University of Nottingham
UK | CHINA | MALAYSIA

METABOLIC ENGINEERING OF PARAGEOBACILLUS THERMOGLUCOSIDASIUS FOR SUSTAINABLE PRODUCTION OF FUELS AND CHEMICALS

Abubakar Madika, B.Sc., M.Sc.

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

September 2022

Dedicated to my dearest wife, Hafsat Tijjani, and my son, Muhammad Abubakar

DECLARATION

Unless otherwise stated or acknowledged, the work presented in this thesis is my own. No part of this thesis has been submitted previously to the University of Nottingham or any other University or Institution for the award of any other Degree or Diploma.

Abubakar Madika

September 2022

ACKNOWLEDGEMENTS

Firstly, all the praises and thanks to the Almighty Allah, The Most Beneficent and The Most Merciful, for His ceaseless grace, mercy and blessings throughout my PhD degree.

I would like to thank my academic supervisors, Prof. Nigel P. Minton and Dr Ying Zhang, for giving me the opportunity to study for a PhD under their supervision and for their kind support, guidance, and encouragement throughout my PhD studies. It is a great privilege and honour for me to work under their tutelage and guidance.

I am also very grateful to my bench supervisors, Dr Lili Sheng and Dr Matthew Lau, for their constant support, never-ending help and advice throughout my research work. Thank you also to all Synthetic Biology Research Centre (SBRC) Nottingham members, both past and present, for their support and encouragement. It is an honour to know, work with, and share such a great working environment with you. I am also thankful to my friends in Nottingham, who have made my stay more enjoyable.

I would also like to acknowledge and thank the Petroleum Technology Development Fund (PTDF), Nigeria, for funding my PhD.

On a personal note, I want to say a special thanks and express my profound gratitude to all my family and friends, particularly my parents, without whom I would not be the person I am today, for their love, care, support, and encouragement. I honestly could not have done this without them.

Finally, thank you to my dearest wife, Hafsat, for her unconditional love, support, inspiration, patience, and encouragement throughout all the stress of the research work and thesis write-up. It certainly would not have been possible for me to come this far and complete this work without her encouragement and support.

ABSTRACT

The continued reliance on fossil fuel reserves to meet global energy and chemical demands is unsustainable and faces challenges such as the limited supply of fossil fuels and environmental concerns due to the release of greenhouse gases (GHGs), the leading cause of climate change. Thus, alternative, more sustainable, environmentally friendly routes are needed to produce fuels and chemicals. One of the most promising and attractive options is the use of microbial fermentative processes to produce chemicals and fuels from renewable feedstocks, mainly lignocellulosic biomass. Presently, ethanol is the leading biofuel produced and used globally, primarily because it can be made by fermentation technology that has been available for a long time. However, there is a growing interest in developing processes for producing alternative, and superior biofuels to ethanol, such as isobutanol which has higher energy density, reduced vapour pressure and lower hygroscopicity. The thermophile *P. thermoglucosidasius* has a significant attraction as a chassis for producing fuels and chemicals from renewable lignocellulosic feedstock. Therefore, the work presented in this thesis focused on the metabolic engineering of *Parageobacillus thermoglucosidasius* NCIMB 11955 for the production of biofuels (ethanol and isobutanol) and the platform chemical 3-hydroxypropionate (3-HP). The lack of adequate and effective genetic tools has hindered the full exploitation of these bacteria and other thermophiles' potential. Hence, the first part of this study involved the characterisation of the pMTL60000 modular vector series. Followed by the successful implementation of a theophylline-responsive CRISPR/Cas9 genome editing tool for the metabolic engineering of *P. thermoglucosidasius* NCIMB 11955 to produce strains capable of producing ethanol efficiently as a primary fermentation product from barley straw hydrolysate. This study also achieved the engineering of *P. thermoglucosidasius* NCIMB 11955 to make the platform chemical 3-HP via the malonyl-CoA pathway using the native acetyl-CoA carboxylase (ACC) and heterologous expression of malonyl-CoA reductase (MCR) and malonate-semialdehyde reductase (MSR) from either *Chloroflexus aurantiacus*, *Metallosphaera sedula*, *Sulfolobus tokodaii* or *Sulfolobus solfataricus*. As part of this goal, the native catabolism of 3-HP in *P. thermoglucosidasius* NCIMB 11955 was investigated, and a strain incapable of 3-HP degradation was generated and represented an ideal chassis for future 3-HP production. Finally, the study also focused on isobutanol production *via* the consolidated bioprocessing (CBP) approach. While attempts and some progress were made to engineer *P. thermoglucosidasius* NCIMB 11955 to produce isobutanol from glucose, unfortunately, a significant titre of isobutanol was not achieved to pave the way for isobutanol production through CBP. Hence, further research efforts will be required to achieve this goal. The engineered strains of *P. thermoglucosidasius* NCIMB 11955 capable of either producing bioethanol, 3-HP or isobutanol could form the basis of a low-cost biomass processing and the production of biobased chemicals or fuels.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF ABBREVIATIONS

