absence of theophylline (Rudolph, Vockenhuber and Suess, 2013). Therefore, it offered the perfect regulation system for expression of Cas9, which is a key factor to achieve genome editing in *C. necator*.

Three pCRISPR plasmids were used to assess the riboswitch-controlled CRISPR system: pCRISPR_gRNA0 which included a control gRNA targeting the *E. coli cadA* gene, pCRISPR_574(+) and pCRISPR_737(-) targeting the *phaC_N* conserved domain within the *phaC* gene, on either strand of the gene. The gRNA 574(+) promotes Cas9 cleavage after the 574th nucleotide of the *phaC* CDS, while the gRNA 737(-) mediates the introduction of a DSB before the 737th nucleotide. All the pCRISPR plasmids harboured homology arms complementary to the flanking regions of the *phaC_N* domain and were individually transformed into *C. necator* H16 to generate H16/pCRISPR_gRNA0, H16/pCRISPR_574(+) and H16/pCRISPR_737(-). A single colony of each transformant strain was inoculated in 10 mL LB, supplemented with 60 μg/mL chloramphenicol (Cm) at 30°C for 96 h. This recovery time allowed native recombination mechanisms to promote allelic exchange between the plasmid-borne homology arms and the genomic sequence flanking the *phaC_N* domain. 100 μL of each culture was diluted in PBS and 100 μL of the appropriate

dilution was spread onto two sets of Cm plates, one with 5 mM theophylline (Theo) for induction of *cas9* expression and the other one without inducer (No Theo). Plates were incubated for 48 h at 30°C and colonies were enumerated.

Comparable CFUs were enumerated with pCRISPR_gRNA0 on Theo and No Theo plates, which suggests the addition of theophylline does not hamper cell growth significantly (**Figure 4.27**). In cells transformed with either pCRISPR_574(+) or pCRISPR_737(-), the induction of Cas9 resulted in the elimination of ~80% of the cells. This clearly highlighted the functionality and the efficiency of the riboswitch-controlled CRISPR system.

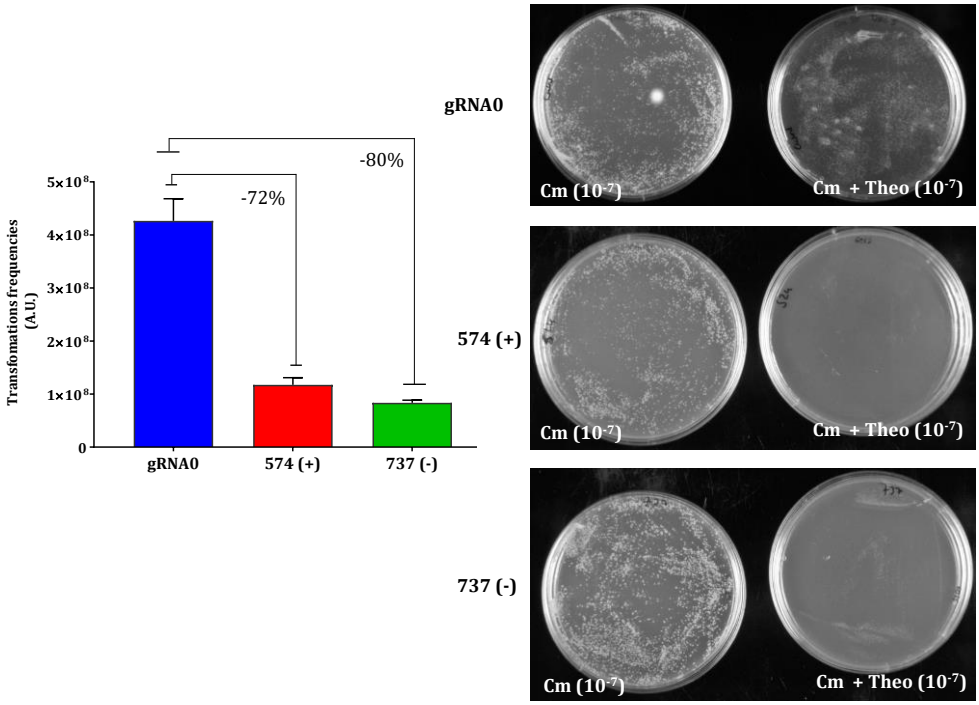


Figure 0.27 – Testing the efficiency of the riboswitch-based CRISPR/Cas9 system in C. necator

A. Transformation frequencies of H16/pCRISPR_gRNA0, H16/pCRISPR_574(+) and H16/pCRISPR_737(-) strains after theophylline-induced expression of Cas9. Transformation frequencies (A.U.) were estimated as follows
TF = $(Cm^R{+}Spec^R\ colonies)/\ (Cm^R\ colonies)$ x10^9.
TF (gRNA0) was set to 100% survival and cleavage efficiencies as the percentage of survival loss in either 574(+) and 737(-) transformants.
The experiment was conducted on two biological replicates per guide.
B. Pictures from a representative experiment where *C. necator* H16 cells were transformed with either pCRISPR_gRNA0, pCRISPR_574(+) or pCRISPR_737(-). The left part of the panel shows the LB agar plates containing Cm that were used to determine total viable cells. The

right part of the panel displays the LB agar plates supplemented with Cm + Theo where Cas9 was induced.

Twelve colonies from each Theo plate and three from each No Theo plate were re-streaked onto a fresh Cm plate for screening. All the colonies were screened by colony PCR for the deletion of *phaC_N* with both internal and external primers, as described previously**.** As expected, all the H16/pCRISPR_gRNA0 colonies generated a wild-type internal PCR band of 500 bp (**Figure 4.28A**). 11 out of 12 H16/pCRISPR_574(+) colonies and 2 out of 12 H16/pCRISPR_737(-) colonies did not produce any band, suggesting a gene deletion. However, after a second confirmation PCR (data not shown), clones 12 (H16/pCRISPR_574(+)), 18 and 19 (H16/pCRISPR_737(-)), initially considered as potential mutants, showed a wild-type profile (data not shown). The low transformation frequencies achieved with pCRISPR_737(-) combined with the results of cPCR screening underpin the sensitivity of Cas9 regarding the sequence and location of the target protospacer. Although the lagging (-) strand of the gene is often considered as more proficient in recombination (Lim, Min and Jung, 2008; Mosberg, Lajoie and Church, 2010), it seems in this particular case, that targeting the (+) strand could be more favourable for allelic exchange (Richardson *et al.*, 2016). The external PCR conducted on the last ten H16/pCRISPR_574(+) colonies resulted in a mixture of bands with various sizes (**Figure 4.28B**): clones 1 and 6 showed a wild-type 3.8 kb band while the other clones generated PCR products of varying sizes, between 3.2 and 3.8 kb. All PCR products were further analysed by Sanger sequencing.

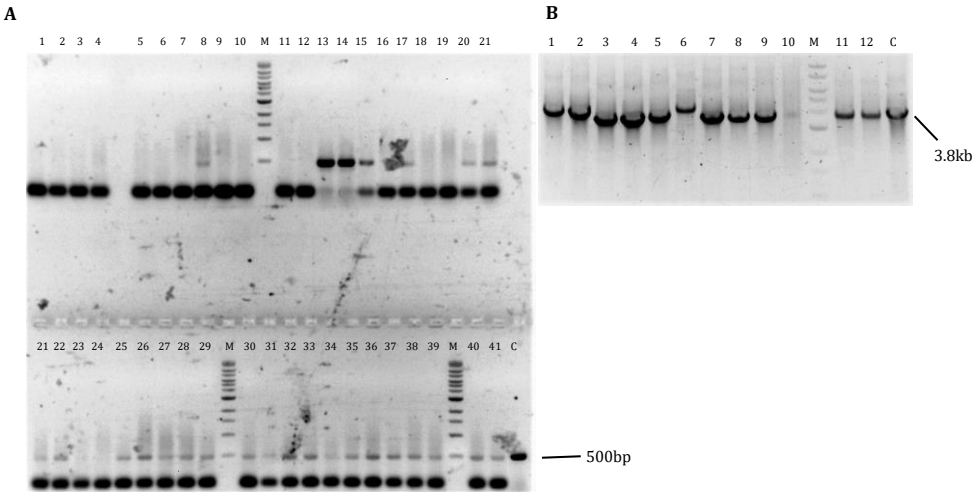


Figure 0.28 – Screening of H16/pCRISPR clones via cPCR utilising PhaC_IN.F/R and PhaC_OUT.F/R primer pairs

A. 1-12: H16/pCRISPR_574(+) from Theo plate; 13-24: H16/pCRISPR_737(-) from Theo plate; 37-38: H16/pCRISPR_574(+) from No Theo plate; 39-40: H16/pCRISPR_737(-) from No Theo plate; 41-42: H16/pCRISPR_gRNA0 from No Theo plate; C: 100 ng of genomic H16 DNA; M: NEB® 2-Log DNA ladder (0.1 – 10.0 kb). Cells were screened for deletion of *phaC_N* with *PhaC*_IN.F/R primers. A 500 bp band was detected in the control and some of the colonies, where CRISPR had failed, while no band suggested the domain had been excised from the chromosome.
B. 1-4: H16/pCRISPR_574(+) 2.1-4; 5-10: H16/pCRISPR_574(+) _1.1-6.
Colonies that did not show the wild-type band after internal cPCR were screened with external *PhaC*_OUT.F/R primers. The wild-type product migrated at 3.8 kb while PCR on the potential mutant cells generated a band of 3.2 kb.

In addition, phenotypic screening was performed on minimum medium supplemented with Nile-Red, to stain the PHB granules and isolate *phaC* mutants from wild-type strains. The colonies H16/pCRISPR_574(+)_1.1, H16/pCRISPR_574(+)_1.3 and H16/pCRISPR_574(+)_2.3 showed a weak intensity under white light with no filter (**Figure 4.29A**), similar to the *phaC* mutant, while the clone H16/pCRISPR_574(+)_1.2 produced the brightest signal under no filter and with UV light (**Figure 4.29B**), demonstrating wild type activity of the *phaC* gene. The phenotypical screen of *phaC* mutants on Nile-Red plates confirmed the results obtained after amplification of the targeted *phaC* region (**Figure 4.29B**) and could be used to establish a robust screening method for the generation of PHB-deficient mutants.

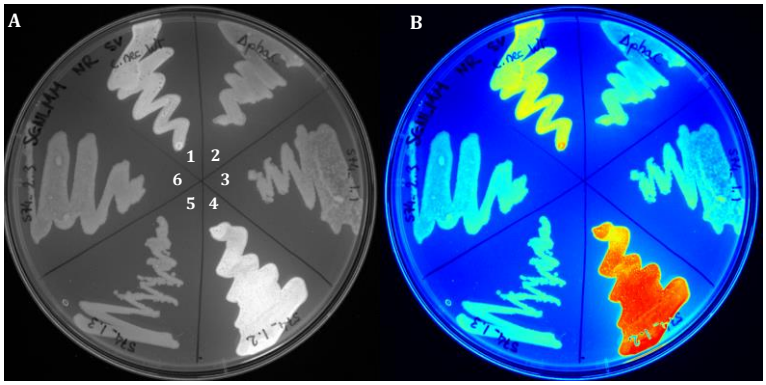


Figure 0.29 – Phenotypical screening of CRISPR-edited ΔphaC mutants on Nile-Red minimum medium agar plates

H16/pCRISPR_574(+) colonies were re-streaked onto fresh 0.4% SG-MM agar plates (ratio carbon/nitrogen = 4) supplemented with 100 ng/mL Nile-Red® and incubated for 24 h at 30°C. To discriminate wild type from mutant individuals, the plate was photographed with (**A**) no filter or (**B**) using a UV transilluminator. The red or green mutants under UV are wild-type PHB-producing strains as the intensity signal reveals the accumulation of polymers under nitrogen deficiency. Consequently, the strains producing a less intense single, like Δ*phaC* control strain, are PHB-negative strains.
1: *C. necator* H16 wild-type strain 2: Δ*phaC* mutant strain; 3: H16/pCRISPR_574(+) _1.1; 4: H16/pCRISPR_574(+) _1.2; 5:H16/pCRISPR_574(+) _1.3; 6: H16/pCRISPR_574(+)_2.3.

Sequencing was utilised to assess selected clones. Five of the ten clones sequenced displayed deletion of the *phaC_N* domain, confirming the generation of CRISPR-edited *C. necator* cells with an editing efficiency of 5 confirmed mutants/10 potential mutants after internal PCR and 5 confirmed mutants /12 screened mutants (~40%) (**Figure 4.30**). Although the CRISPR efficiency is lower than the statistical efficiency of the sucrose-mediate counterselection method (50%) (Lenz, Lauterbach and Frielingsdorf, 2018), these are only preliminary results and much more optimisation is required to develop an efficient CRISPR/Cas9 system in *Cupriavidus* species, including validating the system with other gene targets. The experimental procedure now takes 10 days instead of 3 weeks, the time required to generate a mutant utilising the *sacB*-based method (Lenz, Lauterbach and Frielingsdorf, 2018; Xiong *et al.*, 2018).

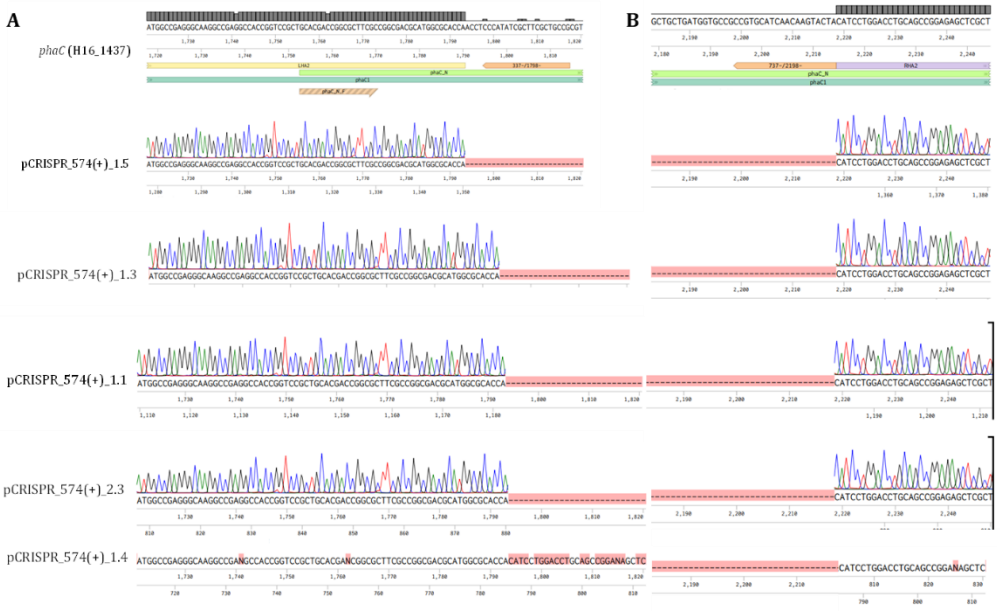


Figure 0.30 – Sequence alignment of H16/pCRISPR_574 cells transformant re-streaked on Cm agar plates

A. Sequence alignments for sequences overlapping the left homology arm LHA2 and the 5' end of *phaC_N*.
B. Sequence alignments for sequences overlapping the 3' end of *phaC_N* and the right homology arm RHA2.
H16 cells transformed with pCRISPR_574(+) were re-streaked on Cm agar plates and screened via cPCR with *PhaC*_IN.F/R and *PhaC*_OUT.F/R primers. The PCR products obtained were gel extracted and prepared for Sanger sequencing using the *phaC*.seq.F.

4.6 Discussion

The CRISPR/Cas9 system is a promising tool to edit genomes rapidly, from a single nucleotide to large clusters, in a highly efficient fashion, paving the way for HTP metabolic engineering and systems biology.

In this study, the application of the CRISPR/Cas9 system was investigated to enhance HTP strain engineering in *Cupriavidus* species, starting with the well-studied chassis *C. necator*.

A series of assays and improvements in the design of the system (**Table 4.1**) included the assessment of:

- A two-plasmid versus a single-plasmid system,
- The archetype *Spycas9* gene and the *Stmcas9* gene, codon-optimised for *Streptomyces* strains,
- The expression of the *cas9* gene on a low-copy number replicon (pVS1 – on pSV500 and pSV501 plasmids) and on a medium-copy number replicon (pBBR1 – on pCRMAGE and pCRISPR),
- The expression of the *cas9* gene under control of the arabinose-inducible promoter P_{BAD} and the theophylline-inducible riboswitch regulating the activity of the ferredoxin promoter $P_{fdx,}$
- The co-expression of a *Cupriavidus* homologous *recET/dnaM* system to enhance recombination efficiencies (pCRMAGE),
- The constitutive (pIJ23119 promoter) and inducible (salicylic acid-inducible p_{nahG} promoter) transcription of sgRNA,
- The addition of an extra termination signal (*rrnBT2),* downstream the sgRNA sequence,
- Multiple gene targets (the genes *staA*, *hsdR, odhA* and *phaC*) on heterologous plasmids and within the genome of *C. necator*,
- Multiple gRNAs targeting forward and reverse strands of the target gene,

- The investigation of NHEJ and HDR pathways,
- Optimisation of homology arms (location, length, sequence) to improve homologous repair.

Table 0.1 – Improvement in design operated in the CRISPR/Cas9 system in C. necator

	Cas9 nuclease and promoter	***recET* recombineering system**	**sgRNA**	**Homology arms**
pSV500	pBAD_SpyCas9	_	pIJ23119	1 kb
pSV501	pBAD_StmCas9	_	pIJ23119	1 kb
pCRMAGE+ pTarget_01	pBAD_StmCas9	p*nahG_recET/dnaM*	pRham	800 bp flanking *phaC*
pCRMAGE+ pTarget_02	pBAD_StmCas9	p*nahG_recET/dnaM*	pRham	800 bp flanking *phaC_N*
pCRMAGE+ pTarget_03	pBAD_StmCas9	p*nahG_recET/dnaM*	pIJ23119	800bp flanking *phaC_N*
pCRMAGE+ pTarget_04	pBAD_StmCas9	p*nahG_recET/dnaM*	pIJ23119 +rrnBT2	800 bp flanking *phaC_N*
pCRISPR	Theo-rb + pfdx_SpyCas9	_	p1339+rrnBT2	800bp flanking *phaC_N*

Theo-rb: theophylline riboswitch.

The native *S. pyogenes* CRISPR/SpyCas9 complex enabled plasmid cleavage *in C. necator* with an efficiency of 88% but could not generate chromosomal modifications, suggesting that the host-mediated repair and low recombination efficiencies in *C. necator* seriously impeded the generation of CRISPR-edited mutants (Jiang *et al.*, 2017).

The utilisation of the riboswitch-based pCRISPR vector helped to develop a functional method for gene editing in *C. necator*. The main difference compared to the other systems described in this study included the addition of an extended incubation time with the donor DNA, which may have increased the proportion of desired recombination events. The transient induction of Cas9 may also have enabled the elimination of non-edited cells. Genetic and phenotypic analysis of the mutants revealed an

editing efficiency of 40% after 4 days of Cas9-free incubation. 100% of the control cells (with *E. coli*-specific sgRNA and *C. necator* targeting homology arms) remained wild type, which established strain-specificity for the guide sequence and tight control of the expression of the Cas9 nuclease.

The first demonstration of an electroporation-based CRISPR/Cas9 genome editing method in *C. necator* was recently published (Xiong *et al.*, 2018). In that study, five genes were edited from the Type I RM system (including the *hsdR* gene), with efficiencies ranging from 78.3 to 100%. However, the system requires extensive incubation time (up to 168 h) and relies on continuing arabinose induction, over the whole incubation period. Although the expression levels of pBAD-controlled *rfp* generated a high fluorescence signal, there was no evidence that, when the inducible system was placed upstream of the *cas9* gene, continuous expression could be maintained over 168 h. A thorough kinetic study of arabinose utilisation has already been conducted in *E. coli* and established a gradual onset of arabinose induction (Fritz *et al.*, 2014). It would be very insightful to perform a similar investigation in *C. necator* to support the results obtained by Xiong *et al,* potentially optimising the utilisation of the arabinose-inducible P_{BAD} promoter in the CRISPR systems developed in the present study.

In conclusion, a deeper understanding of the host's recombination system would be very helpful to increase recombination efficiency. CRISPR must be carefully adapted to the physiology of the host chassis. Although the number of CRISPR-edited bacteria have increased significantly in the last 3 years, protocols and vectors cannot always be readily transferred from one microorganism to another, without extensive adaptation, that sometimes proves quite difficult (Cho *et al.*, 2017; Jiang *et al.*, 2017). Moreover, it is important to use CRISPR as a selective tool rather than an improvement route for recombination (Oh and van Pijkeren, 2014; Penewit *et al.*, 2018). Strains with high recombination efficiencies, like *S. aureus*

(Arnaud, Chastanet and Débarbouillé, 2004), and cyanobacteria (Boyle *et al.*, 2013), have been successfully adapted for CRISPR editing (Wendt *et al.*, 2016; Penewit *et al.*, 2018). The first publication of CRISPR-editing in *C. necator*, suggests that CRISPR/Cas9 can be utilised in *Cupriavidus* species for HTP engineering and highlights the need for more research towards the development of CRISPR systems in these chassis.

The next short-term goal regarding the utilisation of the CRISPR/Cas9 technique *in C. necator* should focus on redesigning the CRMAGE system (**Chapter 3**) and adding the riboswitch-controlled Cas9 vector. Indeed, the co-expression of Cas9 and λ-Red proteins demonstrated precise and efficient genome editing and multiplexgene deletion in *Pseudomonas aeruginosa* PAO1 (Chen *et al.*, 2018). The system could also be tested in the Δ*recJ* strain, which has better transformation efficiencies than the wild type strain. Once an optimised version of the 'CRMAGE' system is fully functional in *C. necator*, it would allow adaption of the method for multiplex editing at HTP scale, with automated successive rounds of adaptation/editing and production of mutant libraries. For example, the in-house CRMAGE system could be utilised to delete the genes involved in the RM systems of the strain (H16_A0006-9, H16_A1814 and PHG170 (H16_A0008 and H16_A0009)) and compare the resulting editing efficiencies with those from the work of Xiong *et al* (Xiong *et al.*, 2018). For CRISPR to replace the cu*rrent sacB* method (Lenz, Lauterbach and Frielingsdorf, 2018), the system must be optimised to reach higher efficiencies (over 50%) and the whole process (from transformation to mutant screening) should ideally not exceed one week. Furthermore, the CRMAGE system could be applied in other *Cupriavidus* species, especially the multi-metal resistant strain *Cupriavidus metallidurans*.

A CRISPRi-based mutation library could also be developed, fine-tuning the expression of genes involved in selected metabolic pathways whilst preserving the integrity of the genome (Larson *et al.*, 2013; Yao *et al.*,

2016; Beuter *et al.*, 2018). Other CRISPR nucleases like Cpf1 or Cas13a could prove more efficient for large fragment deletions (Zetsche *et al.*, 2015) and could prove to be less toxic than Cas9 (Ungerer and Pakrasi, 2016; Li *et al.*, 2018). A CRISPR toolbox for *Cupriavidus* species would be of great interest for a broader use in the engineering of many other metabolic pathways. Ultimately, deciphering the molecular mechanisms and cell physiology in CRISPR-edited cells would enable a better understanding of key recombination pathways in *Cupriavidus* spp.

Engineering improved ethylene production in *Cupriavidus metallidurans*

5.1 Introduction

Ethylene can be synthesised via three biological pathways as detailed in **Chapter 1**, however the ethylene-forming enzyme (EFE) has garnered the most interest, particularly from a biotech perspective (Eckert *et al.*, 2014). The heterologous expression of just one enzyme has been implemented in many organisms.

The enzyme has been successfully introduced into *Escherichia coli* (Lynch *et al.*, 2016), *Saccharomyces cerevisiae*, *Pseudomonas putida*, *Trichoderma viride, Trichoderma reesei*, cyanobacteria and tobacco (**Table 5.1**). Ethylene production in cyanobacteria showed promising results and might pave the way for industrial bioproduction of ethylene from CO_2 and sunlight (Guerrero *et al.*, 2012; Zavřel *et al.*, 2016; Puthan Veetil, Angermayr and Hellingwerf, 2017).

Table 0.1 – Ethylene production rates in the native Kudzu strain and in engineered microbial strains

Organism	Production rate (µmol/gDCW/h)	Growth medium	Reference
Pseudomonas syringae (Kudzu)	312.0	LB + 0.5% glucose	(Ishihara *et al.*, 1996)
E. coli	30.0	M9 + 1% glucose	(Lynch *et al.*, 2016)
S. cerevisiae	1151.5	CBS + 1% glucose + glutamate	(Zhang, Wei and Ye, 2013)
Synechocystis spp. PCC 6803	111.5	BG-11	(Puthan Veetil, Angermayr and Hellingwerf, 2017)
Synechococcus elongatus PCC 7942	84.8	BG-11	(Sakai *et al.*, 1997)
Pseudomonas putida	2859.2	LB	(Wang *et al.*, 2010)
Trichoderma viride	0.093	MM + 2% cellulose + 0.2% peptone	(Tao *et al.*, 2008)
Trichoderma reesei	0.716	MM + 2% wheat straw	(Chen *et al.*, 2010)

DCW: dry cell weight

The archetype EFE(P) was isolated from *Pseudomonas syringae* pv. *phaseolicola* PK2 (Fukuda *et al.*, 1992). The EFE(P) enzyme belongs to the mononuclear nonheme Fe (II)- and 2-oxoglutarate (2OG)-dependent oxygenase superfamily (Martinez and Hausinger, 2015; Zhang *et al.*, 2017a). Like all members of this family, the EFE(P) enzyme catalyses the oxidative decarboxylation of 2OG or alpha-ketoglutarate (AKG) associated with the hydroxylation of L-arginine into succinate, guanidine and Δ-1-pyrroline-5-carboxylate (P5C) (**reaction 1**) (Proshlyakov, McCracken and Hausinger, 2017). But EFE(P) has a unique dual activity and also converts AKG into 3 molecules of CO_2 and ethylene (**reaction 2**) (Martinez and Hausinger, 2016).

(1) L-Arginine + alpha-ketoglutarate (AKG) → Succinate + Δ-1-pyrroline-5-carboxylate (P5C) + Guanidine+ CO_2

(2) Alpha-ketoglutarate (AKG) → Ethylene + 3 CO_2

Recent biochemical and spectroscopic studies have brought insights into the catalytic mechanism of the enzyme and have established a crystal structure, with/without substrate and cofactors bound (**Figure 5.1**) (Martinez and Hausinger, 2016; Martinez *et al.*, 2017; Zhang *et al.*, 2017b).

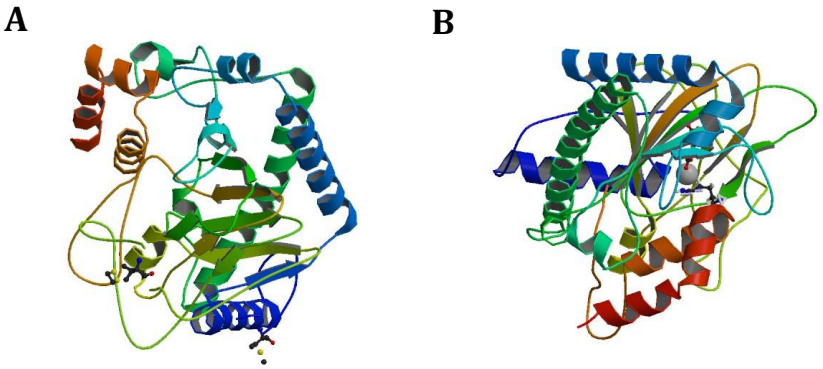


Figure 0.1 – 3D structure of the Ethylene Forming Enzyme (EFE)

Structure of EFE in apo form (**A**) or in complex with ferrous iron, alpha-ketoglutarate and L-arginine (**B**) (Martinez *et al*, 2017).

The EFE enzyme is characterised as an Fe(II)/2OG-dependent oxygenase, because it contains a typical double-stranded β-helix (DSBH) core (Aik *et al.*, 2015) formed of nine β-strands, six from the major β-sheet and three from the minor β-sheet, stabilised by 10 α-helices (Martinez and Hausinger, 2016). Investigation of ethylene production in 757 bacterial strains resulted in the identification of 225 methionine dependent strains and the unique 2OG-dependent ethylene-producing strain, *Pseudomonas syringae* pv. *phaseolicola* PK2, which exhibited the highest ethylene yield, demonstrating the EFE pathway is the most efficient biological route for ethylene production (Nagahama *et al.*, 1992).

Both *C. metallidurans* CH34 and *C. necator* H16 can grow on CO_2 and H_2, making them attractive chassis for sustainable ethylene production from waste gases. To fully engineer improved ethylene production in *Cupriavidus,* HTP genomic engineering tools were developed as detailed in **Chapters 3 and** 4.

The EFE pathway had already been implemented in *C. necator* H16 at the SBRC Nottingham (Alexander Van Hagen, personal communication). Heterologous constitutive expression of the EFE(P) enzyme in the wild-type H16 strain enabled impressive ethylene productivities (up to 327 nmol/OD/mL in Fructose Glycerol Nitrogen [FGN] minimal medium, which is 1.7-fold higher than the highest productivity reported in *E. coli* (Lynch *et al.*, 2016)) and engineering of the enzyme expression as well as metabolic engineering of the TCA cycle, as demonstrated in *E. coli* would undoubtedly improve further ethylene yields in this strain. Extrapolating a similar phenotype in other strains of the *Cupriavidus* genus, **Chapter 4** explores implementation of the EFE biosynthesis pathway in another chassis of the SBRC Nottingham, *Cupriavidus metallidurans*.

The *efe* gene from *Pseudomonas syringae* pv. *phaseolicola (efep)* was successfully cloned and introduced into *C. metallidurans* and ethylene production was detected by gas chromatography (GC). Several growth and

expression parameters were investigated to improve ethylene yield and productivity, such as optimisation of media composition, nutrient supplementation and the development of a plasmid addiction system. The last part of this chapter introduces new HTP tools like global Transcription Machinery Engineering (gTME) and Adaptive Laboratory Evolution (ALE) to circumvent the need for multiple gene knockouts in *C. metallidurans.* Both these methods were implemented in *E. coli* initially to test their robustness in a model strain. Though the main goal was to develop these techniques in *C. metallidurans*.

5.2 Optimisation of the expression of EFE in *C. metallidurans*

5.2.1 Expression of EFE in C. metallidurans

The expression plasmid pEFE(P) was assembled utilising the broad-host-range vector pBBR1, which can replicate in both H16 and CH34. Constitutive expression of the *efe* gene, was driven by the promoter j5 (*efep*, **Figure 5.2**). The j5 promoter is derived from the bacteriophage T5 (Gentz and Bujard, 1985) and has already been successfully utilised in *C. necator* H16 (Gruber *et al.*, 2014), making it a good candidate for high level *efep* expression in CH34.

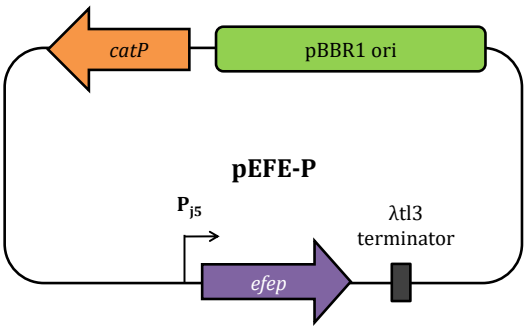


Figure 0.2 – Schematic representation of the plasmid pEFE(P)EFE(P) constitutively expressing the ethylene-forming-enzyme from P. syringae pv. phaseolicola

catP: chloramphenicol acetyltransferase; P_{j5}: promoter from bacteriophage T5; *efep*: ethylene-forming-enzyme from *P. syringae* pv. *phaseolicola*; λtl3: transcription terminator tL3 from phage λ.

The plasmid pEFE(P)EFE(P) was electroporated into CH34 and cells were selected on LB agar plates supplemented with 250 μg/mL chloramphenicol (Cm 250). A single colony of CH34/pEFE(P)EFE(P) was inoculated overnight in either 10 mL LB, 0.4% (w/v) Sodium Gluconate Minimal Salts Medium (0.4% SG-MSM) or 0.4% (w/v) Sodium Gluconate Minimum Medium (0.4% SG-MM) supplemented with 250 μg/mL chloramphenicol. Both rich and minimum media were used to evaluate ethylene production in CH34.

EFE(P) protein expression was assessed in all three of the media tested (**Figure 5.3**). Although some protein was detected in the soluble fraction, a large amount remained in the insoluble fraction, confirming that EFE(P) solubility is a limiting factor in ethylene production (Lynch *et al.*, 2016). Equal amounts of protein were loaded on the gel and slightly higher levels of soluble EFE(P) protein were observed in the LB and 0.4% SG-MSM media. Given that minimum medium has proved more beneficial for ethylene production (Zhang, Wei and Ye, 2013), 0.4% SG-MSM was selected for EFE(P)-catalysed ethylene biosynthesis in CH34.

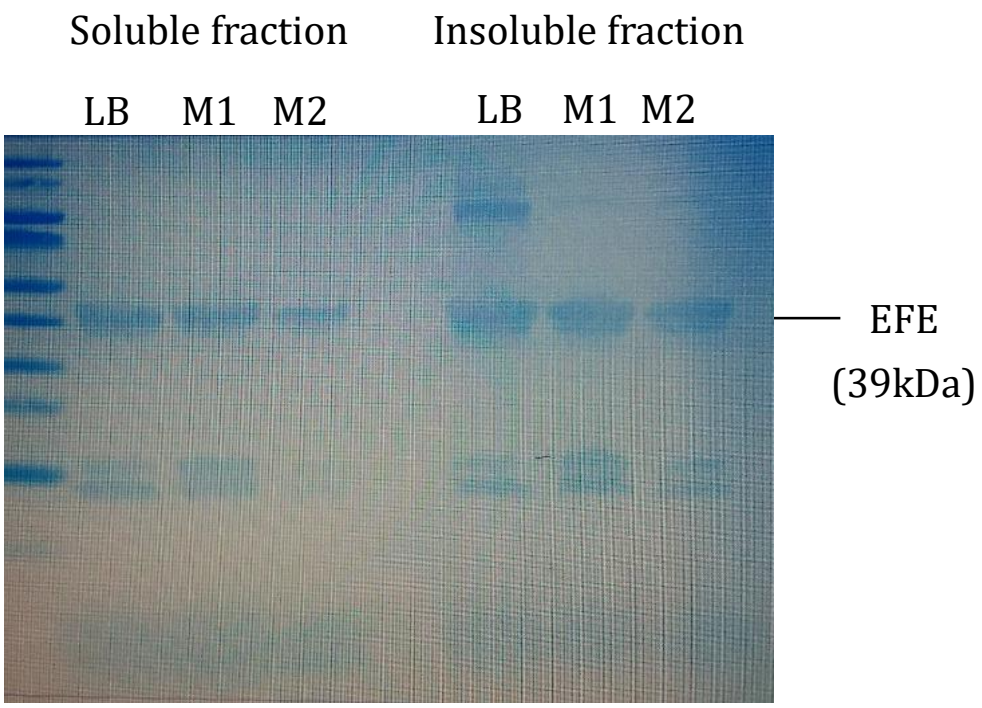


Figure 0.3 – Expression of EFE(P) in CH34 cultured in LB, 0.4% SG-MSM and 0.4% SG-MM

Anti-EFE(P) Western blot for EFE(P) expression in CH34. Cells were grown for 24 h at 30°C in LB, 0.4% SG-MSM (M1) and 0.4% SG-MM (M2) supplemented with 250 µg/mL chloramphenicol. Size was evaluated using the ThemoFischer® PageRuler™ Prestained Protein Ladder.

Ethylene production was monitored by gas chromatography (GC). CH34/pEFE(P) was adapted in 0.4% SG-MSM supplemented with 250 µg/mL chloramphenicol over 48 h at 30°C, then sub-cultured and analysed at appropriate time points to identify the optimal sampling point for peak ethylene production.