%	Percentage
˚C	Degree Celsius
2SPY	Soy Peptone Yeast Extract
2SPYNG	Soy Peptone Yeast Extract No Glycerol
3-HP	3-Hydroxypropionic acid
3-HP/4-HB	3-hydroxypropionate/4-hydroxybutyrate cycle
3-HPA	3-Hydroxypropionaldehyde
5-FOA	5-fluoroorotic acid
5-FOA^R	5-fluoroorotic acid resistance
5-FOMP	5-fluoroorotidine monophosphate
5-FUMP	5-fluorouridine monophosphate
aad9	'encoding spectinomycin 9-adenyltransferase'
ACC	Acetyl-CoA carboxylase
ACE	Allele coupled exchange
Ack	Acetate Kinase
AdhE	'encoding Aldehyde-alcohol Dehydrogenase'
ADHs	Alcohol dehydrogenases
AE	Allelic Exchange
AFEX	Ammonia Fiber expansion
ALS	Acetolactate synthase
ANOVA	Analysis of variance
ApE	A plasmid Editor
ATP	Adenosine triphosphate
Bcf	Billion cubic feet
BLAST	Basic Local Alignment Search tool
bp	Base pairs
BT	Billion tons
C5	Pentose sugar
C6	Hexose sugar
Cas	CRISPR-associated proteins
CBM	*Clostridium* basal media
CBP	Consolidated bioprocessing
CCR	Carbon catabolite repression
CFU	Colony forming units
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
DME	Dimethyl ether
DMSO	Dimethyl sulfoxide
DSB	Double-stranded break
eCGP123	Enhanced Consensus Green Protein Variant 123
EDTA	Ethylenediaminetetraacetic acid

EPB	Electroporation buffer
FOA	Fluoroorotic acid
FTL	Fischer-Tropsch liquids
g	Grams
g	Relative centrifuge force
$g \cdot g^{-1}$	Gram per gram
$g \cdot l^{-1}$	Grams per litre
Gapd	Glyceraldehyde 3-phosphate Dehydrogenase
GHGs	Greenhouse gases
GHs	Glycoside Hydrolases
h	Hour(s)
hbdH	'encoding 3-hydroxyisobutyrate dehydrogenase'
HCL	Hydrochloric acid
HDR	Homology directed repair
hpdH	'encoding 3-HP dehydrogenase'
HPLC	High-performance liquid chromatography
IDT	Integrated DNA Technologies
IEA	International Energy Agency
ilvC	'encoding Ketoacid reductoisomerase'
ilvD	'encoding Dihydroxy acid dehydratase'
Indels	Insertions and deletions
Kan	Kanamycin Adenyltransferase
Kan^R	Kanamycin Resistance
kb	Kilo base
KDC	2-keto acid decarboxylases
KI	Knock-in
KIVD	Ketoisovalerate decarboxylase
KO	Knock-out
KOH	Potassium hydroxide
l	Litre
LB	Luria-Bertani
ldhA	'encoding Lactate Dehydrogenase'
LHA	Left homology arm
M	Moles per litre
Mb d^{-1}	Million barrels per day
Mb	Million barrels
MCR	Malonyl-CoA reductase
MCS	Multiple cloning site
mg	Milligram
min	minute(s)
MJ kg^{-1}	Megajoules per kilogram
MJ l^{-1}	Megajoules per litre
ml	Millitre
mM	Milli moles per litre

mmsA	'encoding (methyl) malonate semialdehyde'
mRNA	Messenger RNA
MSR	Malonate semialdehyde reductase
MT	Million tons
MW	Molecular weight
NAD	Nicotinamide Adenine Dinucleotide
NaOH	Sodium hydroxide
NCBI	National Centre for Biotechnology Information
ng	Nano grams
NHEJ	Non-homologous end-joining
OAA	Oxaloacetate
OD_{600}	Optical density at 600 nm
OMP	Orotidine-5-monophosphate
PAM	Protospacer Adjacent Motif
PCR	Polymerase chain reaction
pdhA	'encoding Pyruvate Dehydrogenase'
pdh^{up}	Pyruvate Dehydrogenase Upregulated
PDO	1,3-propanediol
PEP	Phosphoenolpyruvate
pflB	'encoding Pyruvate Formate Lyase'
PKC	Palm Kernel Cake
Pldh	lactate dehydrogenase promoter
Psi	pound-force per square inch
pyrE	'encoding Orotate Phosphoribosyltransferase'
RBS	Ribosomal binding site
Rbx	Riboswitch
RHA	Right homology arm
RI	Refractive index
rpm	Revolutions per minute
SBRC	Synthetic biology research centre
SD	Standard deviation of mean
s	second
sfGFP	Superfolder green fluorescent protein
sgRNA	Single Guide RNA
SHF	Separate hydrolysis and fermentation
SNP	Single nucleotide polymorphism
SNV	Single Nucleotide Variation
SOC	Super optimal catabolite repression medium
SOEing	Splicing by Overlap Extension
SScF	Simultaneous saccharification and co-fermentation
SSF	Simultaneous saccharification and fermentation
Tb	Trillion barrels
TCA	Tricarboxylic Acid

Tcf	Trillion cubic feet
tracrRNA	Trans-activating CRISPR RNA
tRNA	Transfer RNA
TSA	Tryptone soya agar
U	Units
UMP	Uridine-5-monophosphate
UNFCCC	United Nations Framework Convention on Climate Change
UV	Ultra-violet
V	Volt
v/v	Volume per volume
w/v	Weight per volume
WT	Wild-type
µF	Micro Faraday
µg	Micro gram
µl	Micro litre
µM	Micro moles per litre
Ω	Ohms

CHAPTER 1

1.0 GENERAL INTRODUCTION

1.1 Why the Search for Alternative Fuels

1.1.1 Energy demand

The global demand for energy to meet basic human social and economic needs is increasing rapidly due to growing global populations and industrialisation (Edenhofer *et al.,* 2011). The world relies on power to meet fundamental human needs such as cooking, heating, communication, and transportation, of which the latter is the largest consumer of oil (Edenhofer *et al.,* 2011). According to the International Energy Agency (IEA, 2021), global oil demand is now forecast to rise by 5.5 million barrels per day (Mb d^{-1}) in 2021 and 3.3 Mb d^{-1} in 2022 when it reaches 99.6 Mb d^{-1}, slightly above pre-COVID levels. Hence, meeting global energy demands is a challenge and will continue to be of concern. The energy demand, therefore, underscores the need to continue developing renewable and alternative energy sources to meet the ever-increasing global demand.

1.1.2 Environmental challenges

In the form of coal, oil, and gas, fossil fuels are associated with releasing greenhouse gases (GHGs) into the atmosphere, particularly carbon dioxide (CO_2). GHGs cause global warming, which is of primary environmental concern due to significant changes in the Earth's climate (Asumadu-Sarkodie and Owusu, 2016).

Global energy-related CO_2 emissions from energy combustion and industrial processes increased by 6% in 2021 to their highest ever level of 36.3 gigatonnes. The emission was in part a consequence of the strong recovery of the world's economy from the COVID-19 pandemic, which relied heavily on coal to power that growth (IEA, 2022). Therefore, "*The world must now ensure that the global rebound in emissions in 2021 was a one-off – and that sustainable investments combined with the accelerated*

deployment of clean energy technologies will reduce CO_2 emissions in 2022, keeping alive the possibility of reducing global CO_2 emissions to net-zero by 2050" (IEA, 2022). Therefore, alternative fuels that are carbon neutral and have no net addition of GHG into the atmosphere are needed for environmental and economic sustainability.