In a previous study (Lynch *et al.*, 2016), a standard method for ethylene monitoring identified maximal ethylene production when cell OD_{600} was 0.2-0.3, regardless of the media used. To validate a standard method of ethylene measurement in CH34, ethylene and growth were monitored at frequent intervals over a 96 h period (**Figure 5.4A-B**). Exponential phase started at 4 h and the cells reached late exponential phase at 12 h in the CH34 culture (**Figure 5.4C**), no more than 25% of the maximum yield (obtained after 96 h) could be detected at the end of this

12 h period (**Figure 5.4A** and **D**). However, 62 ± 6.2 nmol of ethylene were measured after 24 h, which was 84% of the maximum ethylene yield (**Figure 5.4D**). According to the Lynch study (Lynch *et al.*, 2016), ethylene measurements were taken at the point where maximal ethylene productivity was reached. In the present study, measurements should then be taken after 72 h (**Figure 5.4D**). However, error bars were important at this point. They were significantly shorter at 24 h and the ethylene productivity at this time point was ~60% of the maximal ethylene productivity. In addition, cells had entered stationary phase and O_2 availability was potentially at a rate-limiting level. Indeed, the O_2 level was depleted in the exponential phase in the Lynch study (Lynch *et al.*, 2016). Therefore, the sampling time point to measure ethylene was determined at 24 h after inoculation.

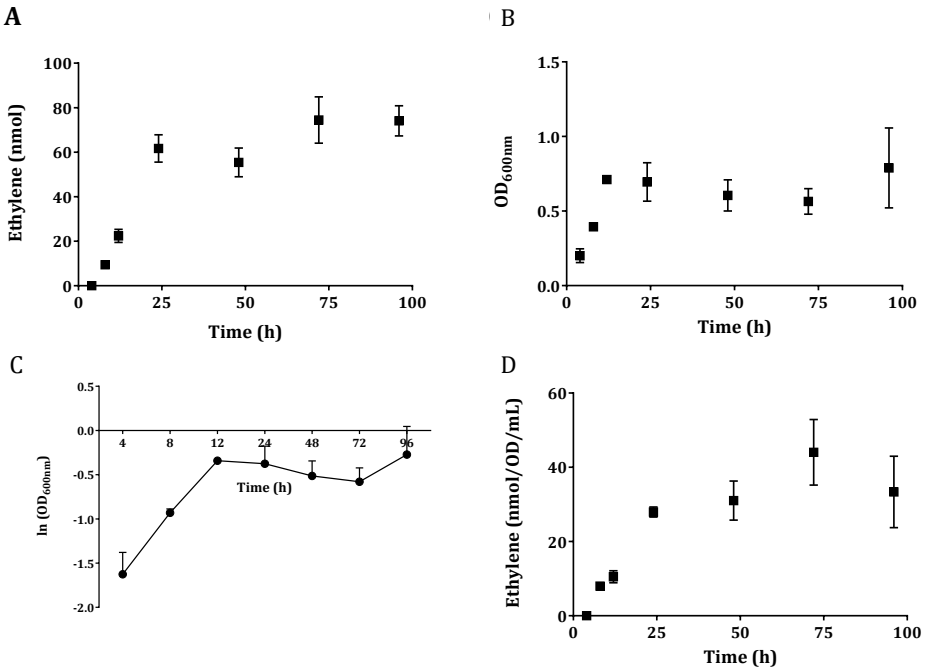


Figure 0.4 – Ethylene yields and OD values of CH34/pEFE(P) grown in 0.4% (w/v) SG-MSM over a 96 h-period

(**A**) Ethylene production, (**B**) ethylene productivity, (**C**) growth course and (**D**) logarithmic (ln) OD values of CH34/pEFE(P) cultivated over 96 h in 0.4% (w/v) SG-MSM medium.

5.2.2 Effect of carbon source on ethylene production

CH34 can grow on a wide variety of different carbon sources including gluconate and succinate (Nies *et al.*, 1987; Rojas *et al.*, 2011), which can impact on growth rate and ethylene production, through cellular metabolism. However, CH34 unlike H16, is not able to grow on glucose, fructose, or galactose, because it lacks adequate glucose uptake systems and does not have a functional 6-phosphofructokinase (Mergeay *et al.*, 1985). The following carbon sources were tested in MSM medium for the growth of both CH34 and H16 expressing *efep*: sodium acetate 5 mM, sodium benzoate 5 mM, sodium gluconate 0.4% (w/v), sodium lactate 5 mM and sodium succinate 0.3%(w/v) (**Table 5.2**).

Table 0.2 – Correspondence mass/volume ratio to molar concentration for the five carbon sources

Carbon source	**% (w/v)**	**mM**
Acetate	0.03	**5**
Benzoate	0.3	**5**
Lactate	0.045	**5**
Gluconate	**0.4**	20
Succinate	**0.3**	25

Mass/volume ratios or molar concentrations used in this study are highlighted in bold.

The most suitable carbon source should allow for high ethylene production without impeding cell growth. **Figure 5.5A** clearly shows that gluconate increases ethylene production (61 ± 4.9 nmol), followed by acetate (39 ± 0.89 nmol) and succinate (33 ± 7.2 nmol). The cell growth yield was 2-3-fold higher in gluconate and succinate than in acetate, benzoate and lactate (**Figure 5.5B**). The greatest ethylene productivities were obtained in acetate, lactate and gluconate (**Figure 5.5C**). Acetate enabled the best rates of ethylene productivity and did not impact on the cell growth as much as lactate. Ethylene productivity in gluconate was very similar to that on acetate, but the use of gluconate allowed cells to enter

stationery phase more quickly than acetate. It has been shown that growth on acetate as a sole carbon source up-regulates genes involved in the TCA cycle, which reach maximum levels in expontential phase (Wolfe, 2005; Han, Inui and Yukawa, 2008). To obtain a high ethylene productivity without hampering growth, acetate was therefore selected as the best carbon source for ethylene production.

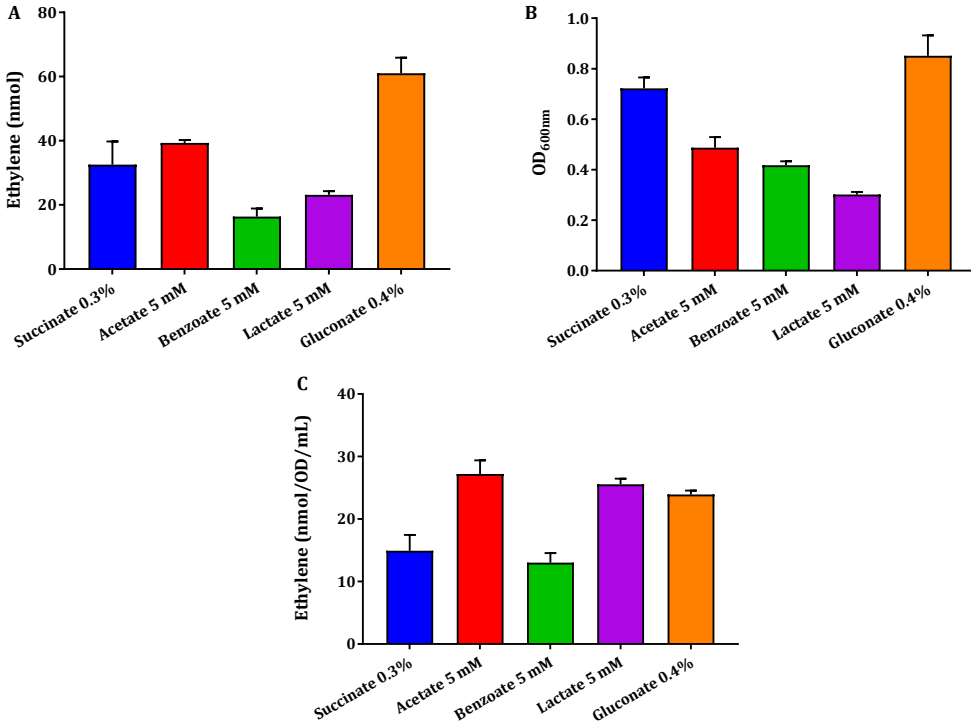


Figure 0.5 – Effect of carbon source on production of ethylene in CH34/pEFE(P)EFE(P)

(**A**) Ethylene production, (**B**) OD_{600} and (**C**) normalised ethylene productivity from a culture of CH34/pEFE(P)EFE(P) grown in MSM medium supplemented with 0.3% (w/v) succinate, 5 mM acetate, 5 mM benzonate, 5 mM lactate or 0.4% (w/v) gluconate.

Ethylene production and growth yield were assessed in incremental concentrations of acetate (5, 10, 15 and 20 mM). Cells grown in 20 mM acetate-MSM did not grow, suggesting that 20 mM acetate could be toxic. The highest amount of ethylene was detected in cultures containing 10 mM acetate (**Figure 5.6A**) whilst 15 mM acetate provided the best growth yield (OD_{600} = 0.85 ± 0.08), which suggests that while excessive flux of acetate enhances growth (**Figure 5.6B**), the flux is not redirected through ethylene

formation. Nonetheless, a 1.2-fold improvement in productivity was observed in 10 mM acetate, compared to ethylene productivity in 5 mM acetate (**Figure 5.6C**). Further optimisation steps were conducted in MSM medium supplemented with 10 mM acetate.

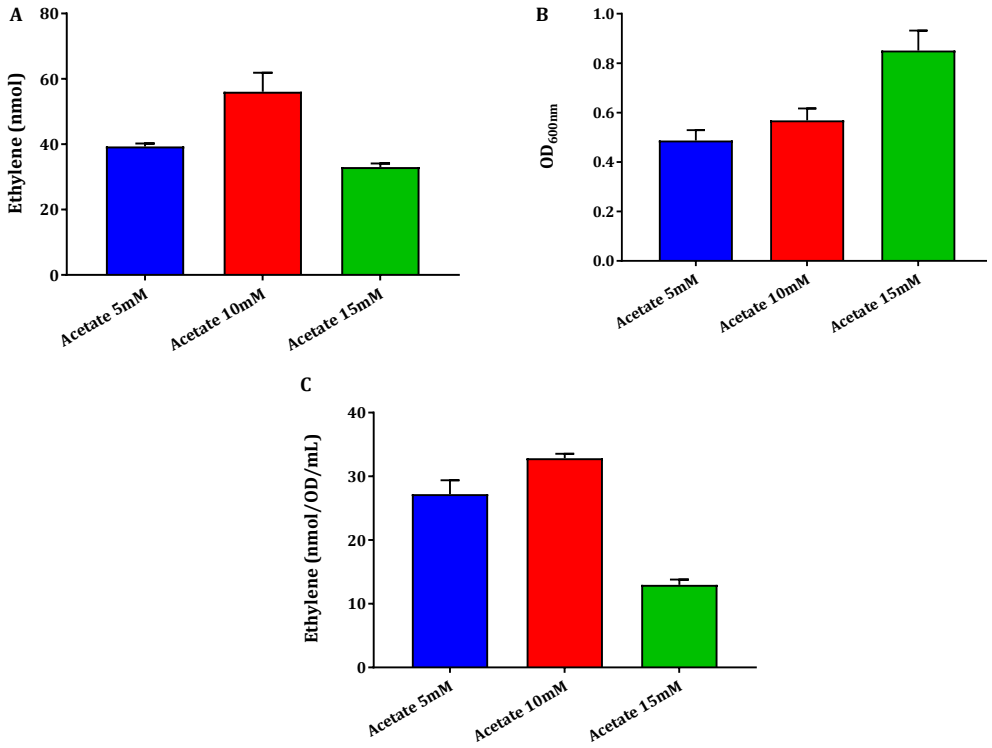


Figure 0.6 – Effect of acetate concentration on the production of ethylene in CH34/pEFE(P)EFE(P)

(**A**) Ethylene production, (**B**) OD_{600} and (**C**) normalised ethylene productivity from a culture of CH34/pEFE(P)EFE(P) grown in MSM medium supplemented with 5, 10 or 15 mM acetate and 250 µg/mL chloramphenicol.

Given the similarity between CH34 and H16, comparable ethylene productivities were expected. Therefore, the effect of the carbon source on ethylene productivity was evaluated in both strains. H16 was also cultivated in FGN (fructose/glycerol/nitrogen) minimum medium and 0.4% (w/v) glycerol, since H16 growth was reported in these carbon sources (Jugder *et al.*, 2015). Surprisingly, the ethylene productivity of H16 was between 3.8- (in benzoate) and 20-fold (in succinate) higher than that of CH34 (**Figure 5.7A-B**). The highest ethylene productivity of wild-type H16

is observed in FGN and reaches 327 nmol/OD/mL, which is almost two fold higher than that of the best recombinant *E. coli* strain (Lynch *et al.*, 2016), suggesting that strain engineering of H16 could generate very high levels of ethylene. Although both H16 and CH34 can both grow on CO_2, CH34 has a much faster growth rate on CO_2 (Prof Alex Conradie, personal communication) and is also able to survive in media with millimolar concentrations of heavy metals (Monchy *et al.*, 2007). Therefore, CH34 could be engineered to generate ethylene from contaminated waste fermentation.

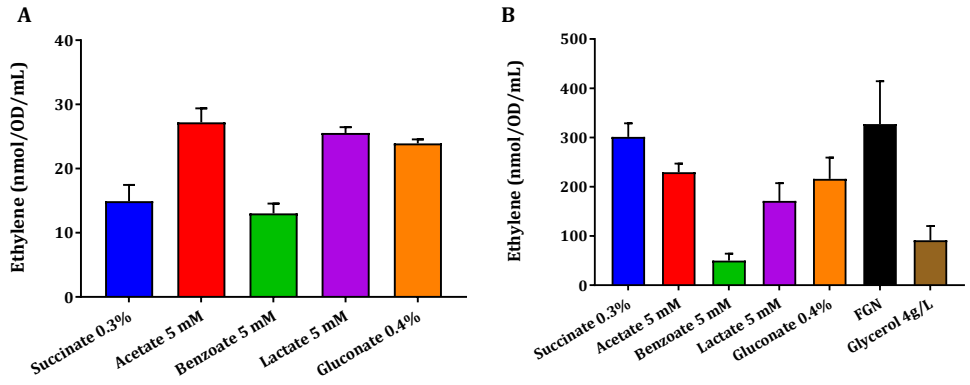


Figure 0.7 - Comparison of the effect of carbon source on ethylene productivity in C. metallidurans CH34 and C. necator H16

Ethylene productivity in (**A**) CH34/pEFE(P)EFE(P) and (**B**) H16/pEFE(P)EFE(P) grown in MSM medium supplemented with 0.3% succinate, 5 mM acetate, 5 mM benzonate, 5 mM lactate, 0.4% (w/v) gluconate and either 0.4% (w/v) glycerol or fructose/glycerol/nitrogen (FGN) and 50 µg/mL (H16) or 250 µg/mL chloramphenicol (CH34).

5.2.3 Effect of nutrient supplementation

Substrate availability can often be a rate-limiting factor in an enzymatic reaction. The effect of EFE substrate supplementation on ethylene production was assessed in CH34/pEFE(P).

Addition of 2 mM AKG and 3 mM L-arginine resulted in a two-fold increase in ethylene production (112 ± 8.41 nmol) and productivity (63 ±

3.1 nmol/OD_{600}/mL) compared to the control culture with no supplement (ethylene production: 56 ± 5.7 nmol; ethylene productivity: 33 ± 0.71 nmol/OD_{600}/mL) (**Figure 5.8A and C**). The supplementation did not affect cell growth except for cells cultivated in 20 mM AKG + 20 mM arginine, which hindered the growth yield by ~30% (**Figure 5.8B**). It is possible that high levels of AKG may affect the expression of key metabolic pathways for nitrogen and carbon (Huergo and Dixon, 2015).

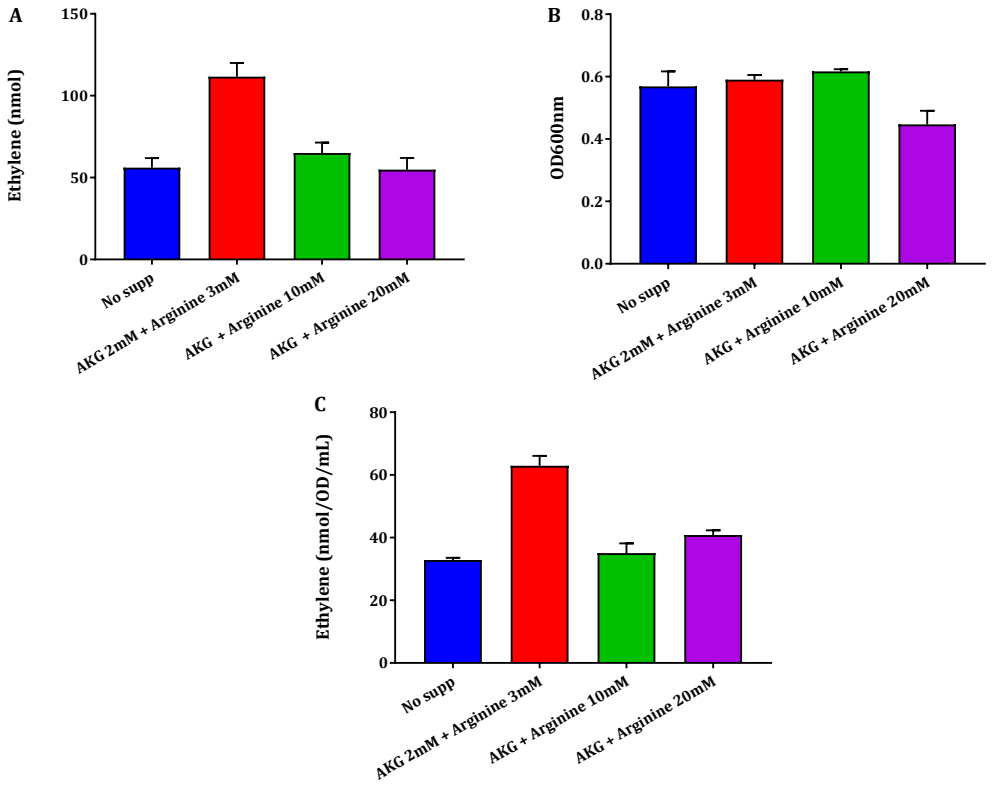


Figure 0.8 – Effect of nutrient supplementation on ethylene production in CH34/pEFE(P)EFE(P)

(**A**) Ethylene production, (**B**) OD_{600} and (**C**) normalised ethylene productivity from a culture of CH34/pEFE(P)EFE(P) grown in MSM medium supplemented with 10 mM acetate, 250 µg/mL chloramphenicol and increasing concentrations of AKG and L-arginine. Cells were grown over 24-48 h at 30°C in MSM medium supplemented with 10 mM acetate and 250 µg/mL chloramphenicol then they were sub-cultured in 10-mL serum bottles to a final volume of 3 mL (starting OD = 0.08), including the appropriate concentration of substrates, and samples were processed as described above. Measures were done in triplicate after 24 h at 30°C, 200 rpm.

These results showed a small increase in ethylene production and productivity in cultures supplemented with 2 mM AKG and 3 mM L-arginine, which confirmed that substrate could be a rate-limiting factor in ethylene

production. This is supported by similar observations in *E. coli* where the greatest improvement in production was observed in culture supplemented with AKG and L-arginine while addition of glutamate, glutamine or proline did not improve ethylene synthesis (Lynch *et al.*, 2016). However, substrate availability is not the only major limiting factor, EFE(P) solubility is also a significant problem, since no significant improvement was observed in the case of 10 and 20 mM AKG and L-arginine addition.

5.2.4 Implementation of a plasmid addiction system

To produce ethylene in a continuous fermentation process, it is necessary to improve the stability of the pEFE(P) plasmid. Indeed, plasmids are often lost in a significant proportion of *Cupriavidus* cells in continuous culture (Prof Alex Conradie, personal communication). To maintain pEFE(P) in CH34 and establish an antibiotic-free ethylene-producing strain of CH34, the Hok/Sok toxin/antitoxin system was added to the pEFE(P) plasmid (**Figure 5.9**). Hok is a host killing toxin protein responsible for depolarisation of the host membrane (Gerdes *et al.*, 1986). Its formation can be inhibited via the small non-coding RNA transcript *sok* (suppressor of killing) (Gerdes, Thisted and Martinussen, 1990). More precisely, *sok* anneals to the *mok* (modulator of killing) mRNA, which regulates *hok* transcription. The RNA duplex is then rapidly degraded by RNAse III and the cell can survive. As *hok* has a longer half-life (approximately 20 min) than *sok* (no more than 30 s), the addiction system is maintained by the cell and, consequently, the expressing plasmid is passed along to the daughter cells.

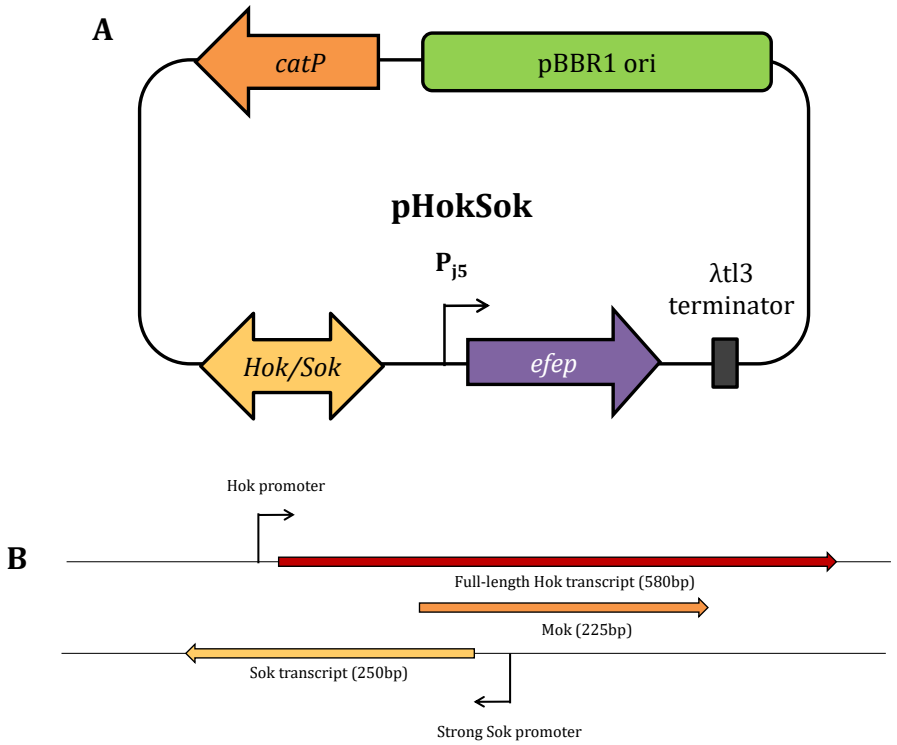


Figure 0.9 – The Hok/Sok system

A. Schematic representation of the pHokSok plasmid.
B. Details of the sequence and features of the Hok/Sok system.
catP: chloramphenicol acetyltransferase; P_{j5}: promoter from bacteriophage T5; *efep*: ethylene-forming-enzyme (*P.syringae pv.phaseolicola*); λtl3: transcription terminator tL3 from phage λ; *hok*: host killing gene; *sok*: suppressor of killing gene; Mok: modulator of killing.

The effect of the Hok/Sok system on plasmid stability was assessed in CH34, utilising the *efep*-expressing plasmids pEFE(P) and pHokSok. The plasmids were compared in 50mL-tubes and continuous fermentation.

After 96 h of growth in 50-mL tubes without any selection pressure, 73% of the CH34 population maintained the pEFE(P) plasmid (**Figure 5.10A**). Over the same period, retention of pHokSok remained maximal (present in 99% of the population). Since ethylene biosynthesis would involve a fermentation process, if transferred to an industrial scale, it was prudent to assess such a stabilisation strategy from the early stages of the process development. Thus, maintenance of both pEFE(P) and

pHokSok was evaluated over 12 days in continuous culture in 1 L bioreactors (750 mL working volume) fed with a gas mixture (78% H, 3% CO_2 and 19% air) and DSMZ 81 medium supplemented with 0.8% (w/v) glycerol. Samples were collected every 24 h, enabling plasmid retention to be assessed in each culture. Interestingly, the stability ratio of pEFE(P) and pHokSok after 96 h was approximately two-fold lower in fermentation than in the tube cultures (**Figure 5.10B**). The high dilution rate in fermentation forced the cells to divide at a constant rate. The low OD values reported in the first hundred hours of fermentation might result from the energetic burden required to replicate the plasmids pEFE(P) and pHokSok (**Figure 5.10C**). After 120 h of culture, 90% of the cells had lost the plasmids and the cell growth could be rescued. Both plasmids were completely lost after 192 h of culture (8 days), when cells entered the stationery phase. There was no significant difference in survival rate and growth between CH34/pEFE(P) and CH34/pHokSok cultures, suggesting the Hok/Sok system was ineffective in fermentation culture.

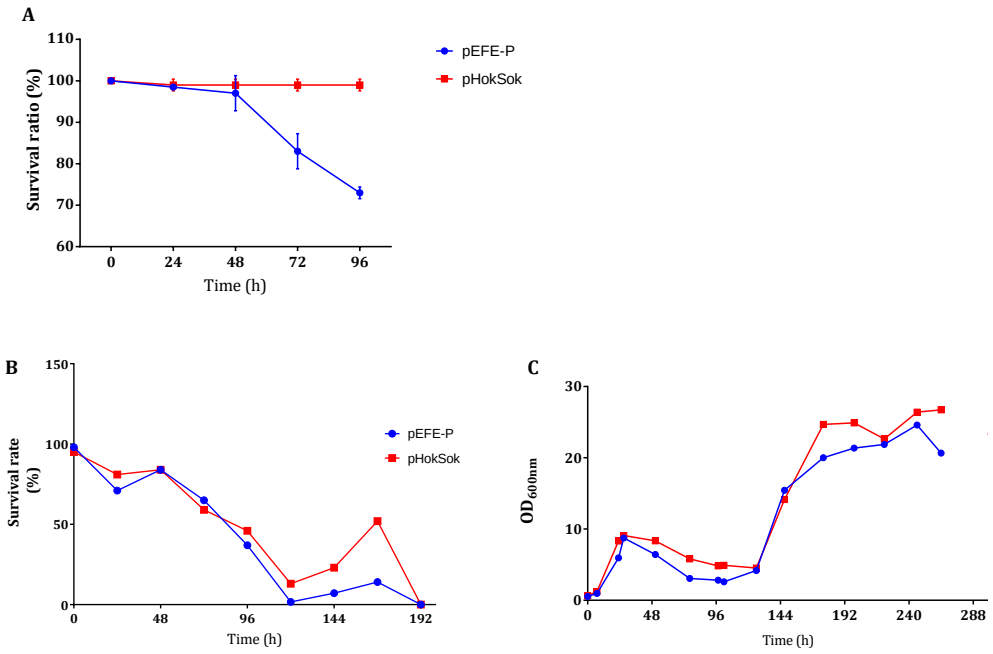


Figure 0.10 – Effect of the Hok/Sok addictive system on stability of the efep-expressing vector pHokSok in CH34

A. Stability or survival ratio of CH34 grown in 50-mL tubes and harbouring pEFE(P) (blue) or pHokSok (red) in the absence of selection pressure over ca. 50 generations. A single colony was incubated overnight in 10 mL LB supplemented with 250 µg/mL chloramphenicol (Cm 250). 100 µL of the culture inoculated 10 mL LB with no antibiotic. 100 µL samples were collected at regular time points, diluted appropriately and spread onto selective (for evaluation of plasmid segregational stability) and non-selective plates (for viable counts).
B. Stability or survival rate of CH34 grown in continuous fermentation over 12 days and harbouring pEFE(P)EFE(P) (blue) or pHokSok (red) in the absence of selection pressure.
C. Growth of CH34 grown in continuous fermentation over 12 days and harbouring pEFE(P)EFE(P) (blue) or pHokSok (red) in the absence of selection pressure.

In summary, the addictive Hok/Sok system did improve retention of the *efep*-expressing plasmids in tube cultures but it did not help maintain the plasmids in continuous fermentation. Therefore, alternative systems must be developed to guarantee a constant expression of EFE(P) along the whole bioproduction process. For example, the hydrogenase transcription factor (*hoxA;* (Lütte *et al.*, 2012)), 2-keto-3-desoxy-6-phosphogluconate-aldolase (KDPG-aldolase; *eda;* (Voss and Steinbüchel, 2006)), xylulose-5-phosphate/fructose-6-phosphate phosphoketolase (*xfp;* (Fleige, Kroll and Steinbüchel, 2011)) and pyrroline-5-carboxylate reductase (*proC;* (Budde *et al.*, 2011)) have been used to complement deletion H16 mutants and

rescue growth in these strains. Heterologous CH34 genes (*hoxA*: Rmet_1542, *eda*: Rmet_4768 and *proC*: Rmet_2938) could potentially be knocked-out in *C. metallidurans* and heterologous expression of these genes on a plasmid backbone could enable the development of a robust plasmid addiction system in CH34.

5.3 Implementation of a gTME strategy

5.3.1 Adaptation of a product-growth coupling in E. coli MG1655

Product-growth coupling is a key principle in metabolic engineering, which makes production of the desired metabolite obligatory for microbial growth (von Kamp and Klamt, 2017) . This approach has two main advantages: (1) it turns growth into a production driver and (2) it helps stabilise metabolite production (Alter, Blank and Ebert, 2018). Many computational tools, like the pioneering Optknock published in 2003, have been developed to identify gene deletions that would lead to overproduction of biochemicals (Burgard, Pharkya and Maranas, 2003). These algorithms suggest straightforward deletion of outcompeting pathways, but they also find nonintuitive reactions that could contribute to a better yield. Growth coupling has already been applied many times in *E. coli* and *S. cerevisiae* to optimise the production of lactate, ethanol or 2,3-butanediol from glycerol, itaconic acid or glucose (Trinh, Unrean and Srienc, 2008; Trinh and Srienc, 2009; Harder, Bettenbrock and Klamt, 2016).

Genome-scale model (GSM) and flux balance analysis (FBA) were utilised to establish a growth couple through P5C to proline in MG1655, expressing EFE(P), to maximise ethylene yield (manuscript in preparation). This work was initially performed in *E. coli,* the model bacteria, and ideal host platform to test the growth couple hypothesis, which can then be implemented in *Cupriavidus* to optimise ethylene production.

Systems biology analyses resulted in the identification of a target gene to connect cell growth to ethylene production in *E. coli*: the *proB* gene, coding for glutamate 5-kinase, part of the proline biosynthesis pathway (**Figure 5.11**). A Δ*proB* proline-auxotroph mutant was generated (Alexander Van Hagen). The heterologous expression of the *efep* gene on a pGEM vector (pGEM-*efep*), complemented the auxotrophy in M9 medium

without the supplementation of proline, via the co-production of the intermediate **Δ**-1-pyrroline-5-carboxylate (P5C). This established the first growth coupling of ethylene formation in *E. coli*. Successive passages in fresh M9 liquid medium were necessary to allow the auxotrophic strain to adapt and reach a growth rate similar to the wild-type strain, harbouring pGEM/*efep* (μ_{max} ~0.32 h^{-1}). In addition, ethylene productivity increased significantly in the Δ*proB* mutant with the pGEM-*efep* plasmid (120 nmol/OD/mL) compared to the wild-type strain (70 nmol/OD/mL).

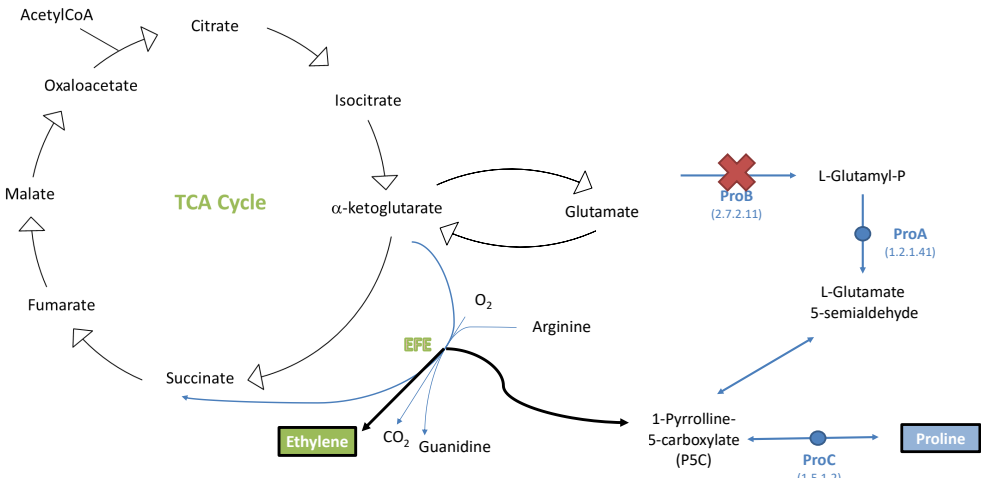


Figure 0.11 – Growth coupling of the activity of EFE with proline formation

Putative metabolic scheme for ethylene production in *E. coli* via the Ethylene-Forming Enzyme (EFE) coupled to proline synthesis via **Δ**-1-pyrroline-5-carboxylate (P5C) after deletion of the *proB* gene.
ProA: Gamma-glutamyl phosphate reductase; ProB: Glutamate 5-kinase and ProC: **Δ**-1-pyrroline-5-carboxylate (P5C) reductase.

5.3.2 Application of the gTME selection process to enhance ethylene production

Global Transcription Machinery Engineering or gTME was developed in the Stephanopoulos lab by Dr Hal Alper in 2006 (Alper and Stephanopoulos, 2007). The technique aims to redirect flux through transcriptional changes by generating a library of *rpoD* mutants. RpoD is the

sigma 70 factor of the RNA polymerase and controls promoter recognition, of most genes under normal growth conditions (Kumar *et al.*, 1994). Altering this factor contributes to the global deregulation of the cell transcriptome, allowing for the quick generation of phenotypical mutants and outcompetes traditional single-gene modification procedures (Tan *et al.*, 2016). As an example, Alper and colleagues used gTME to select for improved lycopene production in *E. coli* (Alper and Stephanopoulos, 2007). After one round of gTME, they obtained a mutant strain producing as much lycopene as a mutant generated by three successive single-genetic perturbations. In the present study, the gTME selection technique was assessed in *E. coli* MG1655 Δ*proB* to elicit improved productivity of ethylene and as a proof of principle to be applied in *Cupriavidus* species.

A plasmid expressing the *rpoD* gene was assembled using a shuttle p15A-pVS1 replicon, transferrable to *E. coli* as well as to *C. metallidurans* CH34. The *rpoD* gene (b3067) and its upstream region were amplified by PCR from the chromosome of MG1655 with an error-prone DNA polymerase to introduce random mutations into the *rpoD* gene and its promoter and the amplicon was ligated into the linearised *prpoD* backbone. The resultant *rpoD* mutant plasmid library (size ca. $2x10^3$) was transformed into the Δ*proB* strain, harbouring the vector pGEM-*efep*, and the best ethylene-producers were selected either based on cell growth or adaptive fermentation (**Figure 5.12**).

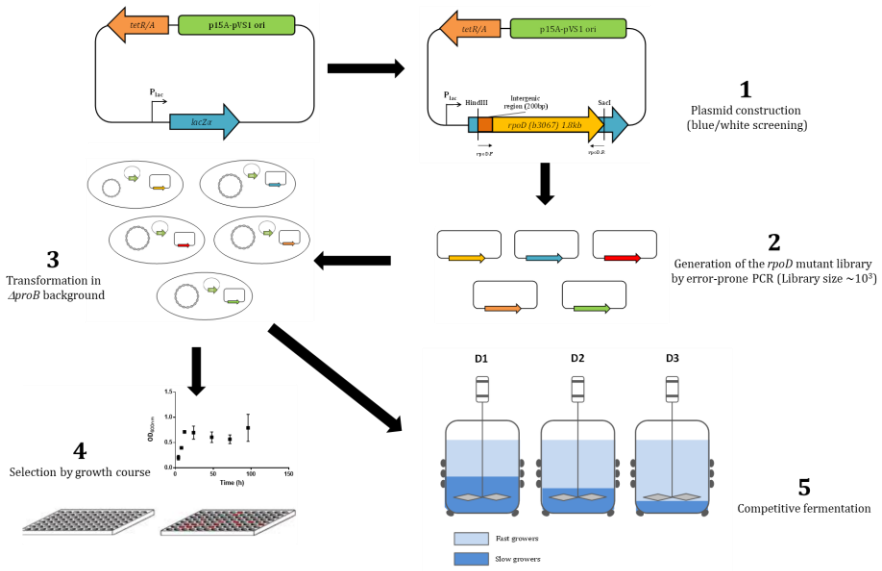


Figure 0.12 – Principle of the gTME applied to the improvement of ethylene production

A library of mutant plasmids, expressing *rpoD* randomly mutated by error-prone (ep-PCR), was transformed in the growth-coupled Δ*proB E. coli* background strain, which also hosts the *efep*-expressing pGEM-*efep* plasmid. All strains obtained were screened for enhanced growth (and improved ethylene productivity) by direct growth monitoring in microtiter plates or after competitive fermentation performed over 18 days at incrementally increasing dilution rates.