1.1.3 Non-renewable resources

Fossil fuels represent about 84.3% of the global energy, with oil, gas, and coal respectively accounting for 33.1%, 24.3%, and 27% of the worldwide energy mix, while 15.7% comes from low-carbon sources (Ritchie and Roser, 2020). Low-carbon sources are the sum of renewable energy sources – such as wind, solar, bioenergy, hydropower, geothermal, wave and tidal (11.4%), and nuclear energy (4.3%) (Ritchie and Roser, 2020). Although more and more energy is produced from renewables every year, the world energy mix is still dominated by oil, gas, and coal and will continue to do so for the foreseeable future. Consequently, this reduces fossil fuel pools, which take millions of years to form. Thus, as fossil fuels are non-renewable, this will create unsustainable circumstances that will, in the end, bring about a possibly irreversible danger to human society (UNFCCC, 2015).

Worldwide, oil, gas, and coal stores are being expanded, yet fossil fuels are limited assets (Abas *et al.,* 2015). Table 1.1 reveals the global hydrocarbon pools, production rates, and exhaustion dates for gas, coal, and oil. Based on the assumption of no future discoveries, the exhaustion dates are calculated based on the current reserves and production rates (Abas *et al.,* 2015).

Table 1.1: World fossil fuel statistics based on confirmed reserves

Fuels	Total reserves	Production per day	Depletion date
Oil	1.689 Tb	86.81 Mb	2066
Gas	6558 Tcf	326 Bcf	2068
Coal	891.531 BT	21.63 MT	2126

Tb=Trillion barrels; Tcf=Trillion Cubic Feet; BT=Billion Tons; Mb=Million barrels; Bcf=Billion Cubic Feet; MT=Million Tons. Adapted from Abas *et al.* (2015).

1.1.4 Energy security

Continual fuel price fluctuations in global markets, geopolitical and military conflicts in oil-producing regions, and the rapid depletion of fossil fuels and unequal distribution of reserves have raised concerns about energy security.

Energy security concerns depend on a constant energy supply, which is essential for running an economy (Kruyt *et al.*, 2009). The relationship between financial development and energy utilisation requires a steady energy supply significant to the political world and financial concern for the developing and developed nations. This is because obstructions would lead to substantial economic challenges (Edenhofer *et al.*, 2011).

Sustainable energy sources such as biofuels could be produced worldwide compared to fossil fuels which are confined to specific regions. These energy sources will reduce the importation of fossil fuels, add to the diversity of energy supply options, lessen a country's vulnerability to price instability, and boost global energy security. A wide range of energy sources, such as biofuels, can help improve energy security (Edenhofer *et al.*, 2011).

1.2 Biofuels

Biofuels, syngas (synthesis gas), renewable natural gas, and hydrogen may represent the most significant renewable energy sources within the foreseeable future. Among these four energy alternatives, biofuels are the most environmental-friendly energy source. The quest for biofuels increases tremendously due to environmental issues associated with fossil fuel usage and energy security (Dennis *et al.*, 2008). Consequently, research into biofuels is being vigorously pursued as they are considered alternatives to conventional energy sources due to advantages such as the generation of acceptable quality exhaust gases and renewability (Bhatti *et al.*, 2008).

Biofuels refer to fuels produced primarily from biomass by living organisms. These fuels could be in the form of gas, liquid or solid. Examples include bioethanol, biodiesel, biomethanol, biohydrogen and methane (Demirbas, 2008). Biofuels are generally grouped into primary and secondary biofuels. Primary biofuels refer to fuels in which the biomass is used in its standard and non-altered structure, for example wood pellets and firewood, whereas secondary biofuels (The state of food and agriculture, 2008) refers to processed fuels in the form of liquids (such as bioethanol or biodiesel), gases (for example, hydrogen, syngas and biogas), and solids (for example, charcoal). Based on the feedstock used in their production, secondary biofuels are classified into first and second-generation biofuels (Nigam and Singh, 2011) and 'advanced biofuels' (Love, 2022).

1.2.1 Classification of liquid biofuels

1.2.1.1 First-generation liquid biofuels

First-generation liquid biofuels refer to liquid fuels that require a moderately simple production procedure. First-generation biofuels are produced from carbohydrate or oil-rich food crops (Nigam and Singh, 2011). Ethanol is the most notable first-generation biofuel produced by sugar fermentation obtained from crops and starch contained in maize kernels, sugar cane and other starchy crops (Larson, 2008). Bioethanol is typically produced using yeast through the fermentation of sugars. Yeast ferment hexose sugars, mainly glucose, into ethanol (Nigam and Singh, 2011). Biodiesel is another outstanding first-generation biofuel produced from oleaginous plants by transesterification processes or cracking. Biodiesel and glycerine as by-product are produced via transesterification using acid, alkaline or enzymatic catalysers and ethanol or methanol (Escobar *et al.,* 2009).

First-generation biofuels are already produced in large-scale in several countries such as Brazil, India, USA, Germany, China, Italy, France, Austria (Escobar *et al.,* 2009). However, production of this type of biofuel is hindered by the feedstock competition with food (Patil *et al.,* 2008), thereby making the cost of production high. Thus, this drawback necessitates the quest for non-food feedstock to produce biofuels (Nigam and Singh, 2011).

1.2.1.2 Second-generation liquid biofuels

Second-generation liquid biofuels are produced primarily from lignocellulosic biomass such as non-food residues of crops or non-edible whole plant biomass. These biofuels can limit the direct competition of food with fuel related to first-generation biofuels as they are produced from non-edible feedstocks (Nigam and Singh, 2011). The feedstock

holds the potential for low production costs, great energy, and advantages to the environment (Larson, 2008). However, second-generation biofuel production involves complex equipment, more significant expenditure per production unit, and numerous facilities to restrict and diminish capital cost scale economies (Stevens *et al.,* 2004). More studies, improvements, and applications of feedstock conversion technologies are needed to accomplish second-generation biofuels' energy and economic outcomes (Nigam and Singh, 2011).

Second-generation biofuels are further divided based on the procedure applied to produce them from lignocellulosic biomass, whether thermochemical or biochemical. Ethanol and butanol as second-generation biofuels are produced via biochemical means. In contrast, refined Fischer-Tropsch liquids (FTL), methanol, pyrolysis oils and dimethyl ether (DME) are made via the thermochemical process (Larson, 2008).

1.2.1.3 Advanced biofuels

Advanced biofuels, sometimes referred to as third- or fourth-generation biofuels, encompass a range of alternative combustible molecules or biomass that are typically derived from microbes (Love, 2022). On the one hand, the term third-generation is often used to describe both the biomass and the fuel precursors derived from microalgae and oleaginous microbes. On the other hand, fourth-generation refers to biofuels from genetically engineered microbial cell factories such as microalgae, bacteria and fungi (Love, 2022). Notwithstanding the fact that microbes produce bulk of the first-generation biofuel used globally and the the enormous research into advanced biofuels, scale up from laboratory demonstration to industrial production and commercial distribution is a significant challenge.

1.2.2 Benefits and challenges of biofuels

Production and utilisation of biofuels are associated with several advantages concerning the environment, energy security, economics, and several challenges that need to be addressed to achieve the desired outcome (Hoekman, 2009). Nigam and Singh (2011) highlighted some potential benefits and challenges of biofuels production and utilisation (Table 1.2).

Table 1.2: Potential benefits and challenges of biofuels

Benefits	Challenges
Energy Security	**Feedstock**
Domestic energy source	Collection network
Locally distributed	Storage facilities
Well connected supply-demand chain	Food-fuel competition
Higher reliability	
Economic stability	**Technology**
Price stability	Pre-treatment
Employment generation	Enzyme production
Rural development	Efficiency improvement
Reduce inter-fuels competition	Technology cost
	Production of value-added co-products
Reduce demand-supply gap	
Open new industrial dimensions	
Land-use change	
Control on the monopoly of fossil-rich states	**Policy**
	Fund for research and development
Environmental gains	Pilot-scale demonstration
Better waste utilisation	Commercial-scale deployment
Reduce local pollution	Policy for biofuels
	Procurement of subsidies on production
Reduce GHGs emissions from energy	
	Tax credits on utilisation and production-consumption of biofuels
Reduction in landfill sites	