A total of 1921 transformants were selected and cultured in 300 µL 0.4% glucose M9 medium (to preserve the growth couple), supplemented with tetracycline (to maintain the p*rpoD* plasmid) in 96-well plates. Growth was monitored over 24 h and compared to that of a control strain, expressing the wild-type *rpoD* gene, on the same plasmid backbone. Strains were selected based on growth rate, mutants with a higher growth rate than the control strain, Δ*proB*/pGEM-*efep* + *prpoD*_wt, were selected for further analysis. The 95 fastest growers were sub-cultured in fresh medium with the control strain. From these candidates, the 8 fastest and slowest strains were further transferred into 10 mL serum bottles for GC analysis. A heterogeneous distribution in productivity was observed, with strains performing less, similarly or better than the control strain (**Figure 5.13**). Nonetheless, low growth yield-strains presented, on average, an ethylene productivity equal to that of the control strain with the unmutated *rpoD* (189 nmol/OD/mL), except strains H11 and G12 that yielded 34% and 32%

more ethylene than Δ*proB*/pGEM-*efep* + *prpoD*_wt. In contrast, 6 out of 8 high-growth yield-strains synthesised less ethylene than the control, but strains H1 and A2 reached 330 and 288 nmol/OD/mL, which represents an increase of 50-75% compared to the control productivity in Δ*proB*/pGEM-*efep*.

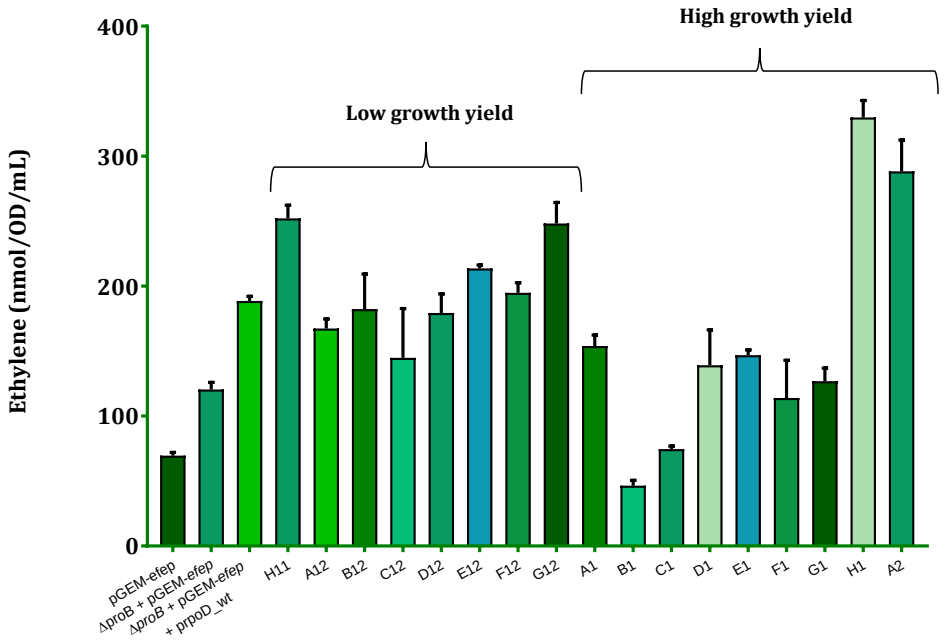


Figure 0.13 – Ethylene productivity in fast-growth isolates after a single round of gTME

H11-G12: Strains with a growth yield similar to the control (OD = 0.67); A1-A2: Strains with the best growth rate (0.8-0.86). Strains are named after the well position on the final selective 96-well plate. The control strain (Δ*proB* + pGEM-*efep* + p*rpoD*_wt) is in H12. Other controls include the pGEM-*efep* plasmid carried either in the wild-type MG1655 *E. coli* strain or in the proline-auxotroph Δ*proB* mutant.
Measures were done in triplicate after 24 h at 30°C, 200 rpm.

The results demonstrate that: (i) the expression of the *rpoD* gene contributed to a 3.7-fold improvement in ethylene productivity compared to the reference strain MG1655/pGEM-*efep*, (ii) the growth couple is not as robust as expected as higher ethylene yields were reported in both fast and slower growing strains. The growth couple biased P5C formation, demonstrating uncoupling of the dual circuit mechanism *in vivo*.

(1) L-Arginine + 2-oxoglutarate → Succinate + L-Δ-1-pyrroline-5-carboxylate (P5C) + Guanidine + CO_2

(2) 2-oxoglutarate → Ethylene + $3CO_2$

In vitro mutagenesis demonstrated that ethylene productivity could be abolished in EFE(P) mutants, however these mutants retained P5C formation (Martinez *et al.*, 2017). Development of an ethylene biosensor will allow higher ethylene producers to be selected directly, rather than via growth couple.

5.3.3 Mutant selection via Adaptive Laboratory Evolution (ALE)

Adaptive laboratory evolution (ALE) process (Portnoy, Bezdan and Zengler, 2011) can be utilised to optimise product formation. ALE was originally utilised as a strain development and optimisation tool (Chatterjee and Yuan, 2006) in industrial applications such as adaptation dynamics (Applebee, Herrgård and Palsson, 2008), population evolution (Tremblay *et al.*, 2011), productive phenotype optimisation (Fong *et al.*, 2005) and tolerance improvement (Portnoy, Herrgård and Palsson, 2008). In ALE experiments, a cell population is maintained in controlled growth conditions, over an extensive number of generations, enabling strains to evolve in response to environmental changes. Adaptation generally creates superior strains, with stable genotype, increased growth rates and production yields (Lee and Palsson, 2010).

To produce a significant improvement in ethylene productivity, the *rpoD* mutant library was subjected to ALE in continuous fermentation (**Figure 5.14**). The whole cell library was utilised to inoculate a 1 L bioreactor (750 mL working volume) and cultivation was carried out over 34 days in 0.8% glycerol M9 medium, supplemented with tetracycline. The dilution rate was increased stepwise to increase selection pressure towards cells with increased growth rates. As initially hypothesised, the higher the

dilution rate, the greater the growth rate, and consequently the higher the ethylene yield and productivity. Although the growth-based selection did not prove entirely robust as a phenotypical screen, it was utilised as a basis for candidate selection in the ALE.

The dilution rate was gradually increased from 0.01 h^{-1} to 0.07 h^{-1}. At steady state, the cell growth rate (h^{-1}) is equal to the dilution rate and cellular metabolism is at an equilibrium (K. Dane Wittrup, 2007). To obtain stably evolved strains, samples were taken from the fermenter just prior to a dilution rate increase, as depicted by the arrows (**Figure 5.14**).

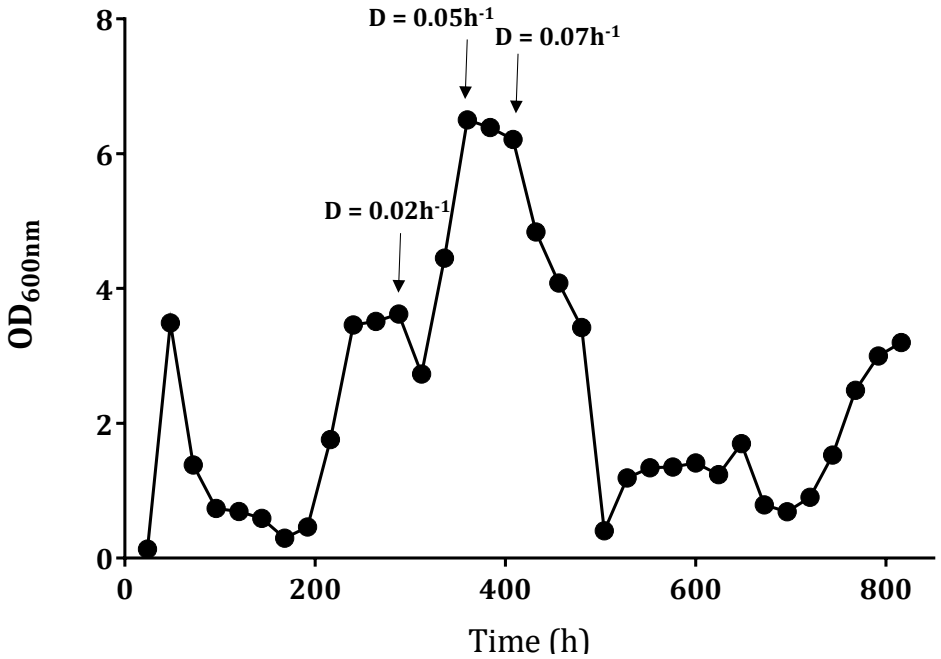


Figure 0.14 – Growth rate of the rpoD mutant library in the ΔproB/pGEM-efep background strain cultivated in continuous fermentation

The dilution rate D was gradually increased from D=0.01 h^{-1} (t=0) to D = 0.02 h^{-1} (12 days), D= 0.05 h^{-1} (15 days) and D = 0.07 h^{-1} (17 days). Isolation of potential overproducing strains was performed from samples collected before each incremental increase of the dilution rate.

Colonies were selected on LB agar plates supplemented with 100 µg/mL ampicillin and 15 µg/mL tetracycline (Amp 100 + Tet 15), selected colonies were then grown in M9 media in small serum bottles and

2 mL samples were taken from the headspace to sample for ethylene (**Figure 5.15**).

The samples taken at D = 0.02 h^{-1} generated 6 increased ethylene-producers out of 25 re-patched colonies, which was the highest number of ethylene-synthesising strains isolated from the ALE process (**Figure 5.15 A-C**). However, only two clones (1 and 4) produced detectable levels of ethylene from isolated colonies at D = 0.05 h^{-1} and D = 0.07 h^{-1}. The other isolated clones either did not survive on the second set of selective LB plates or did not grow when re-inoculated in M9 liquid culture or they did not produce ethylene. The best ethylene producers at each dilution rate were D0.02_15, D0.05_2 and D0.07_14, which generated 282, 305 and 366 nmol/OD/mL of ethylene (**Figure 5.15C**), which represents a 52, 61 and 94% increase in ethylene productivity compared to the control strain Δ*proB*/pGEM-*efep* + *prpoD*_wt. These results show a correlation between increased dilution rate and ethylene productivity (**Figure 5.15D**).

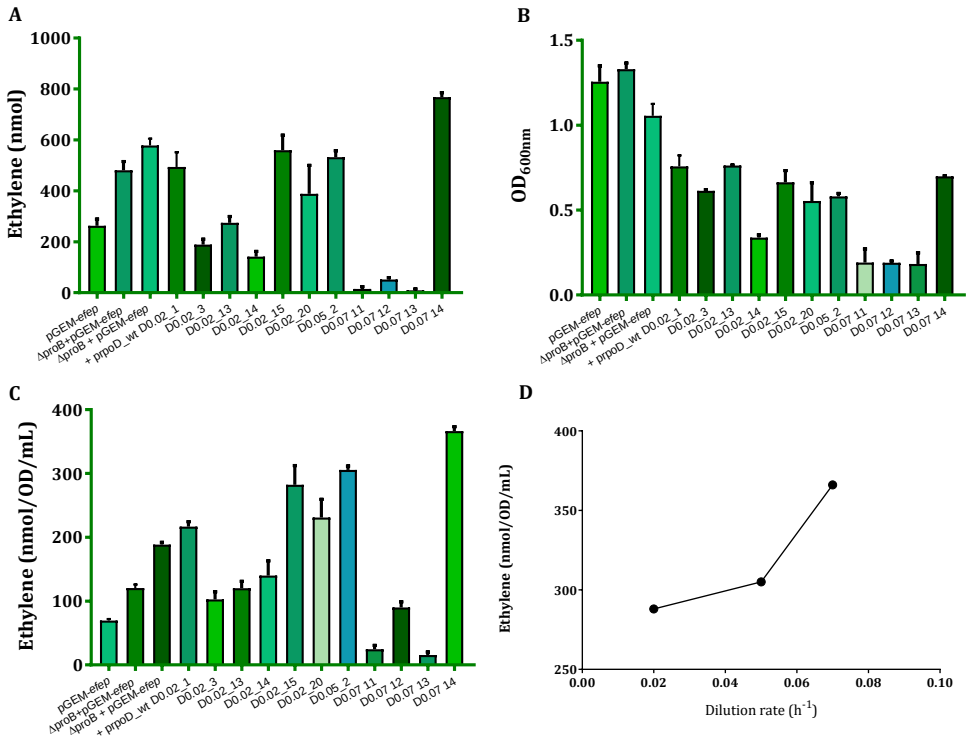


Figure 0.15 – Increased ethylene production using ALE selection

(**A**) Ethylene production, (**B**) OD_{600}, (**C**) ethylene productivity and (**D**) best ethylene productivities at increasing dilution rates from Δ*proB*/pGEM-*efep* + p*rpoD* mutants after ALE selection.

While ALE fermentation remains a powerful technique to quickly generate strains with an enhanced phenotype, the fermentation process can be lengthy and laborious. For these reasons, adaptive evolution of the *rpoD* mutant library was performed in M9 media with 0.4% glucose in a 96-well format on a robotic handling platform. Individual colonies were monitored utilising online OD measurement and faster growing strains were selected for further analysis (**Figure 5.16**).

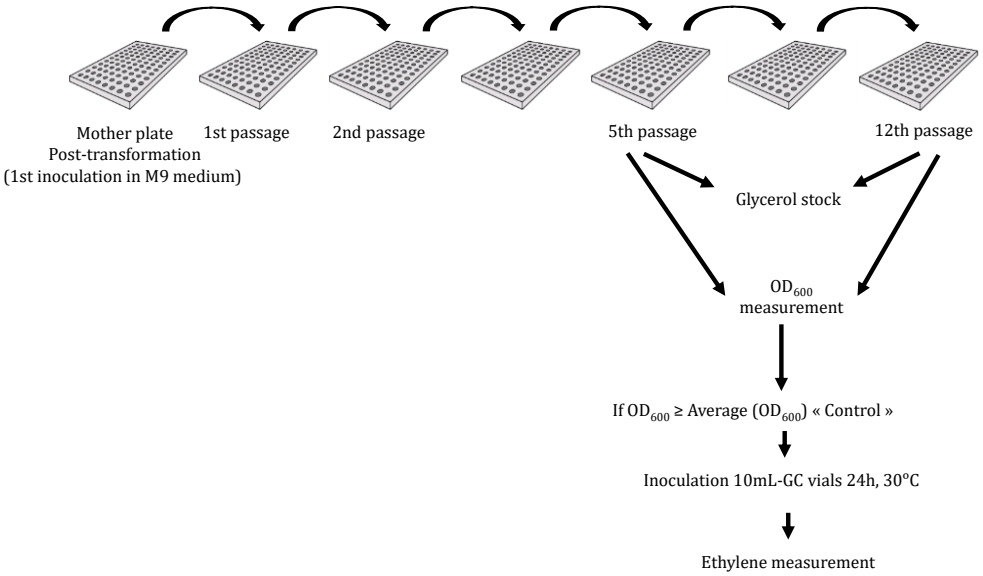


Figure 0.16 – Experimental procedure for semi-automated adaptive evolution in a 96-well plate format.

After the 5th passage, 289 strains were isolated and 25 of those had a growth rate similar to or higher than the control strain Δ*proB* + pGEM-*efep* + p*rpoD*_wt. The candidates with the best growth rates were selected for ethylene measurements. The seven best performing strains selected for GC analysis generated ethylene yields similar or better than the control strain **(Figure 5.17A**). An improved growth rate was confirmed for all seven (**Figure 5.17B**). The best ethylene producing strain generated 439 nmol/OD_{600}/mL, which is the highest ethylene productivity reached in *E. coli* to date (**Figure 5.17C**). It also had a 17% increase in ethylene productivity compared to the best performing strain generated through ALE, D0.07_14. Further analysis still needs to be completed on these strains, but the approach offers a promising platform for HTP ALE. Interestingly, the strains isolated after the 12th passage did not show any improvement in ethylene productivity, despite having an increased growth rate, supporting an uncoupling of the dual circuit mechanism (data not shown).

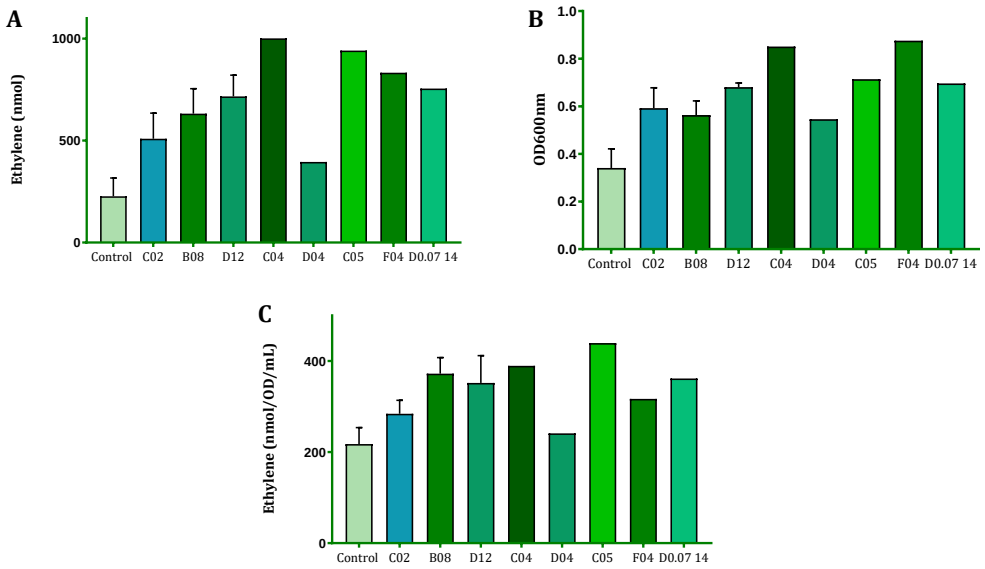


Figure 0.17 – Ethylene productivity in the fastest growing isolates after adaptive evolution in 96-well plates

(**A**) Ethylene production, (**B**) OD_{600}, and (**C**) ethylene productivity from Δ*proB*/pGEM-*efep* + p*rpoD* mutants after ALE selection in 96-well format.
The strains are named after the well position on the final selective 96-well plate. Ethylene productivity is compared to the Control strain, Δ*proB* + pGEM-*efep* + p*rpoD*_wt.
Measurements were done in triplicate after 24 h at 30°C, 200 rpm.

The combination of growth coupling, transcriptional engineering and adaptive evolution has allowed for the generation of a super-producing strain, with increased ethylene productivity (**Figure 5.18**). The final strain Δ*proB*/pGEM-*efep* + p*rpoD*_m (mutated *rpoD*), D0.07_14, generated up to 366 nmol ethylene/OD/mL, yielding a 5.2-fold increase in ethylene productivity compared to the reference strain MG1655/pGEM-*efep* and a 2-fold increase compared to the most significant improvement in *E. coli* ethylene production reported to date (188 nmol ethylene/OD/mL) (Lynch *et al.*, 2016). Preliminary results from the 96-well platform highlight the great potential of this approach, with peak ethylene productivity reaching 439 nmol/OD/mL in C05, which represents a 6.3-fold increase in productivity.

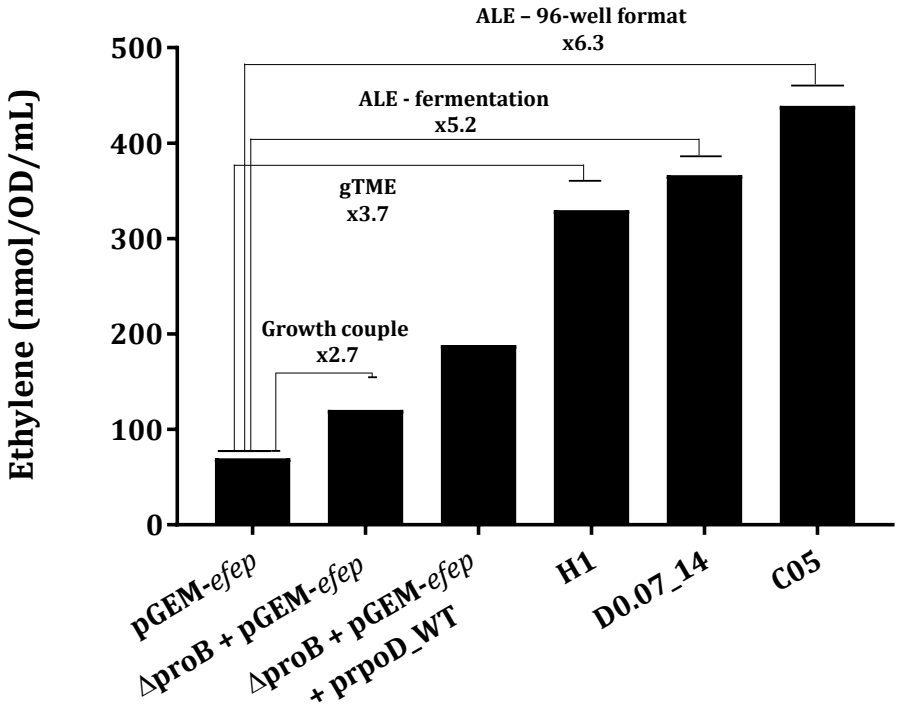


Figure 0.18 – Stepwise optimisation of ethylene productivity in MG1655

Comparison of peak ethylene productivities between evolved strains and the control strain *E. coli* MG1655 harbouring pGEM-*efep.* Strains include the Δ*proB* deletion mutant with either pGEM-*efep* or pGEM-*efep* + *prpoD*_wt and the best producing strains isolated after gTME (strain H1) and ALE selection (D0.07_14 and C05). The fold increase is indicated for each engineering technique.

5.3.4 Phenotypic characterisation of the mutant strains

Selected strains identified from both gTME and ALE fermentation were isolated and the sequence of the *prpoD* plasmid was analysed via Sanger sequencing (**Figure 5.19**). The 'winning' strains isolated at $D = 0.02\ h^{-1}$ and $D = 0.05\ h^{-1}$ possessed only one or two nucleotide substitutions (V255A and A375P in the D0.02_15 mutant and G398D in the D0.05_2 mutant). However, both the mutant strains H1 and D0.07_14, from the gTME library and the ALE, which had the highest ethylene productivity level contained a unique truncated version of the *rpoD* fragment. Both strains had a deletion of approximately 2 kb and 1.7 kb in the *rpoD* gene. Sequencing revealed that the open reading frame coding for RpoD had been disrupted, suggesting the protein was no longer viable.

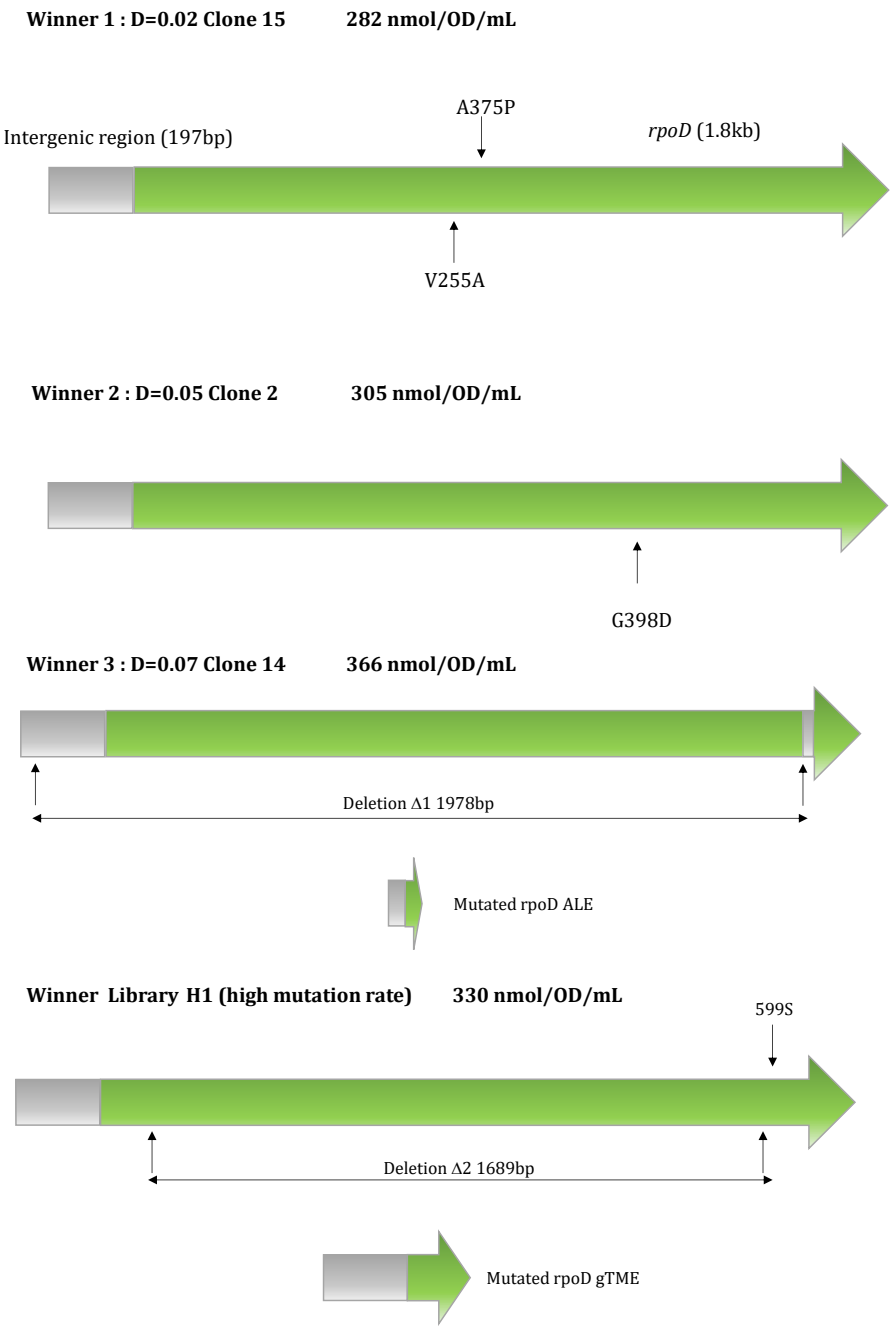


Figure 0.19 – Schematic representation of the identified mutants increasing ethylene productivity

The plasmid DNA from each best ethylene-producing strain identified from gTME and ALE was isolated and the *rpoD* fragment analysed by Sanger sequencing using the primers pLacZα.F/R which bind upstream and downstream of the fragment on the p*rpoD* plasmid. Nucleotide substitutions and deletions are annotated at the sequence locus and truncated fragments were reconstituted schematically.

The truncated *rpoD* sequence from D0.07_14 was amplified via PCR and cloned into the empty *prpoD* backbone to generate the plasmid *prpoD*_truncated. This plasmid was then transformed into Δ*proB*/pGEM-*efep* to create the strain Δ*proB*/pGEM-*efep* + p*rpoD*_truncated. If the *rpoD* deletion in the plasmid was responsible for the observed phenotype, the newly generated Δ*proB*/pGEM-*efep* + *prpoD*_truncated strain should generate a similar ethylene level as D0.07_14, providing the increase in productivity was down to the truncated gene and not due to a mutation in the *E. coli* chromosome. Both strains were tested for ethylene productivity using the GC.

The newly generated Δ*proB*/pGEM-*efep* + *prpoD*_truncated strain produced similar levels of ethylene to the D0.07_14 strain (363 nmol/OD/mL versus 366 nmol/OD/mL; **Figure 5.20A and C**). However, growth was impaired in the Δ*proB*/pGEM-*efep* + *prpoD* strain compared to the other strains (**Figure 5.20B**). The slower growth observed in the evolved strains could be down to adaption, as each of the other strains were repeatedly re-sub-cultured in M9 media, allowing the strains to adapt. Due to time constraints, this did not happen in the evolved strains and prolonged adaptation of the reconstructed strain by successive subcultures in 0.4% (w/v) glucose M9 supplemented with 15 μg/mL tetracycline could have been beneficial to obtain more representative data.

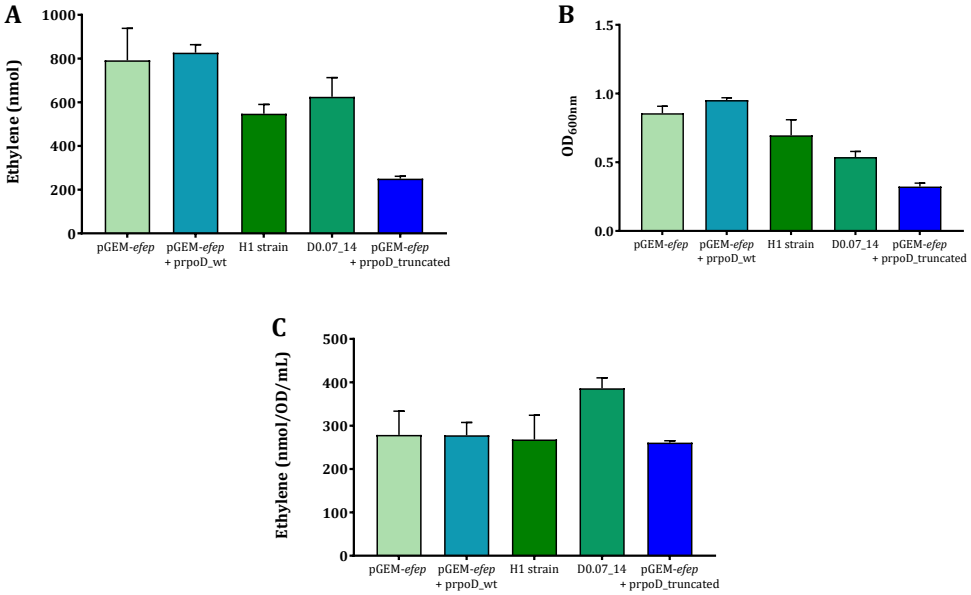


Figure 0.20 – Comparison of ethylene productivity between D0.07_14 and ΔproB/pGEM-efep + prpoD

(A) Ethylene production, (**B**) OD_{600}, and (**C**) ethylene productivity from a culture of Δ*proB*/pGEM-*efep* + pr*poD*_truncated and associated control strains. All strains are auxotroph for proline.
Measures were done in triplicate after 24 h at 30°C, 200 rpm.

5.3.5. Genomic and transcriptomic analysis of the mutant strain D0.07_14

Sequence analysis of the *efep* gene in D0.07_14 confirmed that no mutations occurred within the *efep* gene or the promoter j5. As the sequence of the *rpoD* gene in this strain was extensively disrupted, it was hypothesised that increased ethylene productivity had arisen from genetic variation in the chromosome. To assess this, two separate fermentations were set up, with D0.07_14 and Δ*proB*/pGEM-*efep* + *prpoD*_wt, utilising the same conditions as above. Once steady state had been reached, samples were collected for whole genome sequencing (WGS) and transcriptomic (RNA-seq) analysis. Samples were processed by Genewiz® (South Plain Field, New Jersey, USA).

Four gene or genomic elements were identified with single nucleotide variations (SNV) or small (1-2bp) deletions in both the control and evolved strains (**Table 5.3**).

Two of these modifications were located on mobile or repeat elements and were probably the result of random mutations. The two other mutations affected two genes, *gatC* (b2092) and *glpR* (b3423), which are involved in the regulation of metabolic operons related to the carbohydrate intermediates (galactitol and glycerol-3-phosphate). The nucleic variations recorded in *gatC* and *glpR* are strongly correlated to the utilisation of glycerol and may suggest strain adaptation to the carbon source used in the culture medium. Indeed, modification of the phosphotransferase system had been a positive factor in sugar co-assimilation in *E. coli* (Balderas-Hernández *et al.*, 2011).

Table. 0.3 – Common gene mutations in D0.07_14 and ΔproB/pGEM-efep + prpoD_wt

Gene or genetic element	Locus	Function	Mutation
IS1I	_	Mobile element	SNV
gatC	b2092	phosphotransferase system galactitol-specific EIIC component	2nt-deletion
glpR	b3423	transcriptional repressor of the glycerol-3-phosphate operon	1nt-insertion
REP321j	_	repeat sequence	2nt-insertion

SNV: Single Nucleotide Variation

In addition to the four shared mutations between the two strains, D0.07_14 and *ΔproB*/pGEM-*efep* + *prpoD*_wt, the overproducing clone D0.07_14 possessed mutations in four other genes (**Table 5.4**). A 4-nt-deletion in the *fhuA* gene led to a perturbation in the open-reading frame, possibly giving rise to a truncated and therefore non-functional protein. The *fhuA* gene encodes a ferrichrome outer membrane receptor, involved in iron uptake in a complex with ferrichrome, a cyclic hexapeptide with iron ion affinity (Braun, 2009). FhuA binds and transports the ferrichrome-iron cargo, across the outer membrane into the

cytoplasm. EFE(P) requires an iron atom in its heme-like binding pocket (Martinez *et al.*, 2017). This hypothesis is supported by the transcriptomic analysis of both the control and evolved strains. An increase of 15% in the expression level of the *fhuA* gene was observed in D0.07_14 compared to the control strain (RNA sequencing data discussed below, *p*value = 0.22). The overrepresentation of the Fe (II)-dependent EFE(P) may have depleted the cellular iron level, which resulted into the mutation in the *fhuA* gene sequence to increase the ferrous ion uptake.

Eleven unique variations involving one or two nucleotides have been identified across the sequence of the *cyoB* (b0431) gene. The *cyoB* gene encodes the cytochrome O ubiquitinol oxidase subunit I, included in the cytochrome c oxidase, complex IV of the aerobic respiratory chain, which is upregulated when cells are grown at high aeration (Chepuri *et al.*, 1990). The multiple mutations in the *cyoB* sequence are very likely to be a consequence of the growth conditions in the ALE fermentation. The inactivation of *cyoB* in *Pseudomonas putida* reduced catabolite repression in both LB and defined medium, upregulating the alkane degradation pathway. The multiple gene mutations present in the *cyoB* gene in D0.07_14 might result from adaptation to the growth conditions, e.g. a constant high aeration and a medium containing glycerol as the carbon source.

Both the *alaU* and *rrlA* genes are both involved in protein synthesis and the numerous mutations affecting their sequence might be related to the *rpoD* mutation.

Table 0.4 – Strain-specific gene mutations in D0.07_14

SNV: Single Nucleotide Variation
MNV: Multiple Nucleotide Variation

Gene ID	Locus	Function	Mutation
fhuA	b0150	ferrichrome outer membrane transporter	4nt-deletion (CAGG)
cyoB	b0431	cytochrome O ubiquitinol oxidase subunit I	10 SNV + 1 2nt-MNV
alaU	b3276	Transfer RNA (Ala)	1nt-insertion + 3 2nt-MNV + 3 SNV
rrlA	b3854	23S ribosomal RNA	13 SNV + 3 2nt-MNV + 2 3nt-MNV

The adaptive evolution of D0.07_14 led to many nucleotide variations across four different loci associated with iron flux, the respiratory chain, carbon utilisation and protein synthesis.

The WGS analysis was further supported by transcriptomic analysis (RNA sequencing). The whole cell RNAs were counted and abundance levels were compared between the *rpoD* mutant and wild-type strains. Changes in gene expression are represented as a log2FoldChange with a *p*value that reflects the significance of that log2FoldChange. Seven genes were identified which had a significant difference in gene expression (if *p*value < 0.05) (**Table 5.5 and Figure 5.21**).

The *rpoD* gene was highly overexpressed in the *rpoD* wild-_type strain (Δ*proB*/pGEM-*efep* + *prpoD*_wt) compared to the mutant D0.07_14, which had the truncated *rpoD* gene. This result suggests that the truncated version was either down regulated or not expressed. Alternatively, the chromosomal copy of the *rpoD* gene may also have been downregulated in D0.07_14.

Table 0.5 – Significant differential gene expression in D0.07_14 versus ΔproB/pGEM-efep + prpoD_wt

Gene ID	Locus	Function	log2FoldChange
glpD	b3426	Aerobic glycerol-3-phosphate dehydrogenase	-2.89
nlpI	b3163	Lipoprotein	-1.73
ompX	b0814	Outer membrane protein	-1.72
mntH	b2398	Divalent metal cation transporter	-1.37
ariR	b1166	Putative regulator of acid-resistance	-1.22
cysJ	b2764	Sulfite reductase	-1.04
rpoD	b3067	RNA polymerase sigma factor	4.19

The log2FoldChange is the log-transformed ratio of expression between the two strains analysed. A negative log2FoldChange indicates an overexpression in D0.07_14 compared to the reference strain Δ*proB*/pGEM-*efep* + p*rpoD*_wt whilst a positive log2FoldChange indicates a gene downregulation.

The six other genes identified in the transcriptomic analysis of Δ*proB*/pGEM-*efep* + p*rpoD*_wt and D0.07_14 showed higher expression levels in the evolved strain than in the control (**Figure 5.21**). The significant expression difference in these genes in D0.07_14 could result from strain adaptation to either glycerol as a carbon source (*glpD, ariR*), the increased requirement for iron due to EFE(P) activity (*mntH*), the dilution rate (*cysJ*) or, more generally, stress related to the growth conditions (*nlpI, ompX*).

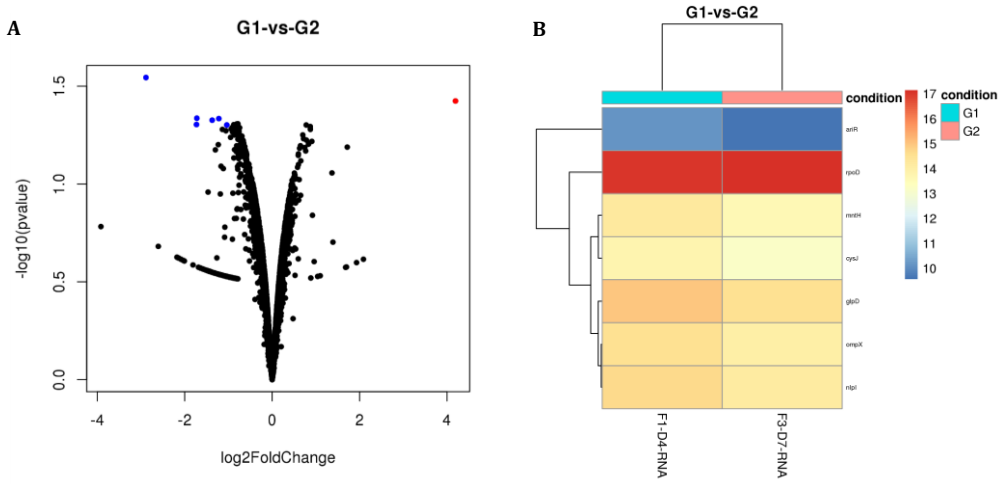


Figure 0.21 – Differential gene expression analysis between D0.07_14 (condition G1) and ΔproB/pGEM-efep + prpoD_wt (condition G2)

A. Volcano plot displaying differentially expressed whole cell RNA in the overproducing strain D0.07_14 and the reference strain Δ*proB*/pGEM-*efep* + p*rpoD*_wt. The log2FoldChange ratio (x axis) and −log10 adjusted p values (y axis) of expression shift between the two strains analysed, were plotted in the form of a volcano plot. Red colour represents significantly up-regulated (right) RNAs and blue colour indicated down-regulated (left) RNAs, respectively (log2FoldChange < 2 or > 2 and adjusted p < 0.05). Black colour represents the small fold change (log2FoldChange < 1.5 or > 1.5) and the non-significant change (-1.5< log2FoldChange <1.5). Total RNAs from one fermentation sample per strain collected at steady state were analysed on the 10xgenomics Chromium™ Single-Cell RNA-Seq platform (Genewiz®, South Plainfield, NJ, USA).
Important: the sample from strain D0.07_14 was designated as condition G1. However, the analysis automatically considered condition G1 as the reference strain and condition G2 as the variant strain. That is why the red spot located on the right top corner of the plot represents the *rpoD* gene: it has a positive 2LogFoldChange in G2 compared to G1, which means overexpression of this gene in Δ*proB*/pGEM-*efep* + *prpoD*_wt compared to D0.07_14. For clarity, a positive 2LogFoldChange is considered as gene downregulation in this particular analysis.
B. Heatmap demonstrating biclustering of the RNAs based on significant differential expression levels calculated from normalised read counts. Red indicates high levels of expression; medium shades represent medium expression levels and blue indicates low levels. Genes were segregated on the adjusted p value and only those associated to a p< 0.05 were considered. Condition G1/F1_D4 RNA refers to the D0.07_14 strain while G2 is associated with the strain Δ*proB*/pGEM-*efep* + p*rpoD*_wt.

Among these upregulated genes, the *glpD* gene showed the greatest log2FoldChange (~-2.89) and is involved in the glycerol degradation pathway (Austin and Larson, 1991). The gene codes for an aerobic glycerol-3-phosphate dehydrogenase. The enzyme is expressed maximally under aerobic growth conditions (Lin, 1987). Thus, the overexpression of this protein was largely affected by the growth conditions imposed on the strain.

The elevated expression of *ariR* in D0.07_14 can also be attributed to the growth conditions. Although AriR is a regulatory protein involved in biofilm formation (in rich medium) and cell motility, it is also involved in acid resistance mediated by an indole molecule (Lee *et al.*, 2007). High expression of *ariR* was observed in *E. coli* K12 cultivated in 0.2% (w/v) glycerol M9 medium (White-Ziegler *et al.*, 2008). The higher mRNA level observed for this gene in the overproducing strain compared to Δ*proB*/pGEM-*efep* + *prpoD*_wt, may be due to a low pH level, perhaps resulting from the formation of fermentative co-products, like acetate, lactate or formate. Alternatively, it could be due to adaptation to the glycerol carbon source.

The *nlpI* gene encodes for a lipoprotein involved in the regulation of peptidoglycan degradation, which greatly affects cell osmosis and outer membrane vesicle budding (Schwechheimer, Rodriguez and Kuehn, 2015). Alteration of the sequence of the *nlpI* gene via insertion of a chloramphenicol resistance cassette altered the shape of *E. coli* K12 cells, with a loss of rod morphology and apparition of ellipsoids (Ohara *et al.*, 1999). The *nlpI* gene may be involved in cell division but the underlying mechanism remains elusive (Nanninga, 1991). Therefore, it was hypothesised that overexpression of the *nlpI* gene was related to environmental stress altering cell structure.

RNA levels for *ompX* were also increased in D0.07_14. OmpX is involved in cell adherence, motility (Otto *et al.*, 2001) and activation of sigma E expression (Mecsas *et al.*, 1993), a transcription factor essential for the extra cytoplasmic stress response (De Las Peñas, Connolly and Gross, 1997). It is plausible that once again the environmental stress induced significantly higher transcription levels of genes involved in the extra cytoplasmic response.

The *mntH* gene was also overexpressed. The proton-dependent metal transporter MntH belongs to the NRAMP orthologous family and has an affinity to Mn^{2+} and Fe^{2+} (Makui *et al.*, 2000). Transcription of *mntH* is dually repressed by the Mn^{2+}-dependent MntR transcriptional regulator and by the Fe^{2+}-associated ferric uptake regulator Fur (Patzer and Hantke, 2001). The transcriptomic analysis also revealed a slight overexpression of both the *mntR* (Log2FoldChange = -0.16) and *fur* (-0.23) genes in D0.07_14, however both were non-significant ($p\text{value}_{mntR}$ = 0.35 and $p\text{value}_{fur}$ = 0.25). Therefore, it could be hypothesised that the increasing need for iron to accommodate the formation of a functional EFE(P) enzyme necessitated a higher iron uptake and stimulated the overexpression of the metal transporter MntH. Alternatively, the overexpression of the *mntH* gene can result from a general cellular response to environmental stress.

Finally the overexpression of the NADPH-dependent sulfite reductase-encoding *cysJ* gene suggests a greater availability of sulfide (H_2S) and $FADH_2$ (Ostrowski *et al.*, 1989). Indeed, CysJ was also identified as the flavin oxidoreductase component of the CysJI sulfite reductase complex (Eschenbrenner, Covès and Fontecave, 1995). This could impact on two energetic pathways. Firstly, sulfide, along with acetyl-L-serine, is a precursor of L-cystein synthesis, catalysed by CysK (Kredich, 2008). In fact, transcriptomic data revealed overexpression of *cysK* (Log2FoldChange = 2.1, *p*value = 0.24). Therefore, it may be assumed that the mutant strain evolved towards an enhanced biosynthesis of cystein, to support growth and protein production. Secondly, CysJ is the flavin subunit of the sulfite reductase; it accepts electrons from NADPH (Kredich, 2008) and catalyses the reduction of FMN and FAD (Eschenbrenner, Covès and Fontecave, 1995). Increased expression of *cysJ* might be interpreted as a cell adaptation strategy to high growth rate and higher demand for cellular material.

In most of the examples listed above, there is a coherent cause-effect relationship between stress conditions imposed by culture settings (high aeration, minimum medium with a single carbon source, increasing growth rate) and the differential RNA levels observed. The cell membrane shapes the cell, protects it from external potential threats and facilitates gas and nutrient exchange between the extracellular matrix and the cellular cytoplasm, where most of the reactions essential for growth and energy maintenance are catalysed. Therefore, it is not surprising to find that most of the genes overexpressed in the evolved strain code for membrane proteins (*glpD, nlpI, ompX, mntH and ariR)*.

Similarly, an *rpoD* mutant library was generated for CH34. The mutated plasmid library was developed and readily available for transformation in CH34 expressing a modified *efep*-expressing plasmid, pHokSok-RFP. This plasmid carried an operon, controlled by the constitutive p_{j5} promoter that included the *efep* gene and the *rfp* reporter gene. Co-expression of EFE(P) and RFP aimed at enabling a direct monitoring of the transcription strength of the *efep* gene (no supportive data available). The study of gTME/ALE-produced ethylene-producing mutants in *E. coli* was conducted to support the trial in CH34. However, due to time constraints, the development of a gTME/ALE-generated CH34 mutant with increased ethylene productivity could not be achieved.

The transcriptomic analysis of D0.07_14 suggested potential target genes to overexpress to improve ethylene productivity in *E. coli* K-12. Though rational engineering would aid in investigating further the role of each gene in the ethylene-overproducing strain, a more thorough adaptive selection procedure, accompanied by a comprehensive mutant analysis, would be more helpful in generating an ethylene-producing super strain. For example, ALE could be performed at higher dilution rates, more isolation at each pseudo steady-state could be performed and a direct biosensor could be developed to readily assess ethylene levels in a high-

throughput manner. Frequent sampling for WGS and RNA-seq covering the whole adaptation culture run could help identify a more genetically diverse mutant pool, representing both a high ethylene productive phenotype and a stabilised genotype. Additionally, a comprehensive metabolomics analysis of the mutants would enable a precise evaluation of the carbon flux across the cell metabolism and especially around the EFE(P) enzyme (TCA cycle, proline synthesis pathway).

5.4 Discussion and future work

The implementation of the EFE pathway via the heterologous expression of *efep* allowed the very first production of ethylene in *C. metallidurans* CH34. The selection of acetate as the optimal carbon source, followed by optimisation of the carbon source concentration and nutrient supplementation, led to a 3-fold increase in ethylene productivity, reaching 63 nmol/OD_{600}/mL in MSM medium enriched with 10 mM acetate, 2 mM AKG and 3 mM arginine. However, H16 cultivated in 5 mM acetate with no supplementation, reached an ethylene productivity four times higher (230 nmol/OD_{600}/mL) than the best performance in CH34, highlighting the physiological differences between both strains despite high genomic similarity.

One of the most straightforward strategies to boost ethylene synthesis in CH34 would be to utilise a range of promoters, RBSs and vector backbones to optimise the transcription of *efep* and translation of EFE(P)EFE(P) in CH34. The synthetic biology toolbox available for H16 (Fukui *et al.*, 2011; Bi *et al.*, 2013; Gruber *et al.*, 2014; Tee *et al.*, 2017; Johnson *et al.*, 2018) could therefore be a good resource for the application of metabolic engineering in CH34. The development of HTP genetic tools, the major driving force of this thesis, could enable flux to be pushed towards ethylene production in CH34, particularly given the tight regulation around the TCA cycle in this strain.

In addition, biomanufacturing of ethylene in continuous fermentation will require the stable expression of *efep*. The Hok/Sok addiction system enabled 100% plasmid retention in tube cultures but could not maintain it longer than 192 h in a fermenter. H16-compatible addiction systems (Voss and Steinbüchel, 2006; Fricke, Kusian and Bowien, 2009; Budde *et al.*, 2011) could be considered for transfer in CH34. Otherwise, chromosomal integration of *efep* would be desirable for stability

purposes, but a lower copy-number of the gene could potentially drastically reduce the ethylene yield, if gene expression is not further optimised (Lynch *et al.*, 2016).

Supplementation in AKG and arginine enabled enhancement of the ethylene yield in CH34, which suggests substrate availability is a rate-limiting factor in ethylene production. Lynch and colleagues optimised ethylene productivity in *E. coli* by knocking competing genes out and overexpressing key enzymes to enhance the flux towards AKG (Lynch *et al.*, 2016) (**Figure 5.22**). The highest ethylene yields were obtained in the Δ*argR/ΔgltBD* (188 ± 13.1 nmol/OD_{600}/mL) deletion mutant and the *icd* overexpressing strain (135 ± 7.9 nmol/OD_{600}/mL). This approach could be utilised in CH34, but it requires an effective HTP toolbox for gene augmentation.

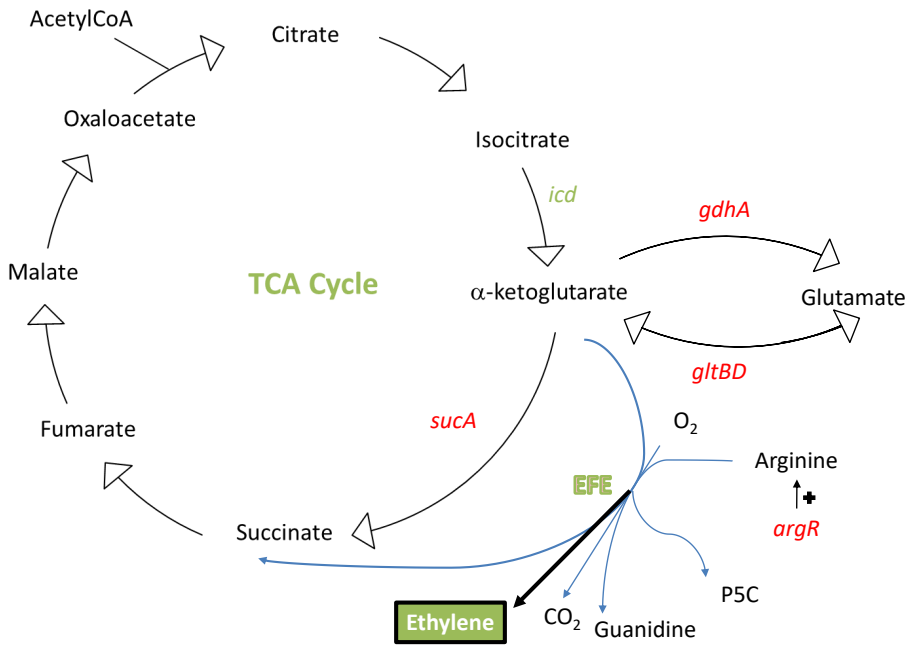


Figure 0.22 – The effect of genetic modifications on ethylene productivity

Putative metabolic scheme for ethylene production in *E. coli* via EFE. Genes responsible for the catalytic steps or regulation relevant targets identified by Lynch and co-workers (Lynch *et al*, 2016) for enhancement of ethylene yield are indicated in red (knockout) and green (overexpression).

icd : isocitrate dehydrogenase; *gdhA* :glutamate dehydrogenase, *gltBD:* glutamate synthase; *argR:* transcriptional regulator of arginine biosynthesis; *sucA*: 2-oxoglutarate decarboxylase.

Despite several attempts to knockout *proB* (Rmet_3103), *proA* (Rmet_2970), *proC* (Rmet_2938) and *sucA* (Rmet_2050) in CH34 utilising the *sacB* counter-selection method, no mutants could be generated. However, the riboswitch-controlled CRISPR system described in **Chapter 4** could be re-targeted against these genes and a CRISPRa system could be developed to enhance the expression level of the *icd* gene (Rmet_2895). Homologous recombination and the CRISPR-based recombineering systems developed in **Chapter 3** and **Chapter 4** of this study have not yet generated a deletion mutant in CH34 and require further optimisation.

To negate the problems associated with metabolic engineering, gTME was tested as an alternative strategy for use in CH34. To test whether the strategy could work for ethylene productivity, the model chassis *E. coli* was utilised initially. The maximum ethylene yield reached 439 nmol ethylene/OD/mL after gTME and ALE selection, which is 6.3-fold increase compared to the primary productivity obtained in the strain and 2.3-fold higher than the best *E. coli* productive strain reported to date (Lynch *et al.*, 2016). Although multi-copies of the *Pseudomonas syringae* pv. *glycinea efe* gene resulted in a significantly better production rate (**Table 5.6**), it should be observed that this rate was reached in LB rather than in minimum medium and that *P. putida* is not able, unlike H16 and CH34 (Pohlmann *et al.*, 2006; Janssen *et al.*, 2010), to grow on CO_2 as a sole carbon source and therefore utilise CO_2 for ethylene formation. Higher yields might have been achieved in *E. coli* if more clones had been screened and a more systematic screening approach had been adopted for strain selection from the gTME library or amongst post-ALE isolates. Additionally, an extended library size would be desirable (at least 10^6 clones) to consider the gTME procedure as high-throughput and maximise output. The gTME method was then applied

to CH34, however due to time constraints the library could not be completed.

Table 0.6 – Comparison of ethylene production rate of engineered strains developed in this study to other engineered ethylene-producing chassis

Organism	Production rate (µmol/gDCW/h)	Growth medium	Reference
Cupriavidus necator H16	28.7	FGN	PC
Cupriavidus metallidurans CH34	10.1	MSM + 10 mM acetate + 2 mM AKG + 3 mM arginine	This study
E. coli MG1655 (ALE in 96-well format)	70.1	M9 + 0.4% glucose	This study
E. coli MG1655 (chemical mutagenesis + ALE)	47.9	M9 + 0.4% glucose	UP
E. coli MG1655	30	M9 + 1% glucose	(Lynch *et al.*, 2016)
Pseudomonas putida	2859.2	LB	(Wang *et al.*, 2010)

Strain evolution and selection based on the growth couple driven through P5C, may be an impediment to the identification of the best producing clones. Indeed, constraining EFE(P) towards the production of P5C, might uncouple the dual circuit mechanism proposed (Martinez and Hausinger, 2016; Martinez *et al.*, 2017). In the common reaction pathway, binding of AKG and arginine induces the formation of a twisted peptide bond at the active site of EFE(P) (**Figure 5.23**).

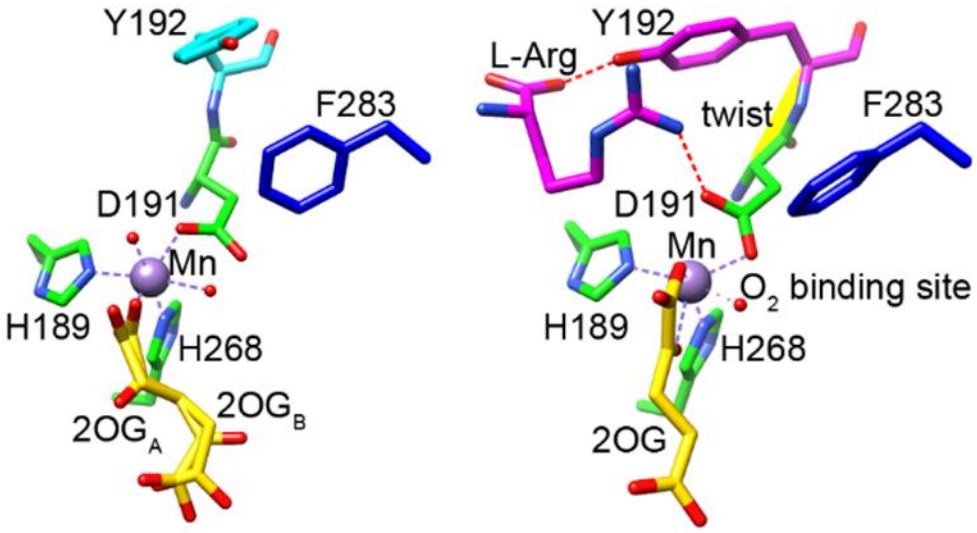


Figure 0.23 – Identification of a twisted peptide bond and conformational changes in EFE·Mn·AKG·Arg

The C1-carboxylate oxygen of AKG (2OG, yellow carbons) binds trans to H268 and the C2-keto oxygen binds opposite D191, thus defining a dioxygen-binding site that points away from the C5 position of the arginine (L-Arg) substrate (magenta carbons). The 2-histidine-1-carboxylate metal-binding motif is shown with green carbons, and the nearby F283 residue is shown in blue. The dioxygen-binding site is illustrated by a water molecule (red sphere). Manganese (Mn) chelation is shown by dashed lines, a key hydrogen bond between NH1of L-Arg and OD1 of D191 is indicated by a red dashed line. The twisted peptide is highlighted in yellow (Martinez *et al*, 2017)

Then the cyclic peroxide-Fe (IV) intermediate is formed upon dioxygen binding to the metallocenter (iron). At this stage, the mechanism diverges into two pathways. In the first branch, the oxidative decarboxylation of AKG is processed along with the hydroxylation of arginine, generating one molecule of succinate, one molecule of P5C and one molecule of guanidine (**Equation 1**). In the second branch, the decarboxylative defragmentation of AKG produces two molecules of ethylene and three molecules of CO_2 (**Equation 2**).

(1) L-Arginine + alpha-ketoglutarate → Succinate + L-Δ-1-pyrroline-5-carboxylate (P5C) + Guanidine+ CO_2

(2) Alpha-ketoglutarate → Ethylene + $3CO_2$

The molecular ratio of equations **(1)** and **(2)** is 2:1 (Fukuda *et al.*, 1992) but the growth-couple may disrupt this equilibrium and push towards P5C production at a cost to ethylene formation. Growth can be useful to monitor phenotype diversity, because it can be readily determined and reflects many factors affecting the cell (Klein-Marcuschamer and Stephanopoulos, 2008). However, it could be misleading in the isolation of the best ethylene producers. A robust phenotype screen must target the desired end-product and be reliable and straightforward.

An ethylene biosensor with a dynamic range is essential to ensure consistent and robust evaluation of ethylene biosynthesis, but this has proved very difficult to establish. Repeated efforts have been made in our group to develop an ethylene biosensor from plants and cyanobacteria. The SynEtr1 protein from the cyanobacterium *Synechocystis* spp. PCC 6803 was the first bacterial ethylene receptor to be characterised (Lacey and Binder, 2016). In this two-component system, ethylene directly binds to an N-terminal transmembrane ethylene binding domain (EBR) formed of three α helices (W. Wang *et al.*, 2006). The C-terminal end of the receptor is a histidine kinase domain, which plays a role in SynEtr1 function , although the signal mechanism of SynEtr1 has not been elucidated yet (Lacey and Binder, 2016). Similar ethylene receptors can be found in plants like the ETR1 receptor in *Arabidopsis thaliana* which has been characterised via heterologous expression in yeast (Schaller, Shiu and Armitage, 2011). Five isoforms of ETR1 have been identified in *A. thaliana* (Hua *et al.*, 1998) and include an EBR similar to that of SynEtr1 (W. Wang *et al.*, 2006) though ethylene binds to ETR1 via a copper cofactor (Rodriguez *et al.*, 1999). The GAF, histidine and receptor domains of plant ethylene receptors are involved in the output signal, related to ethylene binding (Lacey and Binder, 2014). Conformational changes to the receptors, independent of these domains, have been shown to cause alterations in receptor-protein interactions, which resulted in the downregulation of downstream cellular

events (Gao *et al.*, 2003; Huang *et al.*, 2003). Because there is no structural understanding of the ethylene-dependent signalling mechanism, all attempts to use these receptor systems as biosensors have failed.

Alternatively, transcriptional analysis of *E. coli* cultured in the presence of high concentrations of ethylene versus no ethylene allowed for the identification of four genes upregulated in the presence of high ethylene levels (Dr Sam Bryan, personal communication). Expression of an RFP reporter gene controlled by the promoter of each of these genes was utilised in *E.coli* MG1655 but it was not possible to detect ethylene via this reporter system.

In addition, the independent functionality of the two reaction mechanistic pathways of EFE(P) has been demonstrated but no mutant strain has been generated to exclusively form ethylene (Fukuda *et al.*, 1992). Thus, directed evolution of EFE(P) and random mutagenesis are additional strategies to consider to optimise the sequence of *efep* towards better ethylene productivity. Indeed, a version of EFE(P) dedicated to the formation of ethylene without the arginine hydroxylation pathway would be highly desirable. Recent understanding of the reaction mechanism as well as insights in the thermodynamics of the enzyme could pave the way towards a 100% ethylene-producing EFE(P) enzyme (Martinez *et al.*, 2017; M. Li *et al.*, 2018).

In summary, the implementation of genome, transcriptome and strain engineering is a great example of what synthetic and systems biology can deliver towards optimisation of bioproduction of industrially valuable commodities.

General conclusion

6.1 General conclusion

Current methods utilised in *Cupriavidus* spp. for genome editing are laborious and inefficient, limiting the application of this strain beyond PHB production. This study aimed to directly address this problem by developing HTP editing tools, enabling both *C. necator* and *C. metallidurans* to be fully utilised as an SBRC platform for ethylene, isoprene and 3-hydroxybutyric acid (3-HBA) production.

Several HTP genome editing tools were evaluated in *C. necator,* these included λ-Red/RecET recombineering and CRISPR/Cas9. The λ-Red/*recET* system was evaluated utilising an arabinose inducible promoter, driving the phage lambda genes *exo bet* and *gam.* Although this system has been very successful in *E. coli* and closely related strains (Sawitzke *et al.*, 2007; Swingle *et al.*, 2009; Murphy, 2016), it was unsuccessful in *C. necator* and failed to generate any knockouts, despite repeated efforts. **Chapter 3** focused on the implementation of exonuclease/annealing proteins, which do not rely on RecA activity (Ellis *et al.*, 2001). However, it has been reported that the transient co-expression of RecA with a λ-Red recombination system enabled cells to reach better recombination efficiencies (J. Wa*ng et al*, 2006). Future work could investigate the implementation of RecA with the λ-Red system, it would be essential to explore the native recombination pathway, particularly the RecA-dependent recombination system and revisit the recombination protocols based on the results presented in this study.

The Rac prophage encodes both the RecE and RecT proteins, these are functionally equivalent to the Lambda Exo and Beta proteins. Two different sets of RecET homologs were evaluated in *C. necator,* one from a closely related *Cupriavidus* spp (*recET-C*) and one from *P. syringae* (*recET-P*). Although the *recET-C* failed to generate any knockouts in *C. necator,* the *recET-P* strain did result in the insertion of a spectinomycin resistant PCR

cassette into the genome of H16. Although the isolation of a mutant strain could not be confirmed, these preliminary results undoubtedly demonstrate that this system can lead to the incorporation of DNA into the genome. Further work to optimise this system is ongoing.

With the aim to find a better editing technology than heterologous recombination systems, **Chapter 4** sought to implement the CRISPR/Cas9 technique in *C. necator.* The most striking finding following the numerous CRISPR experiments was the low efficiency of the native recombination system, which required an extended outgrowth to allow for effective recombination. However, the increased recovery time may favour a greater population size and might make the screening of recombinants difficult. Therefore, the CRISPR tool, utilised initially as a means to introduce DSBs and put cells in a 'repair or die' situation, proved more efficient as a counter selection tool.

The expression of SpyCas9 under control of a theophylline induced riboswitch, proved most effective and enabled the selection of Δ*phaC* H16 mutants, after four days of recombination using plasmid-borne homology arms as a template. Although the editing efficiency reached only 40%, this CRISPR method remains faster than the standard *sacB*-based protocol, which requires about three weeks to produce a mutant (Lenz, Lauterbach and Frielingsdorf, 2018). Further improvements could remove some of the bottlenecks associated with this method and the recently published pBAD-Cas9 system (Xiong *et al*, 2018). For example, the expression of the *recET-P* system, developed in Chapter 4 will increase the recombination frequency and improve the editing efficiency. Furthermore, a *recET*-based CRMAGE alternative could be generated, which could be transferred to different strains (Ronda *et al.*, 2016). The combination of both CRISPR/Cas9 and a recombinase system has already improved metabolic engineering strategies in both *P. putida* and *C. glutamicum,* utilising ssDNA as a template (Cho *et al.*, 2017; Aparicio, de Lorenzo and Martínez-García, 2018). Future

work could include adapting the ssDNA recombineering technique to generate multiplex mutants in *Cupriavidus* species (Wang *et al.*, 2009). Finally, a two-system plasmid could be utilised to decouple and fine-tune the titre levels of the Cas9 protein and sgRNA.

Future work utilising CRISPR in *C. necator* should include the development of an automated process for the CRISPR-based generation of mutant libraries (Garst *et al.*, 2016). These libraries could then be utilised either to unravel genotype-phenotype relationships or to rewire chassis to control metabolic genes and maximise product titres (Kim *et al.*, 2017). Cpf1 and Cas13a, could also be utilised to develop a CRISPR/Cpf1-based editing method in *Cupriavidus* species (Cox *et al.*, 2017; Yao *et al.*, 2018).

Both recombineering and CRISPR rely on efficient homologous recombination and both **Chapters 3** and **4** clearly demonstrated that this is very inefficient in *C. necator*. Therefore, a better understanding of the molecular mechanisms underlying homologous recombination in these strains, would undoubtedly unlock key elements improving recombineering in this system. Future studies could utilise different approaches to identify the native recombination system. These could include the identification of mutants with enhanced recombination (Clark, 1996), either recombination-deficient or hyper-recombinants, with an increased recombination frequency. Secondly, biochemical analysis of the enzymes/pathways identified in the mutant strains (substrate requirements, reaction conditions, etc.) could be utilised to build a recombination model (Dröge, 1992). Thirdly, bioinformatics analysis could be utilised to identify homologous recombination genes utilising known candidates (Swingle *et al*, 2010). In addition, transcriptomic analysis following the introduction of linear or covalently closed circular DNA, could provide information on the mRNA levels of native genes associated with recombination. Transcriptomic analysis could potentially reveal target genes for overexpression or downregulation to promote enhanced

recombination efficiency. Directed evolution could then be utilised to generate variants with improved enzyme activity.

The HTP genome editing tools evaluated in this thesis were intended to streamline the genetic engineering process in *Cupriavidus* and improve platform chemical production in the SBRC. Ethylene was chosen as an example to test the HTP editing tools and as a prerequisite ethylene productivity was assessed in *C. metallidurans.*

Ethylene is the most produced monomer in the chemical industry with demand increasing, ethylene derivatives include polyethylene and vinyl chloride, both heavily consumed in our modern petrol-based society. Currently the vast majority of ethylene is produced from stream-cracking of ethane or naphtha, this process is extremely detrimental to the environment and is associated with significant levels of CO_2 and other pollutants, greatly contributing to global warming. Ethylene production has been established in model organisms like *E. coli* (Lynch *et al*, 2016), *S. cerevisiae* (Pirkov *et al.*, 2008) and cyanobacteria (Puthan Veetil, Angermayr and Hellingwerf, 2017) (Sakai *et al.*, 1997). However, productivity remains low and requires further optimisation to achieve industrial scale. The work initiated in this study, and presented in **Chapter 5**, aimed to develop *Cupriavidus* spp. as a chassis for bio-ethylene production.

Ethylene productivity in *C. metallidurans* was significantly lower than anticipated (63 nmol/OD_{600}/mL), especially given ethylene productivity in *C. necator* (230 nmol/OD_{600}/mL). It was hypothesised that this could be due to tight regulation of the TCA cycle in *C. metallidurans* and therefore the subsequent availability of alphaketoglutarate and arginine. Given the difficulty in generating knockouts in *Cupriavidus* and the further development needed to exploit the HTP techniques assessed in **Chapter 3** and **4**, gTME and ALE were selected to improve ethylene productivity. Both

techniques were initially assessed in *E. coli*, the model host strain to assess validity of the approach.

gTME was combined for the first time with ALE fermentation, to improve ethylene production initially in *E. coli*, yielding a 2.3-fold increase in ethylene productivity, compared to the highest productivity reported to date (Lynch *et al*, 2016). The next aim was to transfer the methodology to *C. metallidurans*, to unlock the tight regulation around the TCA cycle and improve carbon fluxes through the EFE enzyme. A mutant library was generated for *C. metallidurans,* but time constraints prevented the library from being assessed in this background. This work is ongoing in the laboratory and could yield a significant increase in ethylene productivity in this strain. Thus, gTME and ALE techniques, along with CRISPR-based genome editing, could offer the novel genomic, transcriptomic and metabolic engineering tools to generate improved ethylene production in *Cupriavidus* species.

To exploit the full potential of the gTME and ALE techniques, it will be necessary to define a robust phenotypical screening method to improve the automation procedure and establish a comprehensive –omics analysis. The knowledge built from the analysis of mutant strains will support directed enzyme evolution of the EFE enzyme towards improved ethylene productivity. In that respect, CRISPR-enabled trackable genome engineering (CREATE) could be applied to reconstruct the adaptive evolution experiments in *E. coli*, both furthering the application of CRISPR and productivity of ethylene.

In summary, the implementation of genome, transcriptome and strain engineering serves to illustrate what synthetic and systems biology can deliver in terms of the optimisation of biotechnological processes. The development of sustainable routes for production of chemicals is essential. The HTP tools assessed and developed in this study have contributed

towards developing a toolkit for *Cupriavidus* spp. Recombineering utilising *recET* and the CRISPR system showed great promise, and with further optimisation, could provide a HTP platform for metabolic engineering in *Cupriavidus* spp*.* Furthermore, this study demonstrated the first production of ethylene in *C. metallidurans* and a 2.3 fold improvement of ethylene productivity in *E. coli* utilising gTME and ALE, approaches which can be applied to *Cupriavidus* to increase ethylene productivity in the SBRC.

References

Adams, D. O., Yang, S. F. and Stumpf, P. K. (1979) 'Ethylene biosynthesis: Identification of 1-aminocyclopropane-1-carboxylic acid as an intermediate in the conversion of methionine to ethylene', *Proc. Natl. Acad. Sci. USA*, 76(1), pp. 170–174.

Aik, W. S. *et al.* (2015) 'Introduction to Structural Studies on 2-Oxoglutarate-Dependent Oxygenases and Related Enzymes', pp. 59–94.

Alagesan, S., Minton, N. P. and Malys, N. (2018) '^{13}C-assisted metabolic flux analysis to investigate heterotrophic and mixotrophic metabolism in *Cupriavidus necator* H16', *Metabolomics*, 14(1), p. 9.

Alper, H. and Stephanopoulos, G. (2007) 'Global transcription machinery engineering: A new approach for improving cellular phenotype', *Metab. Eng.*, 9(3), pp. 258–267.

Alter, T. B., Blank, L. M. and Ebert, B. E. (2018) 'Determination of growth-coupling strategies and their underlying principles', *bioRxiv*, p. 258996.

Amghizar, I. *et al.* (2017) 'New Trends in Olefin Production', *Engineering*, 3(2), pp. 171–178.

Ampe, F. and Lindley, N. D. (1995) 'Acetate utilization is inhibited by benzoate in *Alcaligenes eutrophus*: evidence for transcriptional control of the expression of *acoE* coding for acetyl coenzyme A synthetase', *J. Bacteriol.*, 177(20), pp. 5826–5833.

Anders, C. *et al.* (2014) 'Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease', *Nature*, 513(7519), pp. 569–573.

Aparicio, T., de Lorenzo, V. and Martínez-García, E. (2018) 'CRISPR/Cas9-Based Counterselection Boosts Recombineering Efficiency in *Pseudomonas putida*', *Biotechnol. J.*, 13(5), p. 1700161.

Applebee, M. K., Herrgård, M. J. and Palsson, B. Ø. (2008) 'Impact of individual mutations on increased fitness in adaptively evolved strains of *Escherichia coli* .', *J. Bacteriol*., 190(14), pp. 5087–5094.

Aramvash, A. *et al.* (2015) 'Statistical physical and nutrient optimization of bioplastic polyhydroxybutyrate production by *Cupriavidus necator*'. *Int. J. Environ. Sci. Technol.,* 12(7), pp 2307-2316.

Arazoe, T., Kondo, A. and Nishida, K. (2018) 'Targeted Nucleotide Editing Technologies for MicrobialMet. Eng.', *Biotechnol. J*., 13(9), p. 1700596.

Arnaud, M., Chastanet, A. and Débarbouillé, M. (2004) 'New Vector for Efficient Allelic Replacement in Naturally Nontransformable, Low-GC-Content, Gram-Positive Bacteria', *Appl. Environ. Microbiol.*, 70(11), pp. 6887–6891.

Austin, D. and Larson, T. J. (1991) Nucleotide Sequence of the *glpD* Gene Encoding Aerobic sn-Glycerol 3-Phosphate Dehydrogenase of *Escherichia coli* K-12' *J. Bacteriol*., 173(1), pp 101–107.

Balderas-Hernández, V. E. *et al.* (2011) 'Adaptive Evolution of *Escherichia coli* Inactivated in the Phosphotransferase System Operon Improves Co-utilization of Xylose and Glucose Under Anaerobic Conditions', *Appl. Biochem. Biotech.*, 163(4), pp. 485–496.

Barquist, L. *et al.* (2016) 'The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries', *Bioinformatics*, 32(7), pp. 1109–1111.

Barrangou, R. *et al.* (2007) 'CRISPR provides acquired resistance against viruses in prokaryotes', *Science*, 315(5819), pp. 1709–1712.

Bassett, A. R. and Liu, J.-L. (2014) 'CRISPR/Cas9 and Genome Editing in *Drosophila*', *J. Genet. Genomics*, 41(1), pp. 7–19.

Bell, J. C. *et al.* (2012) 'Direct imaging of RecA nucleation and growth on

single molecules of SSB-coated ssDNA', *Nature*, 491(7423), pp. 274–278.

Bernhard, M. *et al.* (1996) The *Alcaligenes eutrophus* Membrane-Bound Hydrogenase Gene Locus Encodes Functions Involved in Maturation and Electron Transport Coupling', *Bacteriol*., 178(15), pp. 4522–4529.

Bertani, G. (1951) 'Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli* ', *J. Bacteriol*., 62(3), pp. 293–300.

Beuter, D. *et al.* (2018) 'Selective Enrichment of Slow-Growing Bacteria in a Metabolism-Wide CRISPRi Library with a TIMER Protein', *ACS Synth. Biol*., 7(12), pp. 2775–2782.

Bi, C. *et al.* (2013) 'Development of a broad-host synthetic biology toolbox for *Ralstonia eutropha* and its application to engineering hydrocarbon biofuel production', *Microb. Cell Fact.*, 12(1), p. 107.

Biebl, H. and Pfennig, N. (1981) 'Isolation of Members of the Family Rhodospirillaceae', in *The Prokaryotes*. Springer Berlin Heidelberg, pp. 267–273.

Boch, J. *et al.* (2009) 'Breaking the code of DNA binding specificity of TAL-type III effectors', *Science*, 326(5959), pp. 1509–1512.

Bogdanove, A. J. and Voytas, D. F. (2011) 'TAL Effectors: Customizable Proteins for DNA Targeting', *Science*, 333(6051), pp. 1843–1846.

Bonde, M. T. *et al.* (2014) 'MODEST: a web-based design tool for oligonucleotide-mediated genome engineering and recombineering', *Nucleic Acids Res.*, 42(1), pp. 408–415.

Bowater, R. and Doherty, A. J. (2006) 'Making Ends Meet: Repairing Breaks in Bacterial DNA by Non-Homologous End-Joining', *PLos Genet*., 2(2), p. 8.

Boyle, N. R. *et al.* (2013) 'Recombineering to homogeneity: extension of

multiplex recombineering to large-scale genome editing', *Biotech. J.*, 8(5), pp. 515–522.

Braun, V. (2009) 'FhuA (TonA), the career of a protein', *J. Bacteriol.*, 191(11), pp. 3431–3436.

Broeren, M. (2013) 'Production of Bio-ethylene', Available at https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2015/IRENA-ETSAP-Tech-Brief-I13-Production_of_Bio-ethylene.pdf

Brouns, S. J. J. *et al.* (2008) 'Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes', *Science*, 321(5891), pp. 960–964.

Budde, C. F. *et al.* (2011) 'Production of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) from plant oil by engineered *Ralstonia eutropha* strains', *Appl. Environ. Microbiol.*, 77(9), pp. 2847–2854.

Bujnicki, J. M. and Rychlewski, L. (2001) 'Identification of a PD-(D/E)XK-like domain with a novel configuration of the endonuclease active site in the methyl-directed restriction enzyme Mrr and its homologs', *Gene*, 267(2), pp. 183–191.

Burdett, V. *et al.* (2001) 'In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair', *Proc. Natl. Acad. Sci. USA*, 98(12), pp. 6765–6770.

Burgard, A. P., Pharkya, P. and Maranas, C. D. (2003) 'Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization', *Biotechnol. Bioeng.*, 84(6), pp. 647–657.

Burstein, D. *et al.* (2016) 'Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems', *Nat. Commun.*, 7, p. 10613.

Bütof, L. *et al.* (2017) 'The components of the unique Zur regulon of

Cupriavidus metallidurans mediate cytoplasmic zinc handling', *J. Bacteriol*., 199(21), pp. e00372-17.

Campo, N. *et al.* (2002) 'Cre-loxP Recombination System for Large Genome Rearrangements in *Lactococcus lactis*', *Appl. Environ. Microbiol.,* 68(5), pp. 2359–2367.

Carroll, D. (2011) 'Genome engineering with zinc-finger nucleases.', *Genetics*, 188(4), pp. 773–782.

Catoni, G. L. (1953) 'S-Adenosylmethionine; a new intermediate formed enzymatically from L-methionine and adenosinetriphosphate', *J. Biol. Chem.*, 204(1), pp. 403–416.

Chakravortty, D., Hansen-Wester, I. and Hensel, M. (2002) '*Salmonella* pathogenicity island 2 mediates protection of intracellular *Salmonella* from reactive nitrogen intermediates', *J. Exp. Med.*, 195(9), pp. 1155–1166.

Chatterjee, R. and Yuan, L. (2006) 'Directed evolution of metabolic pathways', *Trends Biotechnol*., 24(1), pp. 28–38.

Chaveroche, M. K., Ghigo, J. M. and d'Enfert, C. (2000) 'A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans*', *Nucleic Acids Res.*, 28(22), p. E97.

Chen, G.-Q. (2009) 'A microbial polyhydroxyalkanoates (PHA) based bio- and materials industry', *Chem. Soc. Rev*., 38(8), p. 2434.

Chen, G.-Q and Chen, Guo-Qiang (2010) 'Plastics from Bacteria: Natural Functions and Applications', *Microbiology Monographs*, 14, pp. 122-132.

Chen, J. S. *et al.* (2015) 'Production of fatty acids in *Ralstonia eutropha* H16 by engineering β-oxidation and carbon storage.', *PeerJ*., 3, p. 1468.

Chen, Q. *et al.* (2008) 'In vitro CpG methylation increases the transformation

efficiency of *Borrelia burgdorferi* strains harboring the endogenous linear plasmid', *J. Bacteriol.*, 190(24), pp. 7885–91.

Chen, W. *et al.* (2018) 'CRISPR/Cas9-based Genome Editing in *Pseudomonas aeruginosa* and Cytidine Deaminase-Mediated Base Editing in *Pseudomonas* Species.', *iScience*. Elsevier, 6, pp. 222–231.

Chen, X. *et al.* (2010) 'Overexpression of bacterial ethylene-forming enzyme gene in *Trichoderma reesei* enhanced the production of ethylene.', *Int. J. Biol. Sci*, 6(1), pp. 96–106.

Cheng, A. W. *et al.* (2013) 'Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system', *Cell Res.*, 23(10), pp. 1163–1171.

Chepuri, V. *et al.* (1990) 'Recent studies of the cytochrome o terminal oxidase complex of *Escherichia coli* ', *Biochim. Biophys. Acta, Biophys.* 1018(2–3), pp. 124–127.

Cho, J. S. *et al.* (2017) 'CRISPR/Cas9-coupled recombineering forMet. Eng. of *Corynebacterium glutamicum*.',Met. Eng., 42, pp. 157–167.

Choi, K.-H. *et al.* (2005) 'A Tn*7*-based broad-range bacterial cloning and expression system', *Nat. Methods* 2(6), pp. 443–448.

Choi, S. *et al.* (2015) 'Biorefineries for the production of top building block chemicals and their derivatives',*Met. Eng..* 28, pp. 223–239.

Clark, A. J. *et al.* (1984) 'Genes of the RecE and RecF Pathways of Conjugational Recombination in *Escherichia coli* ', *Cold Spring Harbor Symposia on Quantitative Biology*, 49(0), pp. 453–462.

Clark, A. J. (1996) '*recA* mutants of *E. coli* K12: A personal turning point', *BioEssays*, 18(9), pp. 767–772.

Clark, A. J., Satin, L. and Chu, C. C. (1994) 'Transcription of the *Escherichia coli recE* gene from a promoter in Tn*5* and IS50.', *J. Bacteriol.,* 176(22), pp. 7024–31.

Cobb, R. E., Wang, Y. and Zhao, H. (2015) 'High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System', *ACS Synth. Biol*, 4(6), pp. 723–728.

Collard, J. M. *et al.* (1994) 'Plasmids for heavy metal resistance in *Alcaligenes eutrophus* CH34: Mechanisms and applications', *FEMS Microbiol. Rev.*. 14(4), pp. 405–414.

Cong, L. *et al.* (2013) 'Multiplex Genome Engineering Using CRISPR/Cas Systems', *Science*, 339(6121), pp. 819–823.

Cook, A. M. and Schlegel, H. G. (1978) 'Metabolite concentrations in *Alcaligenes eutrophus* H16 and a mutant defective in poly-3-hydroxybutyrate synthesis', *Arch. Microbiol*., 119(3), pp. 231–235.

Cox, D. B. T. T. *et al.* (2017) 'RNA editing with CRISPR-Cas13', *Science*, 6366(358), pp. 1019-1027.

Cramm, R. (2009) 'Genomic view of energy metabolism in *Ralstonia eutropha* H16.', *J. Mol. Microbiol. Biotechnol.,* 16(1–2), pp. 38–52.

D'Agostino, Y. *et al.* (2016) 'A Rapid and Cheap Methodology for CRISPR/Cas9 Zebrafish Mutant Screening', *Mol. Biotechnol.,* 58(1), pp. 73–78.

Datsenko, K. A. and Wanner, B. L. (2000) 'One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products', *Proc. Natl. Acad. Sci. U.S.A* 97(12), pp. 6640–6645.

Datta, S. *et al.* (2008) 'Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their

phages.', *Proc. Natl. Acad. Sci. U.S.A.,* 105(5), pp. 1626–31.

Davis, D. H. *et al.* (1969) 'Proposal to reject the genus *Hydrogenomonas*: Taxonomic implications', *Int J Syst Bacteriol*., 19(4), pp. 375–390.

Delamarre, S. C. and Batt, C. A. (2006) 'Comparative study of promoters for the production of polyhydroxyalkanoates in recombinant strains of *Wautersia eutropha'*, *Appl. Microbiol. Biotechnol.*, 71(5), pp. 668–679.

Della, M. *et al.* (2004) 'Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine.', *Science,* 306(5696), pp. 683–685.

Derbise, A. *et al.* (2003) 'A rapid and simple method for inactivating chromosomal genes in *Yersinia*', *FEMS Immunol. Med. Microbiol.* 38(2), pp. 113–116.

Đermić, D. *et al.* (2005) 'RecBCD enzyme overproduction impairs DNA repair and homologous recombination in *Escherichia coli '*, *Res. Microbiol.,* 156(3), pp. 304–311.

Deveau, H. *et al.* (2008) 'Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*.', *J. Bacteriol.,* 190(4), pp. 1390–400.

DiCarlo, J. E. *et al.* (2013) 'Yeast Oligo-Mediated Genome Engineering (YOGE)', *ACS Synth. Biol.,* 2(12), pp. 741–749.

Diels, L. *et al.* (2009) 'From industrial sites to environmental applications with *Cupriavidus metallidurans '*, *Antonie van Leeuwenhoek*, 96(2), pp. 247–258.

Ding, D. *et al.* (2018) 'Engineering Introns to Express RNA Guides for Cas9- and Cpf1-Mediated Multiplex Genome Editing', *Mol. Plant.* 11(4), pp. 542–552.

Doench, J. G. *et al.* (2016) 'Optimized sgRNA design to maximize activity and

minimize off-target effects of CRISPR-Cas9', *Nat. Biotechnol.,* 34(2), pp. 184–191.

Doudna, J. A. and Charpentier, E. (2014) 'The new frontier of genome engineering with CRISPR-Cas9', *Science*, 346(6213), pp. 1258096–1258096.

Dröge, P. (1992) 'Recombination of nicked DNA knots by gamma delta resolvase suggests a variant model for the mechanism of strand exchange.', *Nucleic Acids Res.,* 20(23), pp. 6159–6166.

Dugar, G. *et al.* (2013) 'High-Resolution Transcriptome Maps Reveal Strain-Specific Regulatory Features of Multiple *Campylobacter jejuni* Isolates', *PLos Genet.* 9(5), p. 1003495.

Echols, H. and Gingery, R. (1968) 'Mutants of bacteriophage λ defective in vegetative genetic recombination', *J. Mol. Biol.,* 34(2), pp. 239–249.

Eckert, C. *et al.* (2014) 'Ethylene-forming enzyme and bioethylene production', *Biotechnol. Biofuels,* 7(1), p. 33.

Ellis, H. M. *et al.* (2001) 'High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides', *Proc. Natl. Acad. Sci. U.S.A.,* 98(12), pp. 6742–6746.

Ely, S. and Wright, A. (1985) 'Maintenance of plasmid pSC101 in *Escherichia coli* requires the host primase.', *J. Bacteriol.,* 164(1), pp. 484–486.

Engler M.J. and Richardson, C. C. (1982) Paul D. Boyer (Ed.) The Enzymes, Vol. XV, pp. 3-29.

Enyeart, P. J. *et al.* (2014) 'Generalized bacterial genome editing using mobile group II introns and Cre-lox', *Mol. Syst. Biol.,* 9(1), pp. 685–685.

Eramo, M. (2012) Global Ethylene Market Outlook: Low Cost Feedstocks

Fuel The Next Wave Of Investments In North America and China. Available at: http://media.corporate-ir.net/media_files/IROL/11/110877/05_Global_Ethylene_Market_Outlook_Eramo.pdf

Eschenbrenner, M., Covès, J. and Fontecave, M. (1995) 'The flavin reductase activity of the flavoprotein component of sulfite reductase from *Escherichia coli* . A new model for the protein structure.', *J. Biol. Chem.*, 270(35), pp. 20550–20555.

Fan, D., Dai, D.-J. and Wu, H.-S. (2012) 'Ethylene Formation by Catalytic Dehydration of Ethanol with Industrial Considerations', *Materials*, 6(1), pp. 101–115.

Feliciello, I. and Chinali, G. (1993) 'A Modified Alkaline Lysis Method for the Preparation of Highly Purified Plasmid DNA from *Escherichia coli* ', *Anal. Biochem.* 212(2), pp. 394–401.

Felton, J. and Wright, A. (1979) 'Plasmid pSC101 replication in integratively suppressed cells requires *dnaA* function.', *Mol. Gen. Genet.,* 175(2), pp. 231–233.

Fleige, C., Kroll, J. and Steinbüchel, A. (2011) 'Establishment of an alternative phosphoketolase-dependent pathway for fructose catabolism in *Ralstonia eutropha* H16', *Appl. Microbiol. Biotechnol., 91*(3), pp. 769–776.

Fong, S. S. *et al.* (2005) 'In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid', *Biotechnol. Bioeng.*, 91(5), pp. 643–648.

Fricke, W. F., Kusian, B. and Bowien, B. (2009) 'The Genome Organization of *Ralstonia eutropha* Strain H16 and Related Species of the Burkholderiaceae', *J. Mol. Microbiol. Biotechnol.*, 16, pp. 124–135.

Fritz, G. *et al.* (2014) 'Single Cell Kinetics of Phenotypic Switching in the

Arabinose Utilization System of *E. coli* ', *PLoS ONE*. 9(2), p. 89532.

Fukuda, H. *et al.* (1989) 'Ethylene production from L-methionine by *Cryptococcus albidus*', *J. Ferment. Bioeng.,* 67(3), pp. 173–175.

Fukuda, H. *et al.* (1992) 'Two reactions are simultaneously catalyzed by a single enzyme: the arginine-dependent simultaneous formation of two products, ethylene and succinate, from 2-oxoglutarate by an enzyme from *Pseudomonas syringae*.', *Biochem. Biophys. Res. Commun.,* 188(2), pp. 483–489.

Fukuda, H., Fujii, T. and Ogawa, T. (1986) 'Preparation of a Cell-free Ethylene-forming System from *Penicillium digitatum*', *Agric. Biol. Chem.,* 50(4), pp. 977–981.

Fukui, T. *et al.* (2011) 'Evaluation of promoters for gene expression in polyhydroxyalkanoate-producing *Cupriavidus necator* H16', *Appl. Microbiol. Biotechnol.*, 89(5), pp. 1527–1536.

Fukui, T. and Doi, Y. (1997) Cloning and Analysis of the Poly(3-Hydroxybutyrate-co-3-Hydroxyhexanoate) Biosynthesis Genes of *Aeromonas caviae*, *J. Bacteriol.,* 179(15), pp. 4821–4830.

Fulco, C. P. *et al.* (2016) 'Systematic mapping of functional enhancer-promoter connections with CRISPR interference.', *Science,* 354(6313), pp. 769–773.

Gai, C. S. *et al.* (2014) 'Insights into bacterial CO_2 metabolism revealed by the characterization of four carbonic anhydrases in *Ralstonia eutropha* H16', *AMB Express*. 4(1), p. 2.

Gaj, T., Gersbach, C. A. and Barbas C. F. (2013) 'ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering'. *Trends Biotechnol.*, 31(7): 397–405.

Gao, Z. *et al.* (2003) 'Localization of the Raf-like kinase CTR1 to the endoplasmic reticulum of *Arabidopsis* through participation in ethylene receptor signaling complexes', *J. Biol. Chem.*, 278(36), pp. 34725–32.

Garst, A. D. *et al.* (2016) 'Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering', *Nat. Biotechnol.*, 35(1), pp. 48–55.

Gasiunas, G. *et al.* (2012) 'Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.', *Proc. Natl. Acad. Sci. U.S.A*. 109(39), pp. 2579-86.

Gay, P. *et al.* (1983) 'Cloning structural gene *sacB*, which codes for exoenzyme levansucrase of *Bacillus subtilis*: expression of the gene in *Escherichia coli* .', *J. Bacteriol.*, 153(3), pp. 1424–31.

Generoso, W. C. *et al.* (2016) 'Simplified CRISPR-Cas genome editing for *Saccharomyces cerevisiae*', *J. Microbiol. Methods,* 127, pp. 203–205.

Gentz, R. and Bujard, H. (1985) 'Promoters recognized by *Escherichia coli* RNA polymerase selected by function: highly efficient promoters from bacteriophage T5.', *J. Bacteriol.*, 164(1), pp. 70–77.

Gerdes, K. *et al.* (1986) 'Mechanism of postsegregational killing by the hok gene product of the *parB* system of plasmid R1 and its homology with the *relF* gene product of the *E. coli relB* operon', *EMBO J.,* 5(8), pp. 2023–2029.

Gerdes, K., Thisted, T. and Martinussen, J. (1990) 'Mechanism of post-segregational killing by the *hok/sok* system of plasmid R1: *sok* antisense RNA regulates formation of a *hok* mRNA species correlated with killing of plasmid-free cells', *Mol. Microbiol.,* 4(11), pp. 1807–1818.

Ghanta, M., Fahey, D. and Subramaniam, B. (2014) 'Environmental impacts of ethylene production from diverse feedstocks and energy sources', *Appl.*

Petrochem. Res., 4(2), pp. 167–179.

Ghysels, S. *et al.* (2018) 'Targeted poly(3-hydroxybutyrate-co-3-hydroxyvalerate) bioplastic production from carbon dioxide', *Bioresour. Technol.* 249, pp. 858–868.

Giacalone, M. J. *et al.* (2006) 'Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system', *BioTechniques*, 40(3), pp. 355–364.

Gorenflo, V. *et al.* (1999) 'Quantification of bacterial polyhydroxyalkanoic acids by Nile red staining', *Appl. Microbiol. Biotechnol.*, 51(6), pp. 765–772.

Goris, J. *et al.* (2001) 'Classification of metal-resistant bacteria from industrial biotopes as *Ralstonia campinensis* sp. *nov*., *Ralstonia metallidurans* sp. nov. and *Ralstonia basilensis*', *Int. J. Syst. Evol. Microbiol*., 51(5), pp. 1773–1782.

Gorski, S. A., Vogel, J. and Doudna, J. A. (2017) 'RNA-based recognition and targeting: sowing the seeds of specificity', *Nat. Rev. Mol. Cell Biol.*, 18(4), pp. 215–228.

Goto, M. *et al.* (1985) 'Ethylene Production by the Kudzu Strains of *Pseudomonas syringae* pv. *phaseolicola* Causing Halo Blight in *Pueraria lobata*', *Plant Cell Physiol.,* 26(1), pp. 141–150.

Goto, M. and Hyodo, H. (1987) 'Ethylene Production by Cell-Free Extract of the Kudzu Strain of *Pseudomonas syringae* pv. *phaseolicola', Plant Cell Physiol.,* 28(3), pp. 405–414.

Grosse, C., Friedrich, S. and Nies, D. H. (2007) 'Contribution of Extracytoplasmic Function Sigma Factors to Transition Metal Homeostasis in *Cupriavidus metallidurans* Strain CH34', *J. Mol. Microbiol. Biotechnol.*, 12(3–4), pp. 227–240.

Gruber, S. *et al.* (2014) 'Versatile and stable vectors for efficient gene expression in *Ralstonia eutropha* H16', *Int. J. Biotechnol.*, 186, pp. 74–82.

Guerrero, F. *et al.* (2012) 'Ethylene Synthesis and Regulated Expression of Recombinant Protein in *Synechocystis* sp. PCC 6803', *PLoS ONE*. 7(11), p. 50470.

Guilinger, J. P. *et al.* (2014) 'Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity', *Nat. Methods,* 11(4), pp. 429–435.

Gupta, R. M. and Musunuru, K. (2014) 'Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9', *J Eur. J. Clin. Invest.,* 124(10), pp. 4154–4161.

Guzman, L. M. *et al.* (1995) 'Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter.', *J. Bacteriol.*, 177(14), pp. 4121–4130.

Hall, S. D., Kane, M. F. and Kolodner, R. D. (1993) 'Identification and characterization of the *Escherichia coli* RecT protein, a protein encoded by the *recE* region that promotes renaturation of homologous single-stranded DNA.', *J. Bacteriol.,* 175(1), pp. 277–287.

Han, S. O., Inui, M. and Yukawa, H. (2008) 'Effect of carbon source availability and growth phase on expression of *Corynebacterium glutamicum* genes involved in the tricarboxylic acid cycle and glyoxylate bypass', *Microbiology*, 154(10), pp. 3073–3083.

Hanahan, D. (1983) 'Studies on transformation of *Escherichia coli* with plasmids', *J. Mol. Biol.,* 166(4), pp. 557–580.

Hanko, E. K. and Minton, N. P. (2017) 'Characterisation of a 3-hydroxypropionic acid-inducible system from *Pseudomonas putida* for

orthogonal gene expression control in *Escherichia coli* and *Cupriavidus necator*' *Sci. Rep*., 1724(7).

Harder, B.-J., Bettenbrock, K. and Klamt, S. (2016) 'Model-based metabolic engineering enables high yield itaconic acid production by *Escherichia coli* ', *Met. Eng.*, 38, pp. 29–37.

Harmsen, P. F. H., Hackmann, M. M. and Bos, H. L. (2014) 'Green building blocks for bio-based plastics', *Biofuels, Bioprod. Biorefin.,* 8(3), pp. 306–324.

Haurwitz, R. E. *et al.* (2010) 'Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease', *Science*. 329(5997), pp. 1355–1358.

Hausinger, R. P. (2015) 'Biochemical Diversity of 2-Oxoglutarate-Dependent Oxygenases' from '*2-Oxoglutarate-Dependent Oxygenases'.* Edited by Christopher Schofield and Robert Hausinger, pp. 1-40.

He, C. and You, F. (2015) 'Toward more cost-effective and greener chemicals production from shale gas by integrating with bioethanol dehydration: Novel process design and simulation-based optimization', *AlChE J.* 61(4), pp. 1209–1232.

Heeb, S. *et al.* (2000) 'Small, Stable Shuttle Vectors Based on the Minimal pVS1 Replicon for Use in Gram-Negative, Plant-Associated Bacteria', *Mol. Plant-Microbe Interact.*, 13(2), pp. 232–237.

Heermann, R., Zeppenfeld, T. and Jung, K. (2008) 'Simple generation of site-directed point mutations in the *Escherichia coli* chromosome using Red/ET Recombination', *Microb. Cell Fact.,* 7(1), p. 14.

Hein, S. and Steinbüchel, A. (1996) 'Cloning and characterization of the *Alcaligenes eutrophus* 2-oxoglutarate dehydrogenase complex.', *FEMS Microbiol. Lett.*, 136(3), pp. 231–238.

Hiom, K. (2009) 'DNA Repair: Common Approaches to Fixing Double-Strand

Breaks', *Current Biology*, 19(13), pp. 523–525.

Hobson, N. *et al.* (2008) 'Generation of a restriction minus enteropathogenic *Escherichia coli* E2348/69 strain that is efficiently transformed with large, low copy plasmids', *BMC Microbiol.,* 8, p. 134.

Hockemeyer, D. *et al.* (2011) 'Genetic engineering of human pluripotent cells using TALE nucleases', *Nat. Biotechnol*, 29(8), pp. 731–734.

Hossain, M. J. *et al.* (2015) 'Genome modifications and cloning using a conjugally transferable recombineering system', *Biotechnol. Rep.,* 8, pp. 24–35.

Hsu, P. D., Lander, E. S. and Zhang, F. (2014) 'Development and applications of CRISPR-Cas9 for genome engineering', *Cell*. 157(6), pp. 1262–1278.

Hu, K. *et al.* (2003) 'Study on gene knockout using red system in *Shigella flexneri*', *Wei sheng wu xue bao = Acta microbiologica Sinica*, 43(6), pp. 740–746.

Hu, S. *et al.* (2014) 'Genome engineering of *Agrobacterium tumefaciens* using the lambda Red recombination system', *Appl. Microbiol. Biotechnol.*, 98(5), pp. 2165–2172.

Hua, J. *et al.* (1998) 'EIN4 and ERS2 Are Members of the Putative Ethylene Receptor Gene Family in *Arabidopsis*', *Plant Cell,* 10(8); pp. 1321–1332.

Huang, H. *et al.* (2015) 'One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*', *Acta Biochim. Biophys. Sin.,* 47(4), pp. 231–243.

Huang, L. C., Wood, E. A. and Cox, M. M. (1997) 'Convenient and reversible site-specific targeting of exogenous DNA into a bacterial chromosome by use of the FLP recombinase: the FLIRT system', *J. Bacteriol.*, 179(19), pp. 6076–6083.

Huang, Y. *et al.* (2003) 'Biochemical and functional analysis of CTR1, a protein kinase that negatively regulates ethylene signaling in *Arabidopsis*', *Plant J.,* 33(2), pp. 221–233.

Huang, Y. *et al.* (2017) 'Recombineering using RecET in *Corynebacterium glutamicum* ATCC14067 via a self-excisable cassette', *Sci. Rep*., 7(1), p. 7916.

Huber G., Iborra S. and Corma A. (2006) 'Synthesis of Transportation Fuels from Biomass: Chemistry, Catalysts, and Engineering', *Chem. Rev*., 106(9), pp. 4044-4098.

Huergo, L. F. and Dixon, R. (2015) 'The Emergence of 2-Oxoglutarate as a Master Regulator Metabolite', *Microbiol. Mol. Biol. Rev.,* 79(4), pp. 419-435.

Im, D.-K. *et al.* (2017) 'Comparison of metabolite profiling of *Ralstonia eutropha* H16 *phaBCA* mutants grown on different carbon sources', *Korean J. Chem. Eng*, 34(3), pp. 797–805.

Ince, J. E. and Knowles, C. J. (1986) 'Ethylene formation by cell-free extracts of *Escherichia coli* ', *Arch. Microbiol.,* 146(2), pp. 151–158.

Ishihara, K. *et al.* (1996) 'Ethylene Production Using a Broad-Host-Range Plasmid in *Pseudomonas syringae* and *Pseudomonas putida*', *J. Ferment. Bioeng.,* 82(5), pp. 509-511.

Ishino, Y. *et al.* (1987) 'Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product', *J. Bacteriol.*, 169(12), pp. 5429–5433.

Ishino, Y., Krupovic, M. and Forterre, P. (2018) 'History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology', *J. Bacteriol.*. 200(7).

Janssen, P. J. *et al.* (2010) 'The Complete Genome Sequence of *Cupriavidus metallidurans* Strain CH34, a Master Survivalist in Harsh and

Anthropogenic Environments', *PLoS ONE*. 5(5), p. 10433.

Jasin, M. (1996) 'Genetic manipulation of genomes with rare-cutting endonucleases', *Trends Genet.* 12(6), pp. 224–228.

Jia, K. *et al.* (2011) 'Group II Intron-Anchored Gene Deletion in *Clostridium*', *PLoS ONE*. 6(1), p. 16693.

Jiang, W. *et al.* (2014) 'Successful Transient Expression of Cas9 and Single Guide RNA Genes in *Chlamydomonas reinhardtii*', *Eukaryotic Cell*, 13(11), pp. 1465–1469.

Jiang, Y. *et al.* (2015) 'Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system', *Appl. Environ. Microbiol.,* 81(7), pp. 2506–2514.

Jiang, Y. *et al.* (2017) 'CRISPR-Cpf1 assisted genome editing of *Corynebacterium glutamicum*', *Nat. Commun.,* 8, p. 15179.

Jinek, M. *et al.* (2012) 'A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity', *Science*, 337(6096), pp. 816–821.

Johansson, N. (2014) 'A study of ethylene production via the 2-oxoglutarate dependent pathway in *S. cerevisiae* - insights from enzyme engineering, metabolic modeling and cultivation studies', *eTheses*.

Johnson, A. O. *et al.* (2018) 'An Engineered Constitutive Promoter Set with Broad Activity Range for*Cupriavidus necator* H16', *ACS Synth. Biol*., 7(8), pp. 1918–1928.

Jugder, B.-E. *et al.* (2015) 'An analysis of the changes in soluble hydrogenase and global gene expression in *Cupriavidus necator* (*Ralstonia eutropha*) H16 grown in heterotrophic diauxic batch culture', *Microb. Cell Fact.,* 14(1), p. 42.

Jugder, B.-E. *et al.* (2016) 'Construction and use of a *Cupriavidus necator* H16 soluble hydrogenase promoter (PSH) fusion to *gfp* (green fluorescent protein).', *PeerJ*. 4, p. 2269.

Julian, D. J. *et al.* (2009) 'Transcriptional activation of MerR family promoters in *Cupriavidus metallidurans* CH34', *Antonie van Leeuwenhoek*, 96(2), pp. 149–159.

Jung, M.-Y. *et al.* (2012) 'Deletion of lactate dehydrogenase in *Enterobacter aerogenes* to enhance 2,3-butanediol production', *Appl. Microbiol. Biotechnol.*, 95(2), pp. 461–469.

Kaiser, A. D. (1955) 'A genetic study of the temperate coli phage.', *Virology*, 1(4), pp. 424–443.

von Kamp, A. and Klamt, S. (2017) 'Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms', *Nat. Commun.* 8, p. 15956.

Kang, S. and Yu, J. (2015) 'Reaction routes in catalytic reforming of poly(3-hydroxybutyrate) into renewable hydrocarbon oil', *RSC Adv.,* 5(38), pp. 30005–30013.

Karberg, M. *et al.* (2001) 'Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria', *Nat. Biotechnol.*, 19(12), pp. 1162–1167.

Karlinsey, J. E. (2007) 'λ-Red Genetic Engineering in *Salmonella enterica* serovar Typhimurium', *Methods Enzymol.,* 421, pp. 199–209.

Kee, R. J., Karakaya, C. and Zhu, H. (2017) 'Process intensification in the catalytic conversion of natural gas to fuels and chemicals', *Proc. Combust. Inst.,* 36(1), pp. 51–76.

Kelly, J. R. *et al.* (2009) 'Measuring the activity of BioBrick promoters using

an in vivo reference standard', *J. Biol. Eng.,* 3(1), p. 4.

Kennaway, C. K. *et al.* (2012) 'Structure and operation of the DNA-translocating type I DNA restriction enzymes.', *Genes Dev.,* 26(1), pp. 92–104.

Khasanov, F. K. *et al.* (1992) 'Homologous recombination between plasmid and chromosomal DNA in *Bacillus subtilis* requires approximately 70 bp of homology', *Mol. Gen. Genet.,* 234(3), pp. 494–497.

Kim, J. *et al.* (2014) 'A versatile and highly efficient method for scarless genome editing in *Escherichia coli* and *Salmonella enterica*', *BMC Biotech.,* 14(1), p. 84.

Kim, S. K. *et al.* (2017) 'CRISPR interference-guided multiplex repression of endogenous competing pathway genes for redirecting metabolic flux in *Escherichia coli*', *Microb. Cell Fact.,* 16(1), p. 188.

Kleihues, L. *et al.* (2000) 'The H_2 sensor of *Ralstonia eutropha* is a member of the subclass of regulatory [NiFe] hydrogenases.', *J. Bacteriol.*, 182(10), pp. 2716–2724.

Klein-Marcuschamer, D. and Stephanopoulos, G. (2008) 'Assessing the potential of mutational strategies to elicit new phenotypes in industrial strains', *Proc. Natl. Acad. Sci. U.S.A.*, 105(7), pp. 2319–2324.

Kniel, L., Winter, O. and Stork, K. (1980) 'Ethylene, keystone to the petrochemical industry.' *Chemie Ingenieur Technik*, 53(2), p. 116.

Kolodner, R., Hall, S. D. and Luisi-DeLuca, C. (1994) 'Homologous pairing proteins encoded by the *Escherichia coli recE* and *recT* genes.', *Mol. Microbiol.*, 11(1), pp. 23–30.

Koppolu, V. and Vasigala, V. K. (2016) 'Role of *Escherichia coli* in Biofuel Production.', *Microbiology insights*, 9, pp. 29–35.

Kourmentza, C. *et al.* (2017) 'Recent Advances and Challenges towards Sustainable Polyhydroxyalkanoate (PHA) Production', *Bioengineering*, 4(4), p. 55.

Kramer M.G. (2016) 'Determination of Plasmid Segregational Stability in a Growing Bacterial Population', In: Hoffman R. (eds) *Bacterial Therapy of Cancer*. *Methods in Molecular Biology,* vol 1409. Humana Press, New York, NY

Kufryk, G. I. *et al.* (2002) 'Transformation of the cyanobacterium *Synechocystis* sp. PCC 6803 as a tool for genetic mapping: optimization of efficiency', *FEMS Microbiol. Let.,* 206(2), pp. 215–219.

Kuijpers, N. G. A. *et al.* (2013) 'One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in *Saccharomyces cerevisiae*.', *FEMS Yeast Res.* 13(8), pp. 769–781.

Kumar, A. *et al.* (1994) 'Role of the Sigma 70 Subunit of *Escherichia coli* RNA Polymerase in Transcription Activation', *J. Mol. Biol.*235(2), pp. 405–413.

Kuzminov, A. (1999) 'Recombinational Repair of DNA Damage in *Escherichia coli* and Bacteriophage λ', *Microbiol. Mol. Biol. Rev.* 63(4), pp. 751–813.

Lacey, R. F. and Binder, B. M. (2014) 'How plants sense ethylene gas — The ethylene receptors', *J. Inorg. Biochem.,*133, pp. 58–62.

Lacey, R. F. and Binder, B. M. (2016) 'Ethylene Regulates the Physiology of the Cyanobacterium *Synechocystis* sp. PCC 6803 via an Ethylene Receptor', *Plant Physiol.,* 171(4), pp. 2798–2809.

Lal, D. *et al.* (2013) '*Cupriavidus metallidurans* : A Modern Alchemist.', *Indian J. Microbiol.,* 53(1), pp. 114–115.

Lambowitz, A. M. and Zimmerly, S. (2011) 'Group II introns: mobile

ribozymes that invade DNA.', *Cold Spring Harb Perspect. Biol.,* 3(8), p. 003616.

Larson, M. H. *et al.* (2013) 'CRISPR interference (CRISPRi) for sequence-specific control of gene expression', *Nat Protoc.,* 8(11), pp. 2180–2196.

De Las Peñas, A., Connolly, L. and Gross, C. A. (1997) 'SigmaE is an essential sigma factor in *Escherichia coli*', *J. Bacteriol.*, 179(21), pp. 6862–6864.

Lee, D.-H. and Palsson, B. Ø. (2010) 'Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a nonnative carbon source, L-1,2-propanediol', *Appl. Environ. Microbiol.* 76(13), pp. 4158-4168.

Lee, J. *et al.* (2007) 'Structure and Function of the *Escherichia coli* Protein YmgB: A Protein Critical for Biofilm Formation and Acid-resistance', *J. Mol. Biol.*, 373(1), pp. 11–26.

Lee, Y.-H., Park, J.-S. and Huh, T.-L. (1997) 'Enhanced biosynthesis of P(3HB-3HV) and P(3HB-4HB) by amplification of.the cloned PHB biosynthesis genes in *Alcaligenes eutrophus*', *Biotechnol. Lett.,* 19(8), pp. 771–774.

Lenz, O. *et al.* (1994) 'The *Alcaligenes eutrophus* H16 *hoxX* gene participates in hydrogenase regulation', *J. Bacteriol.*, 176(14), pp. 4385–4393.

Lenz, O. and Friedrich, B. (1998) 'A novel multicomponent regulatory system mediates H_2 sensing in *Alcaligenes eutrophus*', *Proc. Natl. Acad. Sci. U.S.A.*, 5(21), pp. 12474–12479.

Lenz, O., Lauterbach, L. and Frielingsdorf, S. (2018) 'O_2-tolerant [NiFe]-hydrogenases of *Ralstonia eutropha* H16: Physiology, molecular biology, purification, and biochemical analysis', *Methods Enzymol*. 613, pp. 117–151.

Lesic, B. and Rahme, L. G. (2008) 'Use of the lambda Red recombinase system to rapidly generate mutants in *Pseudomonas aeruginosa*', *BMC Mol. Biol.,* 9(1), p. 20.

Lewandowski, S. (2016) Presentation from the Asia Chemical Conference Ethylene-Global. Available at: https://cdn.ihs.com/www/pdf/Steve-Lewandowski-Big-Changes-Ahead-for-Ethylene-Implications-for-Asia.pdf

Li *et al.* (2018) 'CRISPR-Cpf1-Assisted Multiplex Genome Editing and Transcriptional Repression in *Streptomyces*', *Appl. Environ. Microbiol.*, 84(18), p. 18.

Li, B. *et al.* (2018) 'Enhanced synthesis of poly gamma glutamic acid by increasing the intracellular reactive oxygen species in the *Bacillus licheniformis* Δ1-pyrroline-5-carboxylate dehydrogenase gene *ycgN*-deficient strain', *Appl. Microbiol. Biotechnol.*, 123(102), pp. 10127–10137.

Li, J. *et al.* (2017) 'Ethylene', In: *Hormone Metabolism and Signaling in Plants*, London: Academic Press, pp. 203–241.

Li, K. *et al.* (2014) 'Optimization of Genome Engineering Approaches with the CRISPR/Cas9 System', *PLoS ONE*. 9(8), p. 105779.

Li, M. *et al.* (2018) 'Thermodynamics of iron(II) and substrate binding to the ethylene-forming enzyme', *Biochemistry*. 57(39), pp. 5696-5705.

Li, Y. *et al.* (2015) 'Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 meditated genome editing',*Met. Eng.*, 31, pp. 13–21.

Liboni, L. B. (2012) 'Social and environmental impacts of the sugarcane industry', In: *Profuturo: future studies research journal: trends and strategies*. 4(1), pp. 196-227. Available at: https://revistafuture.org/FSRJ/article/viewFile/101/168.

Lieber, M. R. (2010) 'The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway', *Annu. Rev. Biochem.,* 79, pp. 181–211.

Lieberman, M. *et al.* (1965) 'Ethylene production from methionine',

Biochem. J., 97(2), pp. 449–59.

Lim, S. I., Min, B. E. and Jung, G. Y. (2008) 'Lagging Strand-Biased Initiation of Red Recombination by Linear Double-Stranded DNAs', *J. Mol. Biol.* 384(5), pp. 1098–1105.

Lin, E. (1987) 'Dissimilatory pathways for sugars, polyols, and carboxylates', In: Frederick C. Neidhardt, R. Curtiss, J. l. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. Reznikoff, M. Riley, M. Schaechter, H. E. Umbarger (Editors), Escherichia coli and Salmonella: cellular and molecular biology, 2nd ed., 1, pp. 307–342.

Lindenkamp, N., Volodina, E. and Steinbüchel, A. (2012) 'Genetically modified strains of *Ralstonia eutropha* H16 with β-ketothiolase gene deletions for production of copolyesters with defined 3-hydroxyvaleric acid contents', *Appl. Environ. Microbiol.* 78(15), pp. 5375–5383.

de Lorenzo, V. *et al.* (1990) 'Mini-Tn*5* transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria', *J. Bacteriol.*, 172(11), pp. 6568–6572.

Lovett, S. T. (2012) 'A glimpse of molecular competition', *Nature*, 491(7423), pp. 198–200.

Lovett, S. T., Kolodner, R. D. and Lovett, Susan T. (1989) 'Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the *recJ* gene of *Escherichia coli* ', *Proc. Natl. Acad. Sci. U.S.A* National Academy of Sciences, 86(8), pp. 2627–2631.

Lu, J. *et al.* (2012) 'Studies on the production of branched-chain alcohols in engineered *Ralstonia eutropha*', *Appl. Microbiol. Biotechnol*., 96(1), pp. 283–297.

Lu, S. *et al.* (2002) 'The global regulator ArcA controls resistance to reactive

nitrogen and oxygen intermediates in *Salmonella enterica* serovar Enteritidis', *Infect. Immun.* 70(2), pp. 451–461.

Lütte, S. *et al.* (2012) 'Autotrophic production of stable-isotope-labeled arginine in *Ralstonia eutropha* strain H16', *Appl. Environ. Microbiol.*, 78(22), pp. 7884–7890.

Lynch, S. *et al.* (2016) 'Overcoming substrate limitations for improved production of ethylene in *E. coli* ', *Biotechnol. Biofuels,* 9(1), p. 3.

M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. S. (1989) 'Current protocols in molecular biology', *Mol. Reprod. Dev.,* 1(2), pp. 146.

Makarova, K. S. *et al.* (2015) 'An updated evolutionary classification of CRISPR–Cas systems', *Nat. Rev. Microbiol.*, 13(11), pp. 722–736.

Makui, H. *et al.* (2000) 'Identification of the *Escherichia coli* K-12 Nramp orthologue (MntH) as a selective divalent metal ion transporter', *Mol. Microbiol.,* 35(5), pp. 1065–1078.

Mali, P. *et al.* (2013) 'RNA-guided human genome engineering via Cas9', *Science*. 339(6121), 823–826.

Mali, P., Esvelt, K. M. and Church, G. M. (2013) 'Cas9 as a versatile tool for engineering biology', *Nat. Methods*, 10(10), pp. 957–963.

Martínez-García, E. *et al.* (2014) 'New transposon tools tailored for metabolic engineering of gram-negative Microb. Cell Fact..', *Front. Bioeng. Biotechnol.* 2, p. 46.

Martinez, S. *et al.* (2017) 'Structures and Mechanisms of the Non-Heme Fe(II)- and 2-Oxoglutarate-Dependent Ethylene-Forming Enzyme: Substrate Binding Creates a Twist', *JACS*, 139 (34), pp. 11980-11988.

Martinez, S. and Hausinger, R. P. (2015) 'Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases', *J. Biol. Chem.*, 290(34), pp. 20702–20711.

Martinez, S. and Hausinger, R. P. (2016) 'Biochemical and Spectroscopic Characterization of the Non-Heme Fe(II)- and 2-Oxoglutarate-Dependent Ethylene-Forming Enzyme from *Pseudomonas syringae* pv. *phaseolicola* PK2', *Biochemistry*. 55(43), pp. 5989–5999.

Marx, C. J. and Lidstrom, M. E. (2002) 'Broad-Host-Range *cre-lox* System for Antibiotic Marker Recycling in Gram-Negative Bacteria', *BioTechniques*, 33(5), pp. 1062–1067.

Maurhofer, M. *et al.* (1998) 'Salicylic Acid Biosynthetic Genes Expressed in *Pseudomonas fluorescens* Strain P3 Improve the Induction of Systemic Resistance in Tobacco Against Tobacco Necrosis Virus', *Phytopathology*, 88(7), pp. 678–684.

Mecsas, J. *et al.* (1993) 'The activity of sigma E, an *Escherichia coli* heat-inducible sigma-factor, is modulated by expression of outer membrane proteins', *Genes Dev.*, 7(12B), pp. 2618–2628.

Mergeay, M. *et al.* (1985) '*Alcaligenes eutrophus* CH34 is a facultative chemolithotroph with plasmid-bound resistance to heavy metals', *J. Bacteriol.*, 162(1), pp. 328–334.

Mergeay, M., Houba, C. and Gerits, J. (1978) 'Extrachromosomal inheritance controlling resistance to cadmium, cobalt, copper and zinc ions: evidence from curing in a *Pseudomonas* ', *Archives internationales de physiologie et de biochimie*, 86(2), pp. 440–442.

Meynial-Salles, I., Cervin, M. A. and Soucaille, P. (2005) 'New tool for metabolic pathway engineering in *Escherichia coli* : one-step method to modulate expression of chromosomal genes.', *Appl. Environ. Microbiol.*

71(4), pp. 2140–2144.

Mifune, J., Nakamura, S. and Fukui, T. (2010) 'Engineering of *pha* operon on *Cupriavidus necator* chromosome for efficient biosynthesis of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) from vegetable oil', P*olym. Degrad. Stab.*, 95(8), pp. 1305–1312.

Miller, J. M. and Conn, E. E. (1980) 'Metabolism of hydrogen cyanide by higher plants', *Plant Physiol.* 65(6), pp. 1199–202.

Mo, H. *et al.* (2017) 'Effects of global transcription factor NtcA on photosynthetic production of ethylene in recombinant *Synechocystis* sp. PCC 6803', *Biotechnol. Biofuels*, 10(1), p. 145.

Mohidin Batcha, A. F. *et al.* (2014) 'Biosynthesis of poly(3-hydroxybutyrate) (PHB) by*Cupriavidus necator* H16 from jatropha oil as carbon source', *Bioprocess. Biosyst. Eng.,* 37(5), pp. 943–951.

Mojica, F. J. M. *et al.* (2005) 'Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements', *J. Mol. Evol.,* 60(2), pp. 174–182.

Monchy, S. *et al.* (2007) 'Plasmids pMOL28 and pMOL30 of *Cupriavidus metallidurans* are specialized in the maximal viable response to heavy metals', *J. Bacteriol.*, 189(20), pp. 7417–7425.

Morschbacker, A. (2009) 'Bio-Ethanol Based Ethylene', *Polym. Rev.* 49(2), pp. 79–84.

Mosberg, J. A., Lajoie, M. J. and Church, G. M. (2010) 'Lambda red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate', *Genetics*. 186(3), pp. 791–799.

Moscou, M. J. and Bogdanove, A. J. (2009) 'A Simple Cipher Governs DNA Recognition by TAL Effectors', *Science*, 326(5959), pp. 1501.

Mougiakos, I. *et al.* (2018) 'Hijacking CRISPR-Cas for high-throughput bacterial metabolic engineering: advances and prospects', *Curr. Opin. Biotechnol.,* 50, pp. 146–157.

Munkelt, D., Grass, G. and Nies, D. H. (2004) 'The chromosomally encoded cation diffusion facilitator proteins DmeF and FieF from *Wautersia metallidurans* CH34 are transporters of broad metal specificity', *J. Bacteriol.*, 186(23), pp. 8036–8043.

Murphy, K. C. (2016) 'λ Recombination and Recombineering', *EcoSal Plus*, 7(1).

Murphy, K. C. and Campellone, K. G. (2003) 'Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic *E. coli* ', *BMC Mol. Biol.*, 4(1), p. 11.

Murphy, K. C., Campellone, K. G. and Poteete, A. R. (2000) 'PCR-mediated gene replacement in *Escherichia coli*', *Gene*, 246(1–2), pp. 321–330.

Muyrers, J. P. *et al.* (2000) 'RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners', *Genes Dev.* 14(15), pp. 1971–1982.

Nagahama, K. *et al.* (1991) 'L-Arginine is essential for the formation in vitro of ethylene by an extract of *Pseudomonas syringae*', *J. Gen. Microbiol.* 137, pp. 164–165.

Nagahama, K. *et al.* (1992) 'Classification of ethylene-producing bacteria in terms of biosynthetic pathways to ethylene', *J. Ferment. Bioeng.*, Elsevier, 73(1), pp. 1–5.

Nagy, A. (2000) 'Cre recombinase: the universal reagent for genome tailoring', *Genesis,* 26(2), pp. 99–109.

Nakashima, N. *et al.* (2014) 'Bacterial Cellular Engineering by Genome

Editing and Gene Silencing', *Int. J. Mol. Sci.,* 15(2), pp. 2773–2793.

Nanninga, N. (1991) 'Cell division and peptidoglycan assembly in *Escherichia coli*, *Mol. Microbiol.,* 5(4), pp. 791–795.

Nies, D. *et al.* (1987) 'Cloning of plasmid genes encoding resistance to cadmium, zinc, and cobalt in *Alcaligenes eutrophus* CH34', *J. Bacteriol.*, 169(10), pp. 4865–4868.

Nies, D. H. (2000) 'Heavy metal-resistant bacteria as extremophiles: molecular physiology and biotechnological use of *Ralstonia* sp. CH34', *Extremophiles,* 4(2), pp. 77–82.

Noirot, P. and Kolodner, R. D. (1998) 'DNA strand invasion promoted by *Escherichia coli* RecT protein', *J. Biol. Chem.*, 273(20), pp. 12274–12280.

Noskov, V. N. *et al.* (2015) 'Recombinase-mediated cassette exchange (RMCE) system for functional genomics studies in *Mycoplasma mycoides*', *Biol. Proced. Online,* 17, p. 6.

Obruca, S. *et al.* (2014) 'Application of protease-hydrolyzed whey as a complex nitrogen source to increase poly(3-hydroxybutyrate) production from oils by *Cupriavidus necator*', *Biotechnol. Lett*. 36(4), pp. 775–781.

Ogawa, T. *et al.* (1990) 'The Role of NADH:Fe(III)EDTA Oxidoreductase in Ethylene Formation from 2-Keto-4-Methylthiobutyrate', *J. Ferment. Bioeng.*, 69(5), pp. 287–291.

Oh, J.-H. and van Pijkeren, J.-P. (2014) 'CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*', *Nucleic Acids Res.*, 42(17), p. 131.

Ohara, M. *et al.* (1999) 'Identification and characterization of a new lipoprotein, NlpI, in *Escherichia coli* K-12', *J. Bacteriol.*, 181(14), pp. 4318–4325.

Ostrowski, J. *et al.* (1989) 'Characterization of the flavoprotein moieties of NADPH-sulfite reductase from *Salmonella typhimurium* and *Escherichia coli*. Physicochemical and catalytic properties, amino acid sequence deduced from DNA sequence of *cysJ,* and comparison with NADPH-cytochrom', *J. Biol. Chem.*, 264(27), pp. 15796–15808.

Otto, K. *et al.* (2001) 'Adhesion of Type 1-Fimbriated *Escherichia coli* to Abiotic Surfaces Leads to Altered Composition of Outer Membrane Proteins', *J. Bacteriol.*, 183(8), pp. 2445–2453.

Park, J. M. *et al.* (2010) 'Development of a gene knockout system for *Ralstonia eutropha* H16 based on the broad-host-range vector expressing a mobile group II intron', *FEMS Microbiol. Lett.*, 309(2), p. 193-200.

Patel, M. (2006) 'Medium and Long-term Opportunities and Risks of the Biotechnological Production of Bulk Chemicals from Renewable Resources - The Potential of White Biotechnology'. Final report prepared under the European Commission's GROWTH Programme (DG Research). Available at: http://edepot.wur.nl/183326

Patzer, S. I. and Hantke, K. (2001) 'Dual repression by Fe^{2+}-Fur and Mn^{2+}-MntR of the *mntH*gene, encoding an NRAMP-like Mn^{2+} transporter in *Escherichia coli*', *J. Bacteriol.*, 183(16), pp. 4806–13.

Penewit, K. *et al.* (2018) 'Efficient and Scalable Precision Genome Editing in *Staphylococcus aureus* through Conditional Recombineering and CRISPR/Cas9-Mediated Counterselection', *mBio*. 9(1), e00067-18.

Peoples, O. P. and Sinskey, A. J. (1989) 'Poly-beta-hydroxybutyrate biosynthesis in *Alcaligenes eutrophus* H16. Characterization of the genes encoding beta-ketothiolase and acetoacetyl-CoA reductase.', *J. Biol. Chem.*, 264(26), pp. 15293–15297.

Pfeiffer, D. and Jendrossek, D. (2012) 'Localization of Poly(3-

Hydroxybutyrate) (PHB) Granule-Associated Proteins during PHB Granule Formation and Identification of Two New Phasins, PhaP6 and PhaP7, in *Ralstonia eutropha* H16', *J. Bacteriol.*, 194(21), pp. 5909–5921.

Pingoud, A. and Jeltsch, A. (2001) 'Structure and function of type II restriction endonucleases', *Nucleic Acids Res.*, 29(18), pp. 3705–3727.

Pirkov, I. *et al.* (2008) 'Ethylene production by metabolic engineering of the yeast *Saccharomyces cerevisiae*', *Met. Eng.*,10(5), pp. 276–280.

Pohlmann, A. *et al.* (2006) 'Genome sequence of the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha* H16', *Nat. Biotechnol.*, 24(10), pp. 1257–1262.

Portnoy, V. A., Bezdan, D. and Zengler, K. (2011) 'Adaptive laboratory evolution—harnessing the power of biology forMet. Eng.', *Curr. Opin. Biotechnol.*, 22(4), pp. 590–594.

Portnoy, V. A., Herrgård, M. J. and Palsson, B. Ø. (2008) 'Aerobic fermentation of D-glucose by an evolved cytochrome oxidase-deficient *Escherichia coli* strain.', *Appl. Environ. Microbiol.*, 74(24), pp. 7561–7569.

Posfai, G. *et al.* (1999) 'Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome', *Nucleic Acids Res.*, 27(22), pp. 4409–4415.

Poteete, A. R. (2008) 'Involvement of DNA replication in phage lambda Red-mediated homologous recombination', *Mol. Microbiol.,* 68(1), pp. 66–74.

Pourcel, C., Salvignol, G. and Vergnaud, G. (2005) 'CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies', *Microbiology*, 151(3), pp. 653–663.

Proshlyakov, D. A., McCracken, J. and Hausinger, R. P. (2017) 'Spectroscopic

analyses of 2-oxoglutarate-dependent oxygenases: TauD as a case study', *J. Biol. Inorg. Chem.,* 22(2–3), pp. 367–379.

Quandt, J. and Hynes, M. F. (1993) 'Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria., *Gene*, 127(1), pp. 15–21.

Quintin, M. *et al.* (2016) 'Merlin: Computer-Aided Oligonucleotide Design for Large Scale Genome Engineering with MAGE', *ACS Synth. Biol.*, 5(6), pp. 452–458.

Raberg, M. *et al.* (2014) 'A Closer Look on the Polyhydroxybutyrate- (PHB-) Negative Phenotype of *Ralstonia eutropha* PHB-4', *PLoS ONE*. Public Library of Science, 9(5), p. 95907.

Ramirez, C. L. *et al.* (2008) 'Unexpected failure rates for modular assembly of engineered zinc fingers', *Nat. Methods*, 5(5), pp. 374–375.

Ran, F. A. *et al.* (2013) 'Genome engineering using the CRISPR-Cas9 system', N*at. Protoc.*, pp. 2281–2308.

Reinecke, F. and Steinbüchel, A. (2009) '*Ralstonia eutropha* strain H16 as model organism for PHA metabolism and for biotechnological production of technically interesting biopolymers', *J. Mol. Microbiol. Biotechnol.,* 16(1–2), pp. 91–108.

Reisch, C. R. and Prather, K. L. J. (2015) 'The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli* ', *Sci. Rep*., 5(1), p. 15096.

Reith, F. *et al.* (2009) 'Mechanisms of gold biomineralization in the bacterium *Cupriavidus metallidurans*', *Proc. Natl. Acad. Sci. U.S.A*., 106(42), pp. 17757–17762.

Reznikoff, W. S. (2008) 'Transposon Tn*5*', *Annu. Rev. Genet.* 42(1), pp. 269–

286.

Richardson, C. D. *et al.* (2016) 'Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA', *Nat. Biotechnol*, 34(3), pp. 339–344.

Roberts, R. *et al.* (2003) 'A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes', *Nucleic Acids Res.*, 31(7), pp. 1805–1812.

Roberts, R. J. (1990) 'Restriction enzymes and their isoschizomers.', *Nucleic Acids Res.*, 18(Suppl), pp. 2331–2365.

Rodriguez, F. I. *et al.* (1999) 'A Copper Cofactor for the Ethylene Receptor ETR1 from *Arabidopsis*', *Science*, 283(5404), pp. 996–998.

Rojas, L. A. *et al.* (2011) 'Characterization of the metabolically modified heavy metal-resistant*Cupriavidus metallidurans* strain MSR33 generated for mercury bioremediation.', *PloS One*, 6(3), p. 17555.

Ronda, C. *et al.* (2016) 'CRMAGE: CRISPR Optimized MAGE Recombineering', *Sci. Rep*. 6(1), p. 19452.

Rudolph, M. M., Vockenhuber, M.-P. and Suess, B. (2013) 'Synthetic riboswitches for the conditional control of gene expression in *Streptomyces coelicolor*', *Microbiology*. 159(7), pp. 1416–1422.

Sakai, M. *et al.* (1997) 'Photosynthetic conversion of carbon dioxide to ethylene by the recombinant cyanobacterium, *Synechococcus* sp. PCC 7942, which harbors a gene for the ethylene-forming enzyme of *Pseudomonas syringae*', *J. Ferment. Bioeng.* 84(5), pp. 434–443.

Samai, P. *et al.* (2015) 'Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity', *Cell*. 161(5), pp. 1164–1174.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press, NY.

Santoro, S. W. and Schultz, P. G. (2002) 'Directed evolution of the site specificity of Cre recombinase', *Proc. Natl. Acad. Sci. U.S.A.*, 99(7), pp. 4185–4190.

Sapranauskas, R. *et al.* (2011) 'The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*', *Nucleic Acids Res.*, 39(21), pp. 9275–9282.

Sato, S., Fujiki, T. and Matsumoto, K. (2013) 'Construction of a stable plasmid vector for industrial production of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) by a recombinant *Cupriavidus necator* H16 strain', *J. Biosci. Bioeng.,* 116(6), pp. 677–681.

Sauer, M. (2016) 'Industrial production of acetone and butanol by fermentation-100 years later.', *FEMS Microbiol. Lett.*., 363(13).

Sawitzke, J. A. *et al.* (2007) 'Recombineering: In Vivo Genetic Engineering in *E. coli* , *S. enterica*, and Beyond', in *Methods Enzymol*, 121, pp. 171–199.

Schaller, G. E., Shiu, S.-H. and Armitage, J. P. (2011) 'Two-component systems and their co-option for eukaryotic signal transduction', *Curr. Biol.,* 21(9), pp. 320-330.

Schlegel, Kaltwasser and Gottschalk (1961) 'A submersion method for culture of hydrogen-oxidizing bacteria: growth physiological studies', *Archiv fur Mikrobiologie*, 38, pp. 209–222.

Schlegel, H. G., Lafferty, R. and Krauss, I. (1970) 'The isolation of mutants not accumulating poly-beta-hydroxybutyric acid', *Archiv fur Mikrobiologie*, 71(3), pp. 283–294.

Schleif, R. (2010) 'AraC protein, regulation of the L-arabinose operon in

Escherichia coli, and the light switch mechanism of AraC action', *FEMS Microbiol. Rev.*, 34(5), pp. 779–796.

Schwartz, E. and Friedrich, B. (2001) 'A physical map of the megaplasmid pHG1, one of three genomic replicons in *Ralstonia eutropha* H16.', *FEMS Microbiol. Lett.*, 201(2), pp. 213–219.

Schwechheimer, C., Rodriguez, D. L. and Kuehn, M. J. (2015) 'NlpI-mediated modulation of outer membrane vesicle production through peptidoglycan dynamics in *Escherichia coli*', *MicrobiologyOpen*, 4(3), pp. 375–89.

Serra-Moreno, R. *et al.* (2006) 'Use of the lambda Red recombinase system to produce recombinant prophages carrying antibiotic resistance genes', *BMC Mol. Biol.*, 7(1), p. 31.

Shen, P. and Huang, H. V (1986) 'Homologous recombination in *Escherichia coli*: dependence on substrate length and homology', *Genetics*, 112(3), pp. 441–457.

Shmakov, S. *et al.* (2015) 'Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems', *Mol. Cell,* 60(3), pp. 385–397.

Shmakov, S. *et al.* (2017) 'Diversity and evolution of class 2 CRISPR–Cas systems', *Nat. Rev. Microbiol.*, 15(3), pp. 169–182.

Si, T. *et al.* (2015) 'RNAi-Assisted Genome Evolution in *Saccharomyces cerevisiae* for Complex Phenotype Engineering', *ACS Synth. Biol.*, 4(3), pp. 283–291.

Signer, E. R. and Weil, J. (1968) 'Recombination in bacteriophage lambda. I. Mutants deficient in general recombination.', *J. Mol. Biol.*, 34(2), pp. 261–271.

Simon, R., Priefer, U. and Pühler, A. (1983) 'A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram

Negative Bacteria', *Biotechnology,* 1(9), pp. 784–791.

Simpson, A. E., Skurray, R. A. and Firth, N. (2003) 'A single gene on the staphylococcal multiresistance plasmid pSK1 encodes a novel partitioning system', *J. Bacteriol.*, 185(7), pp. 2143–2152.

del Solar, G. *et al.* (1998) 'Replication and control of circular bacterial plasmids', *Microbiol. Mol. Biol. Rev.,* 62(2), pp. 434–464.

Springael, D., Diels, L. and Mergeay, M. (1994) 'Transfer and expression of PCB-degradative genes into heavy metal resistant *Alcaligenes eutrophus* strains', *Biodegradation*, 5(3–4), pp. 343–357.

Stanisich, V. A., Bennett, P. M. and Richmond, M. H. (1977) 'Characterization of a translocation unit encoding resistance to mercuric ions that occurs on a nonconjugative plasmid in *Pseudomonas aeruginosa*', *J. Bacteriol.*, 129(3), pp. 1227–1233.

Steinmetz, M. *et al.* (1983) '[Genetic analysis of *sacB*, the structural gene of a secreted enzyme, levansucrase of *Bacillus subtilis*', *Mol. Genet. Genomics,* 191(1), pp. 138–144.

Sternberg, S. H. *et al.* (2015) 'Conformational control of DNA target cleavage by CRISPR–Cas9', *Nature*, 527(7576), pp. 110–113.

Sugiura, S., Ohkubo, S. and Yamaguchi, K. (1993) 'Minimal Essential Origin of Plasmid pSC101 Replication: Requirement of a Region Downstream of Iterons', *J. Bacteriol.,* 175(18), pp. 5993–6001.

Sun, J. *et al.* (2018) 'Genome editing and transcriptional repression in *Pseudomonas putida* KT2440 via the type II CRISPR system', *Microb. Cell Fact.*, 17(1), p. 41.

Suzuki, N. *et al.* (2005) 'New multiple-deletion method for the *Corynebacterium glutamicum* genome, using a mutant *lox* sequence', *Appl.*

Environ. Microbiol., 71(12), pp. 8472–8480.

Swingle, B. *et al.* (2009) 'Oligonucleotide recombination in Gram-negative bacteria' *Mol. Microbiol.,* 75(1), pp. 138–148.

Swingle, B. *et al.* (2010) 'Recombineering using RecTE from *Pseudomonas syringae*.', *Appl. Environ. Microbiol.*,76(15), pp. 4960–4968.

Sydow, A. *et al.* (2017) 'Expanding the genetic tool box for *Cupriavidus necator* by a stabilized L-rhamnose inducible plasmid system', *J. Biotechnol.*, 263, pp. 1–10.

Tan, F. *et al.* (2016) 'Using global transcription machinery engineering (gTME) to improve ethanol tolerance of *Zymomonas mobilis*', *Microb. Cell Fact.*, 15, p. 4.

Tao, L. *et al.* (2008) 'Expression of ethylene-forming enzyme (EFE) of *Pseudomonas syringae* pv. *glycinea* in *Trichoderma viride*', *Appl. Microbiol. Biotechnol.*, 80(4), pp. 573–578.

Tee, K. L. *et al.* (2017) 'An Efficient Transformation Method for the Bioplastic-Producing "Knallgas" Bacterium *Ralstonia eutropha* H16', *Biotechnol. J.,* 12(11), p. 1700081.

Thomason, L. C., Costantino, N. and Court, D. L. (2016) 'Examining a DNA Replication Requirement for Bacteriophage λ Red- and Rac Prophage RecET-Promoted Recombination in *Escherichia coli*', *mBio*., 7(5), pp. 1-10.

Thorburn, P. J. *et al.* (2011) 'Environmental impacts of irrigated sugarcane production: Nitrogen lost through runoff and leaching', *Agric. Ecosyst. Environ*., 144(1), pp. 1–12.

Tock, M. R. and Dryden, D. T. (2005) 'The biology of restriction and anti-restriction', *Curr. Opin. Microbiol., 8*(4), pp. 466–472.

Tolmachov, O. *et al.* (2006) 'RecET driven chromosomal gene targeting to generate a RecA deficient *Escherichia coli* strain for Cre mediated production of minicircle DNA', *BMC Biotech.,* 6, p. 17.

Topp, S. *et al.* (2010) 'Synthetic riboswitches that induce gene expression in diverse bacterial species', *Appl. Environ. Microbiol.*, 76(23), pp. 7881–7884.

Tremblay, P.-L. *et al.* (2011) 'A c-type cytochrome and a transcriptional regulator responsible for enhanced extracellular electron transfer in *Geobacter sulfurreducens* revealed by adaptive evolution', *Environ. Microbiol.*, 13(1), pp. 13–23.

Trinh, C. T. and Srienc, F. (2009) 'Metabolic Engineering of *Escherichia coli* for Efficient Conversion of Glycerol to Ethanol', *Appl. Environ. Microbiol.*, 75(21), pp. 6696–6705.

Trinh, C. T., Unrean, P. and Srienc, F. (2008) 'Minimal *Escherichia coli* Cell for the Most Efficient Production of Ethanol from Hexoses and Pentoses', *Appl. Environ. Microbiol.*, 74(12), pp. 3634–3643.

Turan, S. *et al.* (2010) 'Multiplexing RMCE: Versatile Extensions of the Flp-Recombinase-Mediated Cassette-Exchange Technology', *J. Mol. Biol.*, 402(1), pp. 52–69.

Ungerer, J. *et al.* (2012) 'Sustained photosynthetic conversion of CO_2 to ethylene in recombinant cyanobacterium *Synechocystis* 6803', *Energy Environ. Sci.*, 5(10), p. 8998.

Ungerer, J. and Pakrasi, H. B. (2016) 'Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria', *Sci. Rep*., 6(1), p. 39681.

Urnov, F. D. *et al.* (2010) 'Genome editing with engineered zinc finger nucleases', *Nat. Rev. Genet.*, 11(9), pp. 636–646.

Vandamme, P. and Coenye, T. (2004) 'Taxonomy of the genus *Cupriavidus* : a tale of lost and found', *Int. J. Syst. Evol. Microbiol.,* 54(6), pp. 2285–2289.

Vaneechoutte, M. *et al.* (2004) '*Wautersia* gen. nov., a novel genus accommodating the phylogenetic lineage including *Ralstonia eutropha* and related species, and proposal of *Ralstonia syzygii* comb. nov.', *Int. J. Syst. Evol. Microbiol.*, 54(2), pp. 317–327.

Van Vlack, E. R. and Seeliger, J. C. (2015) 'Using riboswitches to regulate gene expression and define gene function in mycobacteria', *Methods Enzymol*., 550, pp. 251–265.

Veetil, P. V., Angermayr, S. A. and Hellingwerf, K. J. (2017) 'Ethylene production with engineered *Synechocystis* sp PCC 6803 strains' *Microb. Cell Fact.*, 16, p. 34.

Volodina, E., Raberg, M. and Steinbüchel, A. (2016) 'Engineering the heterotrophic carbon sources utilization range of *Ralstonia eutropha* H16 for applications in biotechnology', *Crit. Rev. Biotechnol.*, 36(6), pp. 978–991.

Voss, I. and Steinbüchel, A. (2006) 'Application of a KDPG-aldolase gene-dependent addiction system for enhanced production of cyanophycin in *Ralstonia eutropha* strain H16',*Met. Eng.,* 8(1), pp. 66–78.

Waite-Rees, P. A. *et al.* (1991) 'Characterization and expression of the *Escherichia coli* Mrr restriction system', *J. Bacteriol.*, 173(16), pp. 5207–5219.

Wang, H. H. *et al.* (2009) 'Programming cells by multiplex genome engineering and accelerated evolution', *Nature*, 460(7257), pp. 894–898.

Wang, H. H. *et al.* (2012) 'Genome-scale promoter engineering by coselection MAGE', *Nat. Methods*, 9(6), pp. 591–593.

Wang, J.-P. *et al.* (2010) 'Metabolic engineering for ethylene production by

inserting the ethylene-forming enzyme gene (*efe*) at the 16S rDNA sites of *Pseudomonas putida* KT2440', *Bioresour. Technol.*, 101(16), pp. 6404–6409.

Wang, J. *et al.* (2006) 'An improved recombineering approach by adding RecA to lambda Red recombination', *Molecular biotechnology,* 32(1), pp. 43–53.

Wang, S. Y., Adams, D. O. and Lieberman, M. (1982) 'Recycling of 5'-methylthioadenosine-ribose carbon atoms into methionine in tomato tissue in relation to ethylene production', *Plant Physiol.*, 70(1), pp. 117–121.

Wang, W. *et al.* (2006) 'Identification of Important Regions for Ethylene Binding and Signaling in the Transmembrane Domain of the ETR1 Ethylene Receptor of *Arabidopsis*', *Plant Cell*., 10(8), pp. 1321–1332.

Wang, Y. *et al.* (2015) 'Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system', *J. Biotechnol.*, 200, pp. 1–5.

Wards, B. J. and Collins, D. M. (1996) 'Electroporation at elevated temperatures substantially improves transformation efficiency of slow-growing mycobacteria', *FEMS Microbiol. Lett.*, 145, pp. 101-105.

Warming, S. *et al.* (2005) 'Simple and highly efficient BAC recombineering using *galK* selection', *Nucleic Acids Res.*, 33(4), pp. 36.

Watabe, H. *et al.* (1983) 'A new class of site-specific endodeoxyribonucleases. Endo.Sce I isolated from a eukaryote, *Saccharomyces cerevisiae*', *J. Biol. Chem.*, 258(8), pp. 4663–4665.

Wei, T., Cheng, B.-Y. and Liu, J.-Z. (2016) 'Genome engineering *Escherichia coli* for L-DOPA overproduction from glucose', *Sci.Rep*., 6(1), p. 30080.

Wendisch, V. F. *et al.* (2016) 'Updates on industrial production of amino acids using *Corynebacterium glutamicum*', *World J. Microbiol. Biotechnol*.,

32(6), p. 105.

Wendt, K. E. *et al.* (2016) 'CRISPR/Cas9 mediated targeted mutagenesis of the fast growing cyanobacterium *Synechococcus elongatus* UTEX 2973', *Microb. Cell Fact.*, 15, p. 115.

Westra, E. R. *et al.* (2012) 'CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3', *Mol. Cell*, 46(5), pp. 595–605.

White-Ziegler, C. A. *et al.* (2008) 'Low temperature (23C) increases expression of biofilm-, cold-shock- and RpoS-dependent genes in *Escherichia coli* K-12', *Microbiology*, 154(1), pp. 148–166.

Wigley, D. B. (2013) 'Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB', *Nat. Rev. Microbiol.,* 11(1), pp. 9–13.

Wilde, E. (1962) 'Untersuchungen über Wachstum und Speicherstoffsynthese von Hydrogenomonas', *Archiv für Mikrobiologie*, 43(2), pp. 109–137.

Wolfe, A. J. (2005) 'The Acetate Switch', *Microbiol. Mol. Biol. Rev. ',* 69(1), pp. 12–50.

Wollman, E. L. and Jacob, F. (1954) 'Etude génétique d'un bactériophage tempéré d'*Escherichia coli* . II. Mécanisme de la recombinaison génétique', *Annales de l'Institut Pasteur*, 87(6), pp. 674–690.

Wong, Y.-M. *et al.* (2012) 'Biosynthesis and characterization of polyhydroxyalkanoate containing high 3-hydroxyhexanoate monomer fraction from crude palm kernel oil by recombinant *Cupriavidus necator*', *Bioresour. Technol.*, 121, pp. 320–327.

Worrell, E. *et al.* (2000) *Energy Use and Energy Intensity of the U.S. Chemical*

Industry. Available at: https://www.energystar.gov/ia/business/industry/industrial_LBNL-44314.pdf

Wrighton, K. C. and Coates, J. D. (2009) 'Microbial Fuel Cells: Plug-in and Power-on Microbiology These devices already prove valuable for characterizing physiology, modeling electron flow, and framing and testing hypotheses'*Microbe*, 6(4), pp. 281-287.

Wu, Y. *et al.* (2017) 'RecET recombination system driving chromosomal target gene replacement in *Zymomonas mobilis*', *Electronic J. Biotechnol.*, 30, pp. 118–124.

Xiong, B. *et al.* (2018) 'Genome editing of *Ralstonia eutropha* using an electroporation-based CRISPR-Cas9 technique', *Biotechnol. Biofuels,* 11(1), p. 172.

Xiong, W., Morgan, J. A. and Ungerer, J. (2015) 'The plasticity of cyanobacterial metabolism supports direct CO_2 conversion to ethylene Secondary metabolite production from hairy root cultures of *Catharanthus roseus* View project'. *Nat. Plants*, 15053, pp. 1-6.

Yabuuchi, E. *et al.* (1995) 'Transfer of Two *Burkholderia* and An *Alcaligenes* Species to *Ralstonia* Gen. Nov.', *Microbiol. Immunol.,* 39(11), pp. 897–904.

Yan, Q., Du, G. and Chen, J. (2003) 'Biosynthesis of polyhydroxyalkanoates (PHAs) with continuous feeding of mixed organic acids as carbon sources by *Ralstonia eutropha*', *Process Biochem.* 39(3), pp. 387–391.

Yan, Q. and Fong, S. S. (2017) 'Study of in vitro transcriptional binding effects and noise using constitutive promoters combined with UP element sequences in *Escherichia coli*', *J. Biol. Eng.,* 11, p. 33.

Yao, J. and Lambowitz, A. M. (2007) 'Gene targeting in Gram-negative

bacteria by use of a mobile group II intron (Targetron) expressed from a broad-host-range vector', *Appl. Environ. Microbiol.*, 73(8), pp. 2735–2743.

Yao, L. *et al.* (2016) 'Multiple Gene Repression in Cyanobacteria Using CRISPRi', *ACS Synth. Biol.*, 5(3), pp. 207–212.

Yao, R. *et al.* (2018) 'CRISPR-Cas9/Cas12a biotechnology and application in bacteria', *Synth. Syst. Biotechnol.*, 3(3), pp. 135–149.

Yosef, I. *et al.* (2004) 'Restoration of gene function by homologous recombination: from PCR to gene expression in one step', *Appl. Environ. Microbiol.*, 70(12), pp. 7156–7160.

Yu, D. *et al.* (2000) 'An efficient recombination system for chromosome engineering in *Escherichia coli* ', *Proc. Natl. Acad. Sci. U.S.A*, 97(11), pp. 5978–5983.

Zavřel, T. *et al.* (2016) 'A quantitative evaluation of ethylene production in the recombinant cyanobacterium *Synechocystis* sp. PCC 6803 harboring the ethylene-forming enzyme by membrane inlet mass spectrometry', *Bioresour. Technol.*, 202, pp. 142–151.

Zetsche, B. *et al.* (2015) 'Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system.', *Cell*. 163(3), pp. 759–771.

Zhang, C., Wei, Z.-H. and Ye, B.-C. (2013) 'Quantitative monitoring of 2-oxoglutarate in *Escherichia coli* cells by a fluorescence resonance energy transfer-based biosensor', *Appl. Microbiol. Biotechnol.*, 97(18), pp. 8307–8316.

Zhang, G. *et al.* (2015) 'TALENs-Assisted Multiplex Editing for Accelerated Genome Evolution To Improve Yeast Phenotypes', *ACS Synth. Biol.*, 4(10), pp. 1101–1111.

Zhang, L. *et al.* (2011) 'Application of simultaneous saccharification and

fermentation (SSF) from viscosity reducing of raw sweet potato for bioethanol production at laboratory, pilot and industrial scales', *Bioresour. Technol.,* 102(6), pp. 4573–4579.

Zhang, Y. *et al.* (1998) 'A new logic for DNA engineering using recombination in *Escherichia coli* ', *Nat. Genetics*. 20(2), pp. 123–128.

Zhang, Z. *et al.* (2017) 'Structural and stereoelectronic insights into oxygenase-catalyzed formation of ethylene from 2-oxoglutarate', *Proc. Natl. Acad. Sci. U.S.A*., 114(18), pp. 4667–4672.

Ziegler, T. and Rauk, A. (1979) 'A theoretical study of the ethylene-metal bond in complexes between copper(1+), silver(1+), gold(1+), platinum(0) or platinum(2+) and ethylene, based on the Hartree-Fock-Slater transition-state method', *Inorg. Chem.*, 18(6), pp. 1558–1565.

Zimmermann, H. and Walzl, R. (2009) 'Ethylene', in *Ullmann's Encyclopedia of Industrial Chemistry*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co, pp.465-529.

Appendices

Tables

Table A. 1 List of primers used in this study

	Chapter 3 - Recombineering plasmids	
pSLBBR1	pSLBBR1.pBBR1.F	AAACTTTCCTGAAATAACCATGGGTGTGCTACGCCTG AATAAGTG
	pSLBBR1.pBBR1.R	ttttgttaccctaaggtctcgcggtCTGAGGTCTGCCTCGTGAAG
	pSLBBR1.Cm.F	CTTCACGAGGCAGACCTCAGcgagaccttagggtaacaaaa
	Cm.R	gtagccgtcaagttgtcataataaatcctgcaggCGATGCAGGTGG CACTTTTC
	Lambda.F	TTCCCCGAAAAGTgccacctgcatcgcctgcaggATTTATTA TGACAACTTGACG
	Lambda.R	TTAACCATGGATTCTTCGTCTGTTTCTACTGGTATTG GCACAAAC
	I-SceI.F	TGCCAATACCAGtagaaacagacgaggcgcgccAGAATCCAT GGTTAAGACCCAC
	pSLBBR1.I-SceI.R	cgcgtcgacgtcatatggatccgatatcACCCATGACCCATGGTT ATTTCAGGAAAG
	I-SceI.IN.F	CATGGACCACGTATGTCTGC
	I-SceI.IN.R	TGGTTCATGGATGATGGTGG
	araC.seq.F	atgatttgcccaaacaggtc
pSLVS1	pSLVS1.pVS1.F	tagaaatacggtgttttttgttaccctaaggtctcgcggtCTCTAGAAC TAGCTAGAACTAGGATC
	pSLVS1.pVS1.R	gttaccctaaggtctcgcggtCTCTAGAACTAGCTAGAACTAG GATC
	pSLVS1.Cm.F	gatcctagttctagctagttctagaACCGCGAGACCTTAGGGTA AC
	pSLVS1.I-SceI.R	cgcgtcgacgtcatatggatccgatatcACCCATGACCCATGGTT ATTTCAGGAAAG
p*recET-C*	p*recET-C*.recET.F	gggaaagcggccgCTAAGGAGGTTATAAAAAATG
	p*recET-C*.recET.R	tttgggcatatgCTACGTATGTTGCGCTACAG
precET-P	precET-P.BB.F	acccgttttttttgggaattcgagctctaaggaggttataaaaaatgtccgc aagaaacgttgc
	precET-P.BB.R	atttgatgcctggTCATGCTGCCACCTGCGT
	precET-P.recET.F	ggtggcagcatgaCCAGGCATCAAATAAAAC
	precET-P.recET.R	cttaaatgtgaaagtgggtcttaaccatggattcttcgtctgtttTATAA ACGCAGAAAGGCC
pSV200	C1.F	atgggctgggagcttgaagacg
	C1.R	tcacacctgcatcagcttgtcg
pSV242	pSV242.LHA.F	ttcctgcagcccgggggatccactagtttataaATGCAGGCCGCTG CCGGT
	pSV242/442.LHA.R	ggcactcatgcaagcgGATTTGATTGTCTCTCTGCCGTCAC TATTCGAACC
	pSV242/442.I-SceI.F	agagacaatcaaatccgcttgcatgagtgccggcgtgcgtcatgcacgg cgccggcaggcctgcaGAATTCAGGCGTATCACG
	pSV242/442.I-SceI.R	ataaacttatcatccATTACCCTGTTATCCCTAG
	pSV242/442.aadA.F	ggataacagggtaatGGATGATAAGTTTATCACCAC
	pSV242/442.aadA.R	cggctccgggcattgcTTATTTGCCGACTACCTTG
	pSV242/442.RHA.F	gtagtcggcaaataagcaatgcccggagccggttcgaatagtgacggca gagagacaatcaaatcCGCTTGCATGAGTGCCGGC

	pSV242.RHA.R	aattggagctccaccgcggtggcggccgctctagaACTCGGCGCGCGACAGGG
	C2.F	ATGCAGGCCGCTGCCGGT
	C2.R	ACTCGGCGCGCGACAGGG
pSV442	pSV442.LHA.F	actgactctagtttataaatgcaggccgctgccggt
	pSV442.RHA.R	tcacgtgatacttatctagaactcggcgcgcgacagg
pLO3_Δ*recJ*	H16_RS05815.LHA.F	cacctagatccttttaattcgagctcTGCCGCCCGCCCAGCTGG
	H16_RS05815.LHA.R	cccccgggagcacgGGGCGGGGCGGCAGGTGATG
	H16_RS05815.RHA.F	tgccgccccgcccCGTGCTCCCGGGGGCTCC
	H16_RS05815.RHA.R	gcaggtttaaacagtcgactctagatctagaGGCGTCGCACGCGGCGAT
	recJ.OUT.F	CGCGATCACCTGGTGCTG
	recJ.OUT.R	GAACGGAGATCGGAGATGG
	Chapter 4 - CRISPR plasmids	
pSV500	pSLVS1.Bb. F	CGCATCCTCACGATAATATC
	pSLVS1.Bb.R	tcgttttatacctctgaatcaatatc
	Cas9.F	ggttgatattgattcagaggtataaaacgaATGGATAAGAAATACTCAATAGGCTTAG
	Cas9.R	cgcctacccggatattatcgtgaggatgcgTCAGTCACCTCCTAGCTGAC
	I-SceI IN.F	GCTGATCGAACTGAACATCG
	I-SceI IN.R	CCCAGGTGGTTAACACGTTC
pSV501	Cas9.Strep.F	atttttcaacttgcccactgtgcacgtttaaacTTATGACAACTTGACGGCTACATCATTCACTTTTTC
	Cas9.Strep.R	tcgaagcttctgcagacgcgtcgacactagtTCAGTCGCCGCCGAGCTG
pSV711	pSV711.araC.F	tcagtgccaacatagtaagccagtaTTATGACAACTTGACGGC
	pSV711.araC.R	acttcttgtccatTTTTTATAACCTCCTTAGAGC
	pSV711.Cas9ST.F	gaggttataaaaaATGGACAAGAAGTACAGCATCGGC
	pSV711.Cas9ST.R	tcgtgaggatgcgTCAGTCGCCGCCGAGCTG
	pSV711.tl3.F	cggcggcgactgaCGCATCCTCACGATAATATC
	pSV711.tl3.R	cttttgcggggaacgcgtagatctgCTACTGGTATTGGCACAAAC
pCRMAGE	pCRMAGE.pBBR1.F	tcctcggactagcctaccggcgcggcagcgt
	pCRMAGE.pBBR1.R	ctggcaattccggaggtggcacttttcggggaaatgtgc
	pCRMAGE.KanR.F	aaaagtgccacctccggaattgccagctggg
	pCRMAGE.KanR.R	ctagtcagtaagtttaaactcagaagaactcgtcaagaaggc
	pCRMAGE.tl3.F	gttcttctgagtttaaacttactgactagtcgcatcctcacgataatatc
	pCRMAGE.tl3.R	gttttcggattagctactggtattggcacaaac
	pCRMAGE.nahR.F	ccaataccagtagctaatccgaaaacaagccg
	pCRMAGE.nahR.R	attcttcacgagcattcatttttttataaatgtgtctccggctatgtc
pCRISPR	pCRISPR.BB.F	gaatttttcaacttgcccactcctgcaggataaaaaaattgtagataa
	pCRISPR.BB.R	gtcgacgaaaactcctccttaagatttata
pSV600	gRNA.F	aatcgactaacagaacatcggccccggcgagttgc
	gRNA.R	ccgatccccagggcagtgcccgcgattgggcggcc
pTarget_*hsdR*_01	gRNA_01.F	GTCAATGTAGGTTTTAGAGCTAGAAATAGC
	gRNA_01.R	TGGGACATTCTTAATTAATACGACCAGTCTAAAAAG
pTarget_*hsdR_02*	gRNA_02.F	CCACTGTGAAGGTTTTAGAGCTAGAAATAGC
	gRNA_02.R	TCAACCAGTTTAATTAATACGACCAGTCTAAAAAG

pTarget_*hsdR_03*	gRNA_03.F	GGGTTCAATTGTTTTAGAGCTAGAAATAGC
	gRNA_03.R	AGGCGGGCAATTAATTAATACGACCAGTCTAAAAAG
pTarget_HA_*hsdR*	LHA.F	ccgagtcggtgctttttttgaattcGCGGCATCGACGCCAACG
	LHA.R	tgctcgaaccccgcgccgatgcaggcctccatcagcctgcgcgttc
	RHA.F	gaacgcgcaggctgatggaggcctgcatcggcgcggggttcgagca
	RHA.R	ATAACAGGGTAATAGATCTAAGCTTtCCAGTTCCTCTGGCGTGTG
Screening	*hsdR*.IN.F	GGCTATGGCGGCTATCAC
	hsdR.IN.R	GACGACTGTCTTCCACTG
	hsdR.OUT.F	TCAATGAGCAACAGGCAATC
	hsdR.OUT.R	CGATATCCAAGGTCCTGCAC
pTarget_*odhA_01*	gRNA_01.F	TCGACATACGGTTTTAGAGCTAGAAATAGC
	gRNA_01.R	AGAACTGTACTTAATTAATACGACCAGTCTAAAAAG
pTarget_*odhA_02*	gRNA_02.F	GAACTGTACGGTTTTAGAGCTAGAAATAGC
	gRNA_02.R	TTCGACATACTTAATTAATACGACCAGTCTAAAAAG
pTarget_*odhA_03*	gRNA_03.F	CGCCTGACCGGTTTTAGAGCTAGAAATAGC
	gRNA_03.R	GTCCAGGATGTTAATTAATACGACCAGTCTAAAAAG
pTarget_HA_*odhA*	LHA.F	tcggtgctttttttcctgcaggGCGGCAGCATCAGCGTCA
	LHA.R	GTCGCTGGCGTTCAATCCGTCGCACTGATTCACTACCTCCGACCCGCG
	RHA.F	CGCGGGTCGGAGGTAGTGAATCAGTGCGACGGATTGAACGCC
	RHA.R	tgatggagctgcacatgaactcgagATGATGCAGCAGTATCAGAGCAAC
Screening	*odhA*.IN.F	ATGATGCAGCAGTATCAG
	odhA.IN.R	CGGCCACGGTGGGCCATGCC
	odhA.OUT.F	GCCCCGCCTTCGCCGGCG
	odhA.OUT.R	TTCGGAGAGCTGCGGAAC
pTarget_ *phaC*	gRNA_01.F	GCGATATGGGGTTTTAGAGCTAGAAATAGC
	gRNA_01.R	TTCGCTGCCGTTAATTAATACGACCAGTCTAAAAAG
	gRNA_02.F	ACCTGACACGGTTTTAGAGCTAGAAATAGC
	gRNA_02.R	CTTCCATCATTTAATTAATACGACCAGTCTAAAAAG
	gRNA_03.F	tgatgcacggGTTTTAGAGCTAGAAATAGC
	gRNA_03.R	acaagtactaTTAATTAATACGACCAGTCTAAAAAG
pTarget_ HA _ *phaC*	LHA.F	agatctattaccctgttatccctaactagtGCCAGCCAATGGCCACGATG
	LHA.R	caggtccaggatgTGGTGCGCCATGCGTCGC
	RHA.F	gcatggcgcaccaCATCCTGGACCTGCAGCC
	RHA.R	gatatcggatccatatgacgtcgaactagtTGGCCGGCGGGTTGATCA
Screening	*phaC*.IN.F	GCACGACCGGCGCTTCGC
	phaC.IN.R	ATGGCGCACCAGCGAGCTC
	phaC.OUT.F	GAAGCCACGCTGTCACGG
	phaC.OUT.R	CCAGGTCTTGCGGGGTCC
pij23119_EYFP	pij23119_EYFP.F	tgatccgctagtccgaggccgtttaaaccggtattttctccttacg
	pij23119_EYFP.R	tgcggcgatgttaattaaattatacctaggactgag
pHA_lacZα/pTarget_HA _*lacZα*	LHA.F	ccgagtcggtgctttttttgaattcGACGTGCGCCTGATCGCC

	LHA.R	tgcgggcctcAGCTGTTTCCTGTGTGAAATTGTTATCCG
	RHA.F	ggaaacagctGAGGCCCGCACCGATCGC
	RHA.R	tgatggagctgcacatgaactcgagTTCGGGATTTCGGCGCTCC
pTarget *_lacZα*	gRNA.F	AATAGCGAAGGTTTTAGAGCTAGAAATAG
	gRNA.R	ACGCCAGCTGACTAGTATTATACCTAGGAC
Screening	crcZfw	TTTTGAATTCCCGATCTGCATTGCGACG
	crcZrev	TTTTTTCTGCAGCCAATACATAAGCAGATGCCGTGC
	LacZfw	TGACCGCTGGGATCTGCC
	LacZrev	GGCAGATCCCAGCGGTCA
pHA*_las/pTarget _HA_las*	LHA.F	ccgagtcggtgctttttttgaattcTGTTCCTCTGGGAGCTGG
	LHA.R	agaggcaagaaagcttAGCGCTACGTTCTTCTTAAAC
	RHA.F	acgtagcgctaagctttcttgcctctcaggtcgg
	RHA.R	atggagctgcacatgaaCTCGAGacggtggttaccgtcacc
pTarget *_las*	gRNA.F	TCTTCAGGTGGTTTTAGAGCTAGAAATAG
	gRNA.R	TCGGCATCGAACTAGTATTATACCTAGGAC
pHA_lasR/pTarget_HA *_lasR*	LHA.F	ccgagtcggtgctttttttgaattcTGTTCCTCTGGGAGCTGG
	LHA.R	caggtccccgaagcttAGCGCTACGTTCTTCTTAAAC
	RHA.F	acgtagcgctaagcttCGGGGACCTGTCGGCTCG
	RHA.R	atggagctgcacatgaactcgagGACGGAGCGCCGCCTGCT
	gRNA.F	AAATATTCGGGTTTTAGAGCTAGAAATAG
pTarget _lasR	gRNA.R	CCCATATGGAACTAGTATTATACCTAGGAC
Screening	LasR.OUT.R	ATCTGGGTCTTGGCATTGAG
	LasR.IN.F	GGGAGAAGGAAGTGTTGCAG
	LasR.IN.R	ACCCAAATTAACGGCCATA
	Las.OUT.R	GGTCGCCTATCTCGGTATCA
pCRISPR_574 (+)	pCRISPR.574	tataaatcttaaggaggagttttcgtcgacgatgatggaagacctgacacgGTTTTAGAGCTAGAAATA
pCRISPR_737 (-)	pCRISPR.737	tataaatcttaaggaggagttttcgtcgacgtagtacttgttgatgcacggGTTTTAGAGCTAGAAATAGCAAG
	pCRISPR.gRNA.R	tgtacatcgtggccattggctggcactagtGACGTCATAAAAATAAGAAGC
	Chapter 5 - **Ethylene plasmids**	
pEFE(P)	pEFE(P).Kp.F	tccccgaaaagtgccacctgcatcgatttgtttaaacAGCGGATATAAAAACCGTTATTG
	pEFE(P).Kp.R	taggttggtcatttaattaaTATATCTCCTTCTTAAAAAGGTC
	EFE(P).F	agaaggagatatattaattaaATGACCAACCTACAGACTTTCGAGTTG
	EFE(P).R	gtgaggatgcgactagtTCATGAGCCTGTCGCGCG
	pEFE(P).tl3.F	gacaggctcatgaactagtCGCATCCTCACGATAATATC
	pEFE(P).tl3.R	acgaccctgccctgaaccgacgaccgggtcCTACTGGTATTGGCACAAAC
pHokSok	HosSok.F	cgaaaagtgccacctgacgtcgaattcACAACATCAGCAAGGAGAAAGG
	HosSok.R	cgcctcgagggatccAACAAACTCCGGGAGGCAG
	pHS.catP.F	ggagtttgttggatccctcgagGCGAGACCTTAGGGTAAC
	pHS.catP.R	gatgctgaagatcagttgggtgcacAGTGGGCAAGTTGAAAAATTC

prpoD_E	rpoDE.F	AACCTAGGAGCTCTGATTTAACGGCTTAAGTGCCGAAGAGC
	rpoDE.R	TGGAAGCTTTAACGCCTGATCCGGCCTACCGATTAAT
pLacZa_E	lacZa.F	ctcactgatccgctagtccgaggcctcgagGGCTGGCTTAACTATGCG
	lacZa.R	cgatatcggatccatatgacgtcgacgcgtgcgcccaatacgcaaacc

Table A.2– Genomic variants identified from WGS of the strain ΔproB/pGEM-efep + prpoD_wt (Genewiz®)

Gene	Region	Type	Length	Count	Coverage
IS1I	257908	SNV	1	312	315
gatC	2173361..2173362	Deletion	2	971	1017
glpR	3560455^3560456	Insertion	1	1108	1113
REP321j	4296380^4296381	Insertion	2	609	638

Gene	Frequency	Probability	Forward/reverse balance	Average quality
IS1I	99.04761905	1	0.098265896	39.46474359
gatC	95.47689282	1	0.455770851	35.54147347
glpR	99.5507637	1	0.396174863	39.19404332
REP321j	95.45454545	1	0.386897404	39.57032984

Table A.3- Genomic variants identified from WGS of the strain D0.07_14 (Genewiz®)

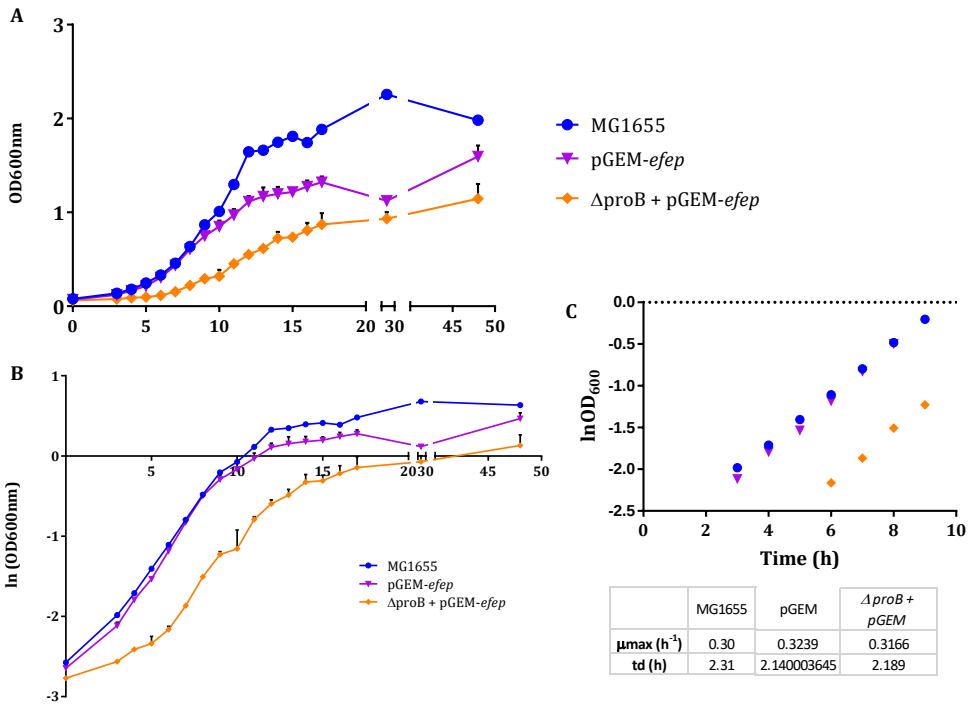
Gene	Region	Type	Length	Count	Coverage
fhuA	167577..167580	Deletion	4	796	875
cyoB	449355	SNV	1	888	1745
	449373	SNV	1	945	1793
	449391	SNV	1	1099	1879
	449435	SNV	1	1188	1985
	449515	SNV	1	893	1724
	449519	SNV	1	906	1726
	449616	SNV	1	1044	1920
	449619	SNV	1	1042	1921
	449624..449625	MNV	2	1043	1948
	449631	SNV	1	1042	1961
	449634	SNV	1	1026	1986
alaU	3426967..3426968	Replacement	2	712	939
	3426970^3426971	Insertion	1	695	920
	3426984..3426985	MNV	2	728	938
	3426990	SNV	1	713	910
	3426996	SNV	1	720	918
	3427008	SNV	1	700	887
	3427027..3427028	MNV	2	720	905
rrlA	4038596	SNV	1	515	801
	4038598	SNV	1	521	845
	4038635	SNV	1	965	1367
	4038641	SNV	1	1003	1416
	4038649	SNV	1	1008	1459

	4038662	SNV	1	1010	1504
	4038668	SNV	1	1005	1523
	4038678..4038679	MNV	2	1017	1557
	4038682..4038683	MNV	2	1025	1556
	4038686..4038688	MNV	3	1020	1595
	4038692	SNV	1	993	1584
	4038696..4038697	MNV	2	1022	1664
	4038700..4038702	MNV	3	1027	1690
	4038708	SNV	1	998	1638
	4038718	SNV	1	1011	1683
	4038726	SNV	1	874	1536
	4039289	SNV	1	1389	1714
	4039299	SNV	1	1605	1963

Gene	Frequency	Probability	Forward/reverse balance	Average quality
fhuA	90.97142857	1	0.38586	37.7052
cyoB	50.88825215	1	0.4169	39.1081
	52.70496375	1	0.45423	39.1291
	58.48855774	1	0.43328	39.0064
	59.8488665	1	0.44652	38.3754
	51.79814385	1	0.44301	38.8578
	52.49130939	1	0.44435	39.2693
	54.375	1	0.38769	39.477
	54.24258199	1	0.39497	38.9165
	53.54209446	1	0.37939	38.8562
	53.13615502	1	0.38138	39.6008
	51.66163142	1	0.38356	39.6842
alaU	75.82534611	1	0.25949	38.9379
	75.54347826	1	0.2619	39.082
	77.6119403	1	0.28084	39.3289
	78.35164835	1	0.28601	39.1992
	78.43137255	1	0.28311	39.3097
	78.91770011	1	0.28942	38.94
	79.55801105	1	0.28138	38.9716
rrlA	64.29463171	1	0.32012	38.5786
	61.65680473	1	0.32481	39.0173
	70.59253841	1	0.49033	38.5513
	70.83333333	1	0.49371	39.4487
	69.08841672	1	0.48143	39.0516
	67.15425532	1	0.45512	39.2327
	65.98818122	1	0.45648	38.9025
	65.31791908	1	0.37904	39.0395

	65.87403599	1	0.38534	39.2795
	63.94984326	1	0.3851	39.0725
	62.68939394	1	0.38099	39.3696
	61.41826923	1	0.38498	39.3996
	60.76923077	1	0.38426	38.3605
	60.92796093	1	0.38306	38.6222
	60.07130125	1	0.38054	39.4679
	56.90104167	1	0.35708	39.2769
	81.03850642	1	0.26813	39.3398
	81.76260825	1	0.34576	39.362

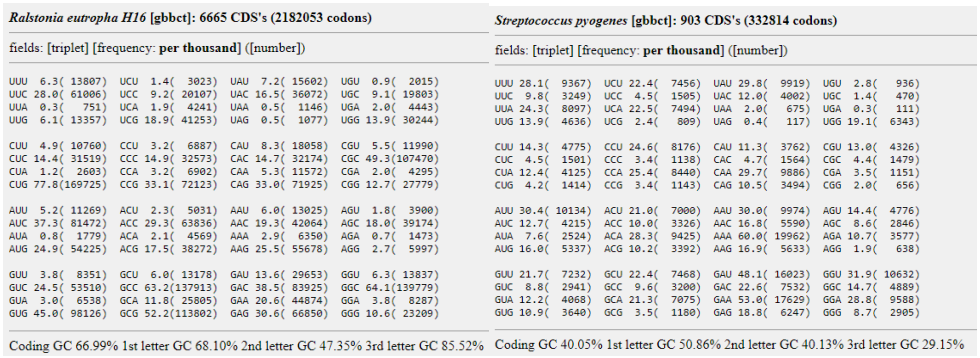
Figures



	MG1655	pGEM	*ΔproB + pGEM*
μmax (h^{-1})	0.30	0.3239	0.3166
td (h)	2.31	2.140003645	2.189

Figure A.1 – Growth course of ethylene-producing strains

Growth course (**A**) and associated logarithmic transformation of OD_{600} values (**B**) of the strains MG1655, MG1655/pGEM-*efep* and MG1655 Δ*proB*/pGEM-*efep* cultivated in 50 mL 0.4% glucose-M9 medium in 250 mL baffled flasks at 30°C, shaking 48h.
(**C**) The exponential phase was identified between t = 3h and t = 9h for the strains MG1655 and MG1655/pGEM-*efep* while the exponential phase of Δ*proB*/pGEM-*efep* was delayed between t = 6h and t = 9h.



Ralstonia eutropha H16 **[gbbct]: 6665 CDS's (2182053 codons)**

fields: [triplet] [frequency: **per thousand**] ([number])

```
UUU  6.3( 13807)  UCU  1.4(  3023)  UAU  7.2( 15602)  UGU  0.9(  2015)
UUC 28.0( 61006)  UCC  9.2( 20107)  UAC 16.5( 36072)  UGC  9.1( 19803)
UUA  0.3(   751)  UCA  1.9(  4241)  UAA  0.5(  1146)  UGA  2.0(  4443)
UUG  6.1( 13357)  UCG 18.9( 41253)  UAG  0.5(  1077)  UGG 13.9( 30244)

CUU  4.9( 10760)  CCU  3.2(  6887)  CAU  8.3( 18058)  CGU  5.5( 11990)
CUC 14.4( 31519)  CCC 14.9( 32573)  CAC 14.7( 32174)  CGC 49.3(107470)
CUA  1.2(  2603)  CCA  3.2(  6902)  CAA  5.3( 11572)  CGA  2.0(  4295)
CUG 77.8(169725)  CCG 33.1( 72123)  CAG 33.0( 71925)  CGG 12.7( 27779)

AUU  5.2( 11269)  ACU  2.3(  5031)  AAU  6.0( 13025)  AGU  1.8(  3900)
AUC 37.3( 81472)  ACC 29.3( 63836)  AAC 19.3( 42064)  AGC 18.0( 39174)
AUA  0.8(  1779)  ACA  2.1(  4569)  AAA  2.9(  6350)  AGA  0.7(  1473)
AUG 24.9( 54225)  ACG 17.5( 38272)  AAG 25.5( 55678)  AGG  2.7(  5997)

GUU  3.8(  8351)  GCU  6.0( 13178)  GAU 13.6( 29653)  GGU  6.3( 13837)
GUC 24.5( 53510)  GCC 63.2(137913)  GAC 38.5( 83925)  GGC 64.1(139779)
GUA  3.0(  6538)  GCA 11.8( 25805)  GAA 20.6( 44874)  GGA  3.8(  8287)
GUG 45.0( 98126)  GCG 52.2(113802)  GAG 30.6( 66850)  GGG 10.6( 23209)
```

Coding GC 66.99% 1st letter GC 68.10% 2nd letter GC 47.35% 3rd letter GC 85.52%

Streptococcus pyogenes **[gbbct]: 903 CDS's (332814 codons)**

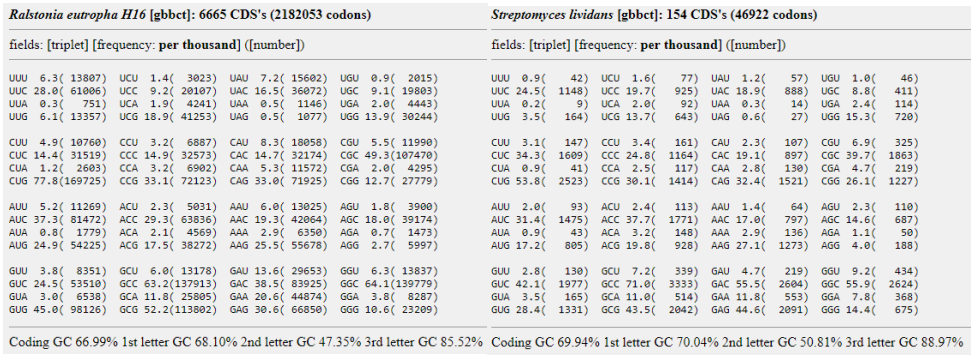
fields: [triplet] [frequency: **per thousand**] ([number])

```
UUU 28.1(  9367)  UCU 22.4(  7456)  UAU 29.8(  9919)  UGU  2.8(   936)
UUC  9.8(  3249)  UCC  4.5(  1505)  UAC 12.0(  4002)  UGC  1.4(   470)
UUA 24.3(  8097)  UCA 22.5(  7494)  UAA  2.0(   675)  UGA  0.3(   111)
UUG 13.9(  4636)  UCG  2.4(   809)  UAG  0.4(   117)  UGG 19.1(  6343)

CUU 14.3(  4775)  CCU 24.6(  8176)  CAU 11.3(  3762)  CGU 13.0(  4326)
CUC  4.5(  1501)  CCC  3.4(  1138)  CAC  4.7(  1564)  CGC  4.4(  1479)
CUA 12.4(  4125)  CCA 25.4(  8440)  CAA 29.7(  9886)  CGA  3.5(  1151)
CUG  4.2(  1414)  CCG  3.4(  1143)  CAG 10.5(  3494)  CGG  2.0(   656)

AUU 30.4( 10134)  ACU 21.0(  7000)  AAU 30.0(  9974)  AGU 14.4(  4776)
AUC 12.7(  4215)  ACC 10.0(  3326)  AAC 16.8(  5590)  AGC  8.6(  2846)
AUA  7.6(  2524)  ACA 28.3(  9425)  AAA 60.0( 19962)  AGA 10.7(  3577)
AUG 16.0(  5337)  ACG 10.2(  3392)  AAG 16.9(  5633)  AGG  1.9(   638)

GUU 21.7(  7232)  GCU 22.4(  7468)  GAU 48.1( 16023)  GGU 31.9( 10632)
GUC  8.8(  2941)  GCC  9.6(  3200)  GAC 22.6(  7532)  GGC 14.7(  4889)
GUA 12.2(  4068)  GCA 21.3(  7075)  GAA 53.0( 17629)  GGA 28.8(  9588)
GUG 10.9(  3640)  GCG  3.5(  1180)  GAG 18.8(  6247)  GGG  8.7(  2905)
```

Coding GC 40.05% 1st letter GC 50.86% 2nd letter GC 40.13% 3rd letter GC 29.15%

Figure A.2 – Codon usage tables of C. necator H16 (formerly R. eutropha H16) and S. pyogenes

Codon usage tables from the Codon Usage Database (https://www.kazusa.or.jp/codon/) developed from NCBI-GenBank Flat File Release 160.0 (15/6/2007) data source.



Ralstonia eutropha H16 **[gbbct]: 6665 CDS's (2182053 codons)**

fields: [triplet] [frequency: **per thousand**] ([number])

```
UUU  6.3( 13807)  UCU  1.4(  3023)  UAU  7.2( 15602)  UGU  0.9(  2015)
UUC 28.0( 61006)  UCC  9.2( 20107)  UAC 16.5( 36072)  UGC  9.1( 19803)
UUA  0.3(   751)  UCA  1.9(  4241)  UAA  0.5(  1146)  UGA  2.0(  4443)
UUG  6.1( 13357)  UCG 18.9( 41253)  UAG  0.5(  1077)  UGG 13.9( 30244)

CUU  4.9( 10760)  CCU  3.2(  6887)  CAU  8.3( 18058)  CGU  5.5( 11990)
CUC 14.4( 31519)  CCC 14.9( 32573)  CAC 14.7( 32174)  CGC 49.3(107470)
CUA  1.2(  2603)  CCA  3.2(  6902)  CAA  5.3( 11572)  CGA  2.0(  4295)
CUG 77.8(169725)  CCG 33.1( 72123)  CAG 33.0( 71925)  CGG 12.7( 27779)

AUU  5.2( 11269)  ACU  2.3(  5031)  AAU  6.0( 13025)  AGU  1.8(  3900)
AUC 37.3( 81472)  ACC 29.3( 63836)  AAC 19.3( 42064)  AGC 18.0( 39174)
AUA  0.8(  1779)  ACA  2.1(  4569)  AAA  2.9(  6350)  AGA  0.7(  1473)
AUG 24.9( 54225)  ACG 17.5( 38272)  AAG 25.5( 55678)  AGG  2.7(  5997)

GUU  3.8(  8351)  GCU  6.0( 13178)  GAU 13.6( 29653)  GGU  6.3( 13837)
GUC 24.5( 53510)  GCC 63.2(137913)  GAC 38.5( 83925)  GGC 64.1(139779)
GUA  3.0(  6538)  GCA 11.8( 25805)  GAA 20.6( 44874)  GGA  3.8(  8287)
GUG 45.0( 98126)  GCG 52.2(113802)  GAG 30.6( 66850)  GGG 10.6( 23209)
```

Coding GC 66.99% 1st letter GC 68.10% 2nd letter GC 47.35% 3rd letter GC 85.52%

Streptomyces lividans **[gbbct]: 154 CDS's (46922 codons)**

fields: [triplet] [frequency: **per thousand**] ([number])

```
UUU  0.9(    42)  UCU  1.6(    77)  UAU  1.2(    57)  UGU  1.0(    46)
UUC 24.5(  1148)  UCC 19.7(   925)  UAC 18.9(   888)  UGC  8.8(   411)
UUA  0.2(     9)  UCA  2.0(    92)  UAA  0.3(    14)  UGA  2.4(   114)
UUG  3.5(   164)  UCG 13.7(   643)  UAG  0.6(    27)  UGG 15.3(   720)

CUU  3.1(   147)  CCU  3.4(   161)  CAU  2.3(   107)  CGU  6.9(   325)
CUC 34.3(  1609)  CCC 24.8(  1164)  CAC 19.1(   897)  CGC 39.7(  1863)
CUA  0.9(    41)  CCA  2.5(   117)  CAA  2.8(   130)  CGA  4.7(   219)
CUG 53.8(  2523)  CCG 30.1(  1414)  CAG 32.4(  1521)  CGG 26.1(  1227)

AUU  2.0(    93)  ACU  2.4(   113)  AAU  1.4(    64)  AGU  2.3(   110)
AUC 31.4(  1475)  ACC 37.7(  1771)  AAC 17.0(   797)  AGC 14.6(   687)
AUA  0.9(    43)  ACA  3.2(   148)  AAA  2.9(   136)  AGA  1.1(    50)
AUG 17.2(   805)  ACG 19.8(   928)  AAG 27.1(  1273)  AGG  4.0(   188)

GUU  2.8(   130)  GCU  7.2(   339)  GAU  4.7(   219)  GGU  9.2(   434)
GUC 42.1(  1977)  GCC 71.0(  3333)  GAC 55.5(  2604)  GGC 55.9(  2624)
GUA  3.5(   165)  GCA 11.0(   514)  GAA 11.8(   553)  GGA  7.8(   368)
GUG 28.4(  1331)  GCG 43.5(  2042)  GAG 44.6(  2091)  GGG 14.4(   675)
```

Coding GC 69.94% 1st letter GC 70.04% 2nd letter GC 50.81% 3rd letter GC 88.97%

Figure A.3 - Codon usage tables of C. necator H16 (formerly R. eutropha H16) and S. lividans

Codon usage tables from the Codon Usage Database (https://www.kazusa.or.jp/codon/) developed from NCBI-GenBank Flat File Release 160.0 [15/7/07] data source. *S.lividans* was one of the three *Streptomyces* species in which Ryan Cobb and his team successfully developed a CRISPR/Cas9 system, based on a codon-optimised *cas9* (Cobb, Wang and Zhao, 2015)

Figure A.4– Sequence alignment of the CRISPR-based ΔphaC mutants

The left homology arm is highlighted in red and the right homology arm appears in blue.

```
phaC            1
AGCGCTGCATACCGTCCGGTAGGTCGGGAAGCGTGCAGTGCCGAGGCGGATTCCCGCATT
574_+__1.1      1
AGCGCTGCATACCGTCCGGTAGGTCGGGAAGCGTGCAGTGCCGAGGCGGATTCCCGCATT
574_+__1.3      1
AGCGCTGCATACCGTCCGGTAGGTCGGGAAGCGTGCAGTGCCGAGGCGGATTCCCGCATT
574_+__1.4      1
AGCGCTGCATACCGTCCGGTAGGTCGGGAAGCGTGCAGTGCCGAGGCGGATTCCCGCATT
574_+__1.5      1
```

```
AGCGCTGCATACCGTCCGGTAGGTCGGGAAGCGTGCAGTGCCGAGGCGGATTCCCGCATT
574_+__2.3    1
AGCGCTGCATACCGTCCGGTAGGTCGGGAAGCGTGCAGTGCCGAGGCGGATTCCCGCATT


phaC          61
GACAGCGCGTGCGTTGCAAGGCAACAATGGACTCAAATGTCTCGGAATCGCTGACGATTC
574_+__1.1    61
GACAGCGCGTGCGTTGCAAGGCAACAATGGACTCAAATGTCTCGGAATCGCTGACGATTC
574_+__1.3    61
GACAGCGCGTGCGTTGCAAGGCAACAATGGACTCAAATGTCTCGGAATCGCTGACGATTC
574_+__1.4    61
GACAGCGCGTGCGTTGCAAGGCAACAATGGACTCAAATGTCTCGGAATCGCTGACGATTC
574_+__1.5    61
GACAGCGCGTGCGTTGCAAGGCAACAATGGACTCAAATGTCTCGGAATCGCTGACGATTC
574_+__2.3    61
GACAGCGCGTGCGTTGCAAGGCAACAATGGACTCAAATGTCTCGGAATCGCTGACGATTC


phaC          121
CCAGGTTTCTCCGGCAAGCATAGCGCATGGCGTCTCCATGCGAGAATGTCGCGCTTGCCG
574_+__1.1    121
CCAGGTTTCTCCGGCAAGCATAGCGCATGGCGTCTCCATGCGAGAATGTCGCGCTTGCCG
574_+__1.3    121
CCAGGTTTCTCCGGCAAGCATAGCGCATGGCGTCTCCATGCGAGAATGTCGCGCTTGCCG
574_+__1.4    121
CCAGGTTTCTCCGGCAAGCATAGCGCATGGCGTCTCCATGCGAGAATGTCGCGCTTGCCG
574_+__1.5    121
CCAGGTTTCTCCGGCAAGCATAGCGCATGGCGTCTCCATGCGAGAATGTCGCGCTTGCCG
574_+__2.3    121
CCAGGTTTCTCCGGCAAGCATAGCGCATGGCGTCTCCATGCGAGAATGTCGCGCTTGCCG


phaC          181
GATAAAAGGGGAGCCGCTATCGGAATGGACGCAAGCCACGGCCGCAGCAGGTGCGGTCGA
574_+__1.1    181
GATAAAAGGGGAGCCGCTATCGGAATGGACGCAAGCCACGGCCGCAGCAGGTGCGGTCGA
574_+__1.3    181
GATAAAAGGGGAGCCGCTATCGGAATGGACGCAAGCCACGGCCGCAGCAGGTGCGGTCGA
574_+__1.4    181
GATAAAAGGGGAGCCGCTATCGGAATGGACGCAAGCCACGGCCGCAGCAGGTGCGGTCGA
574_+__1.5    181
GATAAAAGGGGAGCCGCTATCGGAATGGACGCAAGCCACGGCCGCAGCAGGTGCGGTCGA
574_+__2.3    181
GATAAAAGGGGAGCCGCTATCGGAATGGACGCAAGCCACGGCCGCAGCAGGTGCGGTCGA


phaC          241
GGGCTTCCAGCCAGTTCCAGGGCAGATGTGCCGGCAGACCCTCCCGCTTTGGGGGAGGCG
574_+__1.1    241
GGGCTTCCAGCCAGTTCCAGGGCAGATGTGCCGGCAGACCCTCCCGCTTTGGGGGAGGCG
574_+__1.3    241
GGGCTTCCAGCCAGTTCCAGGGCAGATGTGCCGGCAGACCCTCCCGCTTTGGGGGAGGCG
574_+__1.4    241
GGGCTTCCAGCCAGTTCCAGGGCAGATGTGCCGGCAGACCCTCCCGCTTTGGGGGAGGCG
574_+__1.5    241
GGGCTTCCAGCCAGTTCCAGGGCAGATGTGCCGGCAGACCCTCCCGCTTTGGGGGAGGCG
574_+__2.3    241
```

```
GGGCTTCCAGCCAGTTCCAGGGCAGATGTGCCGGCAGACCCTCCCGCTTTGGGGGAGGCG


phaC        301
CAAGCCGGGTCCATTCGGATAGCATCTCCCCATGCAAAGTGCCGGCCAGGGCAATGCCCG
574_+__1.1  301
CAAGCCGGGTCCATTCGGATAGCATCTCCCCATGCAAAGTGCCGGCCAGGGCAATGCCCG
574_+__1.3  301
CAAGCCGGGTCCATTCGGATAGCATCTCCCCATGCAAAGTGCCGGCCAGGGCAATGCCCG
574_+__1.4  301
CAAGCCGGGTCCATTCGGATAGCATCTCCCCATGCAAAGTGCCGGCCAGGGCAATGCCCG
574_+__1.5  301
CAAGCCGGGTCCATTCGGATAGCATCTCCCCATGCAAAGTGCCGGCCAGGGCAATGCCCG
574_+__2.3  301
CAAGCCGGGTCCATTCGGATAGCATCTCCCCATGCAAAGTGCCGGCCAGGGCAATGCCCG


phaC        361
GAGCCGGTTCGAATAGTGACGGCAGAGAGACAATCAAATCATGGCGACCGGCAAAGGCGC
574_+__1.1  361
GAGCCGGTTCGAATAGTGACGGCAGAGAGACAATCAAATCATGGCGACCGGCAAAGGCGC
574_+__1.3  361
GAGCCGGTTCGAATAGTGACGGCAGAGAGACAATCAAATCATGGCGACCGGCAAAGGCGC
574_+__1.4  361
GAGCCGGTTCGAATAGTGACGGCAGAGAGACAATCAAATCATGGCGACCGGCAAAGGCGC
574_+__1.5  361
GAGCCGGTTCGAATAGTGACGGCAGAGAGACAATCAAATCATGGCGACCGGCAAAGGCGC
574_+__2.3  361
GAGCCGGTTCGAATAGTGACGGCAGAGAGACAATCAAATCATGGCGACCGGCAAAGGCGC


phaC        421
GGCAGCTTCCACGCAGGAAGGCAAGTCCCAACCATTCAAGGTCACGCCGGGGCCATTCGA
574_+__1.1  421
GGCAGCTTCCACGCAGGAAGGCAAGTCCCAACCATTCAAGGTCACGCCGGGGCCATTCGA
574_+__1.3  421
GGCAGCTTCCACGCAGGAAGGCAAGTCCCAACCATTCAAGGTCACGCCGGGGCCATTCGA
574_+__1.4  421
GGCAGCTTCCACGCAGGAAGGCAAGTCCCAACCATTCAAGGTCACGCCGGGGCCATTCGA
574_+__1.5  421
GGCAGCTTCCACGCAGGAAGGCAAGTCCCAACCATTCAAGGTCACGCCGGGGCCATTCGA
574_+__2.3  421
GGCAGCTTCCACGCAGGAAGGCAAGTCCCAACCATTCAAGGTCACGCCGGGGCCATTCGA


phaC        481
TCCAGCCACATGGCTGGAATGGTCCCGCCAGTGGCAGGGCACTGAAGGCAACGGCCACGC
574_+__1.1  481
TCCAGCCACATGGCTGGAATGGTCCCGCCAGTGGCAGGGCACTGAAGGCAACGGCCACGC
574_+__1.3  481
TCCAGCCACATGGCTGGAATGGTCCCGCCAGTGGCAGGGCACTGAAGGCAACGGCCACGC
574_+__1.4  481
TCCAGCCACATGGCTGGAATGGTCCCGCCAGTGGCAGGGCACTGAAGGCAACGGCCACGC
574_+__1.5  481
TCCAGCCACATGGCTGGAATGGTCCCGCCAGTGGCAGGGCACTGAAGGCAACGGCCACGC
574_+__2.3  481
TCCAGCCACATGGCTGGAATGGTCCCGCCAGTGGCAGGGCACTGAAGGCAACGGCCACGC
```

```
phaC        541
GGCCGCGTCCGGCATTCCGGGCCTGGATGCGCTGGCAGGCGTCAAGATCGCGCCGGCGCA
574_+__1.1  541
GGCCGCGTCCGGCATTCCGGGCCTGGATGCGCTGGCAGGCGTCAAGATCGCGCCGGCGCA
574_+__1.3  541
GGCCGCGTCCGGCATTCCGGGCCTGGATGCGCTGGCAGGCGTCAAGATCGCGCCGGCGCA
574_+__1.4  541
GGCCGCGTCCGGCATTCCGGGCCTGGATGCGCTGGCAGGCGTCAAGATCGCGCCGGCGCA
574_+__1.5  541
GGCCGCGTCCGGCATTCCGGGCCTGGATGCGCTGGCAGGCGTCAAGATCGCGCCGGCGCA
574_+__2.3  541
GGCCGCGTCCGGCATTCCGGGCCTGGATGCGCTGGCAGGCGTCAAGATCGCGCCGGCGCA


phaC        601
GCTGGGTGATATCCAGCAGCGCTACATGAAGGACTTCTCAGCGCTGTGGCAGGCCATGGC
574_+__1.1  601
GCTGGGTGATATCCAGCAGCGCTACATGAAGGACTTCTCAGCGCTGTGGCAGGCCATGGC
574_+__1.3  601
GCTGGGTGATATCCAGCAGCGCTACATGAAGGACTTCTCAGCGCTGTGGCAGGCCATGGC
574_+__1.4  601
GCTGGGTGATATCCAGCAGCGCTACATGAAGGACTTCTCAGCGCTGTGGCAGGCCATGGC
574_+__1.5  601
GCTGGGTGATATCCAGCAGCGCTACATGAAGGACTTCTCAGCGCTGTGGCAGGCCATGGC
574_+__2.3  601
GCTGGGTGATATCCAGCAGCGCTACATGAAGGACTTCTCAGCGCTGTGGCAGGCCATGGC


phaC        661
CGAGGGCAAGGCCGAGGCCACCGGTCCGCTGCACGACCGGCGCTTCGCCGGCGACGCATG
574_+__1.1  661
CGAGGGCAAGGCCGAGGCCACCGGTCCGCTGCACGACCGGCGCTTCGCCGGCGACGCATG
574_+__1.3  661
CGAGGGCAAGGCCGAGGCCACCGGTCCGCTGCACGACCGGCGCTTCGCCGGCGACGCATG
574_+__1.4  661
CGAGGGCAAGGCCGAGGCCACCGGTCCGCTGCACGACCGGCGCTTCGCCGGCGACGCATG
574_+__1.5  661
CGAGGGCAAGGCCGAGGCCACCGGTCCGCTGCACGACCGGCGCTTCGCCGGCGACGCATG
574_+__2.3  661
CGAGGGCAAGGCCGAGGCCACCGGTCCGCTGCACGACCGGCGCTTCGCCGGCGACGCATG


phaC        721
GCGCACCAACCTCCCATATCGCTTCGCTGCCGCGTTCTACCTGCTCAATGCGCGCGCCTT
574_+__1.1  721
GCGCACCA--------------------------------------------------
574_+__1.3  721
GCGCACCA--------------------------------------------------
574_+__1.4  721
GCGCACCA--------------------------------------------------
574_+__1.5  721
GCGCACCA--------------------------------------------------
574_+__2.3  721
GCGCACCA--------------------------------------------------


phaC        781
GACCGAGCTGGCCGATGCCGTCGAGGCCGATGCCAAGACCCGCCAGCGCATCCGCTTCGC
```

```
574_+__1.1  729 ----------------------------------------------
-------------
574_+__1.3  729 ----------------------------------------------
-------------
574 +  1.4  729 ----------------------------------------------
-------------
574_+__1.5  729 ----------------------------------------------
-------------
574_+__2.3  729 ----------------------------------------------
-------------


phaC        841
GATCTCGCAATGGGTCGATGCGATGTCGCCCGCCAACTTCCTTGCCACCAATCCCGAGGC
574_+__1.1  729 ----------------------------------------------
-------------
574_+__1.3  729 ----------------------------------------------
-------------
574_+__1.4  729 ----------------------------------------------
-------------
574_+__1.5  729 ----------------------------------------------
-------------
574_+__2.3  729 ----------------------------------------------
-------------


phaC        901
GCAGCGCCTGCTGATCGAGTCGGGCGGCGAATCGCTGCGTGCCGGCGTGCGCAACATGAT
574_+__1.1  729 ----------------------------------------------
-------------
574_+__1.3  729 ----------------------------------------------
-------------
574_+__1.4  729 ----------------------------------------------
-------------
574_+__1.5  729 ----------------------------------------------
-------------
574_+__2.3  729 ----------------------------------------------
-------------


phaC        961
GGAAGACCTGACACGCGGCAAGATCTCGCAGACCGACGAGAGCGCGTTTGAGGTCGGCCG
574_+__1.1  729 ----------------------------------------------
-------------
574_+__1.3  729 ----------------------------------------------
-------------
574_+__1.4  729 ----------------------------------------------
-------------
574_+__1.5  729 ----------------------------------------------
-------------
574_+__2.3  729 ----------------------------------------------
-------------


phaC       1021
CAATGTCGCGGTGACCGAAGGCGCCGTGGTCTTCGAGAACGAGTACTTCCAGCTGTTGCA
574_+__1.1  729 ----------------------------------------------
-------------
```

```
574_+__1.3  729 ----------------------------------------------
-------------
574_+__1.4  729 ----------------------------------------------
-------------
574 +  1.5  729 ----------------------------------------------
-------------
574_+__2.3  729 ----------------------------------------------
-------------
```

```
phaC       1081
GTACAAGCCGCTGACCGACAAGGTGCACGCGCGCCCGCTGCTGATGGTGCCGCCGTGCAT
574_+__1.1  729 ------------------------------------------------
-------------
574 +  1.3  729 ------------------------------------------------
-------------
574_+__1.4  729 ------------------------------------------------
-------------
574_+__1.5  729 ------------------------------------------------
-------------
574_+__2.3  729 ------------------------------------------------
-------------


phaC       1141
CAACAAGTACTACATCCTGGACCTGCAGCCGGAGAGCTCGCTGGTGCGCCATGTGGTGGA
574_+__1.1  729 ------------
CATCCTGGACCTGCAGCCGGAGAGCTCGCTGGTGCGCCAT--------
574_+__1.3  729 ------------
CATCCTGGACCTGCAGCCGGAGAGCTCGCTGGTGCGCCAT--------
574_+__1.4  729 ------------
CATCCTGGACCTGCAGCCGGAGAGCTCGCTGGTGCGCCATGTGGTGGA
574_+__1.5  729 ------------
CATCCTGGACCTGCAGCCGGAGAGCTCGCTGGTGCGCCAT--------
574_+__2.3  729 ------------
CATCCTGGACCTGCAGCCGGAGAGCTCGCTGGTGCGCCAT--------


phaC       1201
GCAGGGACATACGGTGTTTCTGGTGTCGTGGCGCAATCCGGACGCCAGCATGGCCGGCAG
574_+__1.1      ------------------------------------------------
-------------
574_+__1.3      ------------------------------------------------
-------------
574_+__1.4  777 GCAGGGACATACGGTGTTTCTGGTGT----------------------
-------------
574_+__1.5      ------------------------------------------------
-------------
574_+__2.3      ------------------------------------------------
-------------


phaC       1261
CACCTGGGACGACTACATCGAGCACGCGGCCATCCGCGCCATCGAAGTCGCGCGCGACAT
574_+__1.1      ------------------------------------------------
-------------
574_+__1.3      ------------------------------------------------
-------------
574_+__1.4      ------------------------------------------------
-------------
574_+__1.5      ------------------------------------------------
-------------
574_+__2.3      ------------------------------------------------
-------------


phaC       1321
CAGCGGCCAGGACAAGATCAACGTGCTCGGCTTCTGCGTGGGCGGCACCATTGTCTCGAC
```

```
574_+__1.1      ----------------------------------------------
-------------
574_+__1.3      ----------------------------------------------
-------------
574_+__1.4      ----------------------------------------------
-------------
574_+__1.5      ----------------------------------------------
-------------
574_+__2.3      ----------------------------------------------
-------------


phaC       1381
CGCGCTGGCGGTGCTGGCCGCGCGCGGCGAGCACCCGGCCGCCAGCGTCACGCTGCTGAC
574_+__1.1      ----------------------------------------------
-------------
574_+__1.3      ----------------------------------------------
-------------
574_+__1.4      ----------------------------------------------
-------------
574_+__1.5      ----------------------------------------------
-------------
574_+__2.3      ----------------------------------------------
-------------


phaC       1441
CACGCTGCTGGACTTTGCCGACACGGGCATCCTCGACGTCTTTGTCGACGAGGGCCATGT
574_+__1.1      ----------------------------------------------
-------------
574_+__1.3      ----------------------------------------------
-------------
574_+__1.4      ----------------------------------------------
-------------
574_+__1.5      ----------------------------------------------
-------------
574_+__2.3      ----------------------------------------------
-------------


phaC       1501
GCAGTTGCGCGAGGCCACGCTGGGCGGCGGCGCCGGCGCGCCGTGCGCGCTGCTGCGCGG
574_+__1.1      ----------------------------------------------
-------------
574_+__1.3      ----------------------------------------------
-------------
574_+__1.4      ----------------------------------------------
-------------
574_+__1.5      ----------------------------------------------
-------------
574_+__2.3      ----------------------------------------------
-------------


phaC       1561
CCTTGAGCTGGCCAATACCTTCTCGTTCTTGCGCCCGAACGACCTGGTGTGGAACTACGT
574_+__1.1      ----------------------------------------------
-------------
```

```
574_+__1.3      ---------------------------------------------
-------------
574_+__1.4      ---------------------------------------------
-------------
574 +  1.5      ---------------------------------------------
-------------
574_+__2.3      ---------------------------------------------
-------------
```

```
phaC       1621
GGTCGACAACTACCTGAAGGGCAACACGCCGGTGCCGTTCGACCTGCTGTTCTGGAACGG
574_+__1.1      ------------------------------------------------
-------------
574 +  1.3      ------------------------------------------------
-------------
574_+__1.4      ------------------------------------------------
-------------
574_+__1.5      ------------------------------------------------
-------------
574_+__2.3      ------------------------------------------------
-------------


phaC       1681
CGACGCCACCAACCTGCCGGGGCCGTGGTACTGCTGGTACCTGCGCCACACCTACCTGCA
574_+__1.1      ------------------------------------------------
-------------
574_+__1.3      ------------------------------------------------
-------------
574_+__1.4      ------------------------------------------------
-------------
574_+__1.5      ------------------------------------------------
-------------
574_+__2.3      ------------------------------------------------
-------------


phaC       1741
GAACGAGCTCAAGGTACCGGGCAAGCTGACCGTGTGCGGCGTGCCGGTGGACCTGGCCAG
574_+__1.1      ------------------------------------------------
-------------
574_+__1.3      ------------------------------------------------
-------------
574_+__1.4      ------------------------------------------------
-------------
574_+__1.5      ------------------------------------------------
-------------
574_+__2.3      ------------------------------------------------
-------------


phaC       1801
CATCGACGTGCCGACCTATATCTACGGCTCGCGCGAAGACCATATCGTGCCGTGGACCGC
574_+__1.1      ------------------------------------------------
-------------
574_+__1.3      ------------------------------------------------
-------------
574_+__1.4      ------------------------------------------------
-------------
574_+__1.5      ------------------------------------------------
-------------
574_+__2.3      ------------------------------------------------
-------------


phaC       1861
GGCCTATGCCTCGACCGCGCTGCTGGCGAACAAGCTGCGCTTCGTGCTGGGTGCGTCGGG
```

```
574_+__1.1      ----------------------------------------------
-------------
574_+__1.3      ----------------------------------------------
-------------
574 +  1.4      ----------------------------------------------
-------------
574_+__1.5      ----------------------------------------------
-------------
574_+__2.3      ----------------------------------------------
-------------


phaC        1921 CCATATCGCCGGTGTGATCAACCCGCCGGCCA
574_+__1.1       --------------------------------
574_+__1.3       --------------------------------
574_+__1.4       --------------------------------
574_+__1.5       --------------------------------
574_+__2.3       --------------------------------
```

Figure A0.4 – Sequence alignment of the efe gene from ALE isolated DproB/pGEM-efep + prpoD mutants

The promoter p15 is in red while the start codon of the *efe* gene is highlighted in yellow.

```
pGEM-efep     1
TTGACAGCTAGCTTCAGTCCTAGGTACTGTGCTAGCAACCTGAATTCACTAGTTTAACTT
D0.02_15      1
TTGACTGCTAGCTTCNNTCCTAGGNACAGTGCTAGCTTAATGAATTCACTAGTTTAACTT
D0.05_2       1
TTGACGGCTAGCTTCAGTCCTAGGTACAGTGCTAGCTTAATGAATTCACTAGTTTAACTT
D0.07_14      1
TTGACGGCTAGCTTCAGTCCTAGGTACAGTGCTAGCTTAATGAATTCACTAGTTTAACTT


pGEM-efep    61
TAAGAAGGAGATATACATATGACCAACCTACAGACTTTCGAGTTGCCTACCGAGGTAACC
D0.02_15     61
TAGAAAGGAGATATACATATGACCAACCTACAGACTTTCGAGTTGCCTACCGAGGTAACC
D0.05_2      61
TAAGAAGGAGATATACATATGACCAACCTACAGACTTTCGAGTTGCCTACCGAGGTAACC
D0.07_14     61
TAAGAAGGAGATATACATATGACCAACCTACAGACTTTCGAGTTGCCTACCGAGGTAACC


pGEM-efep   121
GGCTGCGCCGCCGATATCTCATTGGGAAGGGCGCTGATCCAAGCCTGGCAAAAAGATGGC
D0.02_15    121
GGCTGCGCCGCCGATATCTCATTGGGAAGGGCGCTGATCCAAGCCTGGCAAAAAGATGGC
D0.05_2     121
GGCTGCGCCGCCGATATCTCATTGGGAAGGGCGCTGATCCAAGCCTGGCAAAAAGATGGC
D0.07_14    121
```

GGCTGCGCCGCCGATATCTCATTGGGAAGGGCGCTGATCCAAGCCTGGCAAAAAGATGGC

pGEM-*efep* 181
ATTTTTCAGATCAAGACCGATAGTGAGCAGGATCGCAAAACGCAGGAAGCAATGGCTGCT
D0.02_15 181
ATTTTTCAGATCAAGACCGATAGTGAGCAGGATCGCAAAACGCAGGAAGCAATGGCTGCT
D0.05_2 181
ATTTTTCAGATCAAGACCGATAGTGAGCAGGATCGCAAAACGCAGGAAGCAATGGCTGCT
D0.07_14 181
ATTTTTCAGATCAAGACCGATAGTGAGCAGGATCGCAAAACGCAGGAAGCAATGGCTGCT

pGEM-*efep* 241
AGCAAGCAGTTTTGCAAGGAACCGCTGACTTTTAAGAGTAGCTGCGTTAGCGATCTGACC
D0.02_15 241
AGCAAGCAGTTTTGCAAGGAACCGCTGACTTTTAAGAGTAGCTGCGTTAGCGATCTGACC
D0.05_2 241
AGCAAGCAGTTTTGCAAGGAACCGCTGACTTTTAAGAGTAGCTGCGTTAGCGATCTGACC
D0.07_14 241
AGCAAGCAGTTTTGCAAGGAACCGCTGACTTTTAAGAGTAGCTGCGTTAGCGATCTGACC

pGEM-*efep* 301
TACAGCGGCTATGTTGCGTCAGGCGAGGAAGTCACAGCTGGTAAACCTGATTTCCCTGAA
D0.02_15 301
TACAGCGGCTATGTTGCGTCAGGCGAGGAAGTCACAGCTGGTAAACCTGATTTCCCTGAA
D0.05_2 301
TACAGCGGCTATGTTGCGTCAGGCGAGGAAGTCACAGCTGGTAAACCTGATTTCCCTGAA
D0.07_14 301
TACAGCGGCTATGTTGCGTCAGGCGAGGAAGTCACAGCTGGTAAACCTGATTTCCCTGAA

pGEM-*efep* 361
ATCTTCACTGTCTGCAAGGACTTGTCGGTAGGCGATCAGCGTGTAAAAGCCGGCTGGCCT
D0.02_15 361
ATCTTCACTGTCTGCTAGGACTTGTCGGTAGGCGATCAGCGTGTAAAAGCCGGCTGGCCT
D0.05_2 361
ATCTTCACTGTCTGCAAGGACTTGTCGGTAGGCGATCAGCGTGTAAAAGCCGGCTGGCCT
D0.07_14 361
ATCTTCACTGTCTGCAAGGACTTGTCGGTAGGCGATCAGCGTGTAAAAGCCGGCTGGCCT

pGEM-*efep* 421
TGCCATGGTCCGGTGCCATGGCCAAATAACACCTATCAGAAAAGCATGAAGACCTTCATG
D0.02_15 421
TGCCATGGTCCGGTGCCATGGCCAAATAACACCTATCAGAAAAGCATGAAGACCTTCATG
D0.05_2 421
TGCCATGGTCCGGTGCCATGGCCAAATAACACCTATCAGAAAAGCATGAAGACCTTCATG
D0.07_14 421
TGCCATGGTCCGGTGCCATGGCCAAATAACACCTATCAGAAAAGCATGAAGACCTTCATG

pGEM-*efep* 481
GAAGAGCTGGGTTTAGCGGGCGAACGGTTGCTCAAACTGACAGCGCTCGGCTTTGAACTA
D0.02_15 481
GAAGAGCTGGGTTTAGCGGGCGAACGGTTGCTCAAACTGACAGCGCTCGGCTTTGAACTA
D0.05_2 481

```
GAAGAGCTGGGTTTAGCGGGCGAACGGTTGCTCAAACTGACAGCGCTCGGCTTTGAACTA
D0.07_14   481
GAAGAGCTGGGTTTAGCGGGCGAACGGTTGCTCAAACTGACAGCGCTCGGCTTTGAACTA


pGEM-efep  541
CCCATCAACACGTTCACCGACTTAACTCGCGATGGTTGGCACCACATGCGTGTATTACGC
D0.02_15   541
CCCATCAACACGTTCACCGACTTAACTCACGATGGTTGGCACCACATGCGTGTATTACGC
D0.05_2    541
CCCATCAACACGTTCNCNGACTTAACTCGCGATGGTTGGCACCACATGCGTGTATTACGC
D0.07_14   541
CCCATCAACACGTTCACCGACTTAACTCGCGATGGTTGGCACCACATGCGTGTATTACGC


pGEM-efep  601
TTCCCGCCCCAAACATCCACGCTGTCCCGTGGAATTGGTGCGCACACTGACTATGGGTTG
D0.02_15   601
TTCCCGCCCCAAACATCCACGGATTTCCGTGGAATTGGTGCGCACACTGACTATGGGTTG
D0.05_2    601
TTCCCGCCCCAAACATCCACGCTGTCCCGTGGAATTGGTGCGCACACTGACTATGGGTTG
D0.07_14   601
TTCCCGCCCCAAACATCCACGCTGTCCCGTGGAATTGGTGCGCACACTGACTATGGGTTG


pGEM-efep  661
TTGGTAATTGCCGCTCAGGACGATGTTGGTGGCTTATATATTCGCCCTCCAGTCGAGGGA
D0.02_15   661
TTGGTAATTCCCGATCAGGACGATATCGGTGGATTATATATTCGCCCTCCAGTCGAGGGA
D0.05_2    661
TTGGTAATTGCCGCTCAGGACGATGTTGGTGGCTTATATATTCGCCCTCCAGTCGAGGGA
D0.07_14   661
TTGGTAATTGCCGCTCAGGACGATGTTGGTGGCTTATATATTCGCCCTCCAGTCGAGGGA
```

```
pGEM-efep  721
GAGAAGCGTAATCGTAACTGGTTGCCTGGTGAGAGCTCAGCAGGCATGTTTGAGCACGAT
D0.02_15   721
GAGAAGCGTAATCGTAACTGGTTGCCTGGTGAGAGCTCAGCAGGCATGTTTGAGCACGAT
D0.05_2    721
GAGAAGCGTAATCGTAACTGGTTGCCTGGTGAGAGCTCAGCAGGCATGTTTGAGCACGAT
D0.07_14   721
GAGAAGCGTAATCGTAACTGGTTGCCTGGTGAGAGCTCAGCAGGCATGTTTGAGCACGAT


pGEM-efep  781
GAACCTTGGACCTTCGTGACGCCCACCCCAGGCGTGTGGACAGTTTTCCCAGGTGATATC
D0.02_15   781
GAACCTTGGACCTTCGTGACGCCCACCCCAGGCGTGTGGACAGTTTTCCCAGGTGATATC
D0.05_2    781
GAACCTTGGACCTTCGTGACGCCCACCCCAGGCGTGTGGACAGTTTTCCCAGGTGATATC
D0.07_14   781
GAACCTTGGACCTTCGTGACGCCCACCCCAGGCGTGTGGACAGTTTTCCCAGGTGATATC


pGEM-efep  841
TTGCAGTTCATGACCGGCGGCCAGCTGCTTTCCACTCCGCACAAGGTTAAGCTCAATACC
D0.02_15   841
TTGCAGTTCATGACCGGCGGCCAGCTGCTTTCCACTCCGCACAAGGTTAAGCTCAATACC
D0.05_2    841
TTGCAGTTCATGACCGGCGGCCAGCTGCTTTCCACTCCGCACAAGGTTAAGCTCAATACC
D0.07 14   841
TTGCAGTTCATGACCGGCGGCCAGCTGCTTTCCACTCCGCACAAGGTTAAGCTCAATACC


pGEM-efep  901
CGCGAACGTTTCGCCTGCGCTTATTTTCATGAGCCTAATTTTGAAGCATCCGCCTATCCG
D0.02_15   901
CGCGAACGTTTCGCCTGCGCTTATTTTCATGAGCCTAATTTTGAAGCATCCGCCTATCCG
D0.05_2    901
CGCGAACGTTTCGCCTGCGCTTATTTTCATGAGCCTAATTTTGAAGCATCCGCCTATCCG
D0.07_14   901
CGCGAACGTTTCGCCTGCGCTTATTTTCATGAGCCTAATTTTGAAGCATCCGCCTATCCG


pGEM-efep  961
TTGTTCGAGCCCAGCGCCAATGAGCGTATTCATTATGGTGAGCACTTTACCAACATGTTT
D0.02_15   961 TTGTTCGAGCCCAGCGCCAATGAGCGTATTCATTATGGTGAGCACTT-
------------
D0.05_2    961
TTGTTCGAGCCCAGCGCCAATGAGCGTATTCATTATGGTGAGCACTTTACCAACATGTTT
D0.07_14   961
TTGTTCGAGCCCAGCGCCAATGAGCGTATTCATTATGGTGAGCACTTTACCAACATGTTT


pGEM-efep 1021
ATGCGTTGCTATCCAGATCGGATCACCACCCAGAGGATCAACAAGGAGAATCGCCTGGCG
D0.02_15  1008 ----------------------------------------------
------------
D0.05_2   1021 ATGCGTTGCTATCCAGATCG---------------------------
------------
D0.07_14  1021 ATGCGTTGCTATCCAGATCG---------------------------
------------
```

```
pGEM-efep 1081
CACTTGGAGGACTTGAAGAAGTATTCGGACACCCGCGCGACAGGCTCATGA
D0.02_15  1008 ----------------------------------------------------
--T
D0.05_2   1041 ----------------------------G---------------------
-AT
D0.07_14  1041 ----------------------------------------------------
-GA
```