Adapted from Nigam and Singh (2011)

1.3 Biobutanol as an Alternative to Bioethanol

Butanol is a four-carbon alcohol (C_4H_9OH) that has found application in the cosmetics and pharmaceutical industries as a solvent. It is also called biobutanol when produced from the fermentation of sugar feedstock through biological means (Durre, 2007). Furthermore, butanol occurs in four isomers, namely *n*-butanol (normal-butanol), 2-butanol (secondary-butanol), i-butanol (iso-butanol) and t-butanol (ter-butanol), all of which contain a similar amount of energy (Grana *et al.*, 2010). However, only *n*-butanol and isobutanol are produced via biological processes (Ramey, 2004).

As a possible replacement fuel for bioethanol, biobutanol (*n*-butanol and isobutanol) have gained increasing attention due to several advantages above ethanol. These include high energy content (29 $MJ.l^{-1}$), low volatility (0.53-1.17 kPa), and less hygroscopicity (Durre, 2007; Lee *et al.*, 2008; Nigam and Singh, 2011). Butanol also has high number of carbon atoms (2x) compared to ethanol, as a result, more energy is present in butanol than in ethanol (approximately 30% excess energy) (Durre, 2007; Wackett, 2008). Furthermore, butanol takes longer to combust in the motor engine than ethanol because of its high boiling point compared to ethanol. Additionally, it is less corrosive and more appropriate for distribution through existing petroleum pipelines. Butanol has low Reid vapour pressure than ethanol (7.5 times lower), hence less evaporative and volatile (Morone and Pandey, 2014). In addition, butanol can be blended in higher proportions with petroleum than ethanol for use in existing cars without moderation because the air-fuel ratio and energy content is nearer to oil (Sarathy *et al.*, 2012; Campos-Fernandez *et al.*, 2012). In contrast to different alcohols, the Environmental Energy Company (US) affirmed that butanol could be used to replace petroleum with no alterations to vehicle engines (Brekke, 2007). The comparison of the properties of gasoline, isobutanol, *n*-

butanol, and ethanol is presented in Table 1.3. The production of isobutanol via amino acid biosynthetic pathway is discussed in Chapter 6.

Table 1.3: Properties of gasoline, isobutanol, *n*-butanol and ethanol

Properties	Gasoline	Isobutanol	*n*-butanol	Ethanol
Energy density (MJ/L)	32	29	29	19.6
Air-fuel ratio	14.6	11.2	11.2	9.0
Vapour pressure (kPa) at 20 °C	0.7-207	1.17	0.53	7.58
Vapour pressure of mixture with gasoline (kPa)	53.8-103.4	46.9	44.1	138
Heat of vaporisation (MJ/Kg)	0.36	0.57	0.43	0.92
Research octane number	91-99	113	96	129
Motor octane number	81-89	94	78	102
Boiling point at 1 atm (°C)	210-235	108	118	78
Freezing temperature (°C)	<-60	-180	-89.5	-114.5
Hygroscopicity	Low	Low	Low	High
Compatibility with existing infrastructure	Yes	Yes	Yes	No

(Al-Hasan and Al-Momany, 2008; Hewitt and Lokare, 2016; Aziz and Gozan, 2017)

1.4 3-hydroxypropionic Acid

According to the United States Department of Energy list issued in 2004 and revised in 2010, the platform chemical, 3-hydroxy propionic acid (3-HP; MW 90.08; $C_3H_6O_3$) is recognised as one of the top 12 value-added platform chemicals that can be derived from renewable biomass (Werpy and Petersen, 2004; Matsakas *et al.*, 2018). 3-HP serves as one of the critical building block chemicals. However, the production of platform chemicals from fossil resources is associated with environmental problems, sustainability, and limited fossil fuel resources. For these reasons, the development of

alternative production methods from renewable resources has gained increased attention to replace fossil resources (Matsakas *et al.,* 2018).

Chemically, 3-HP is a three-carbon, non-chiral organic compound and a structural isomer of lactic acid (2-hydroxypropionic acid). 3-HP has two functional groups, a carboxyl group and a hydroxyl group at the β-position, making it a versatile platform for producing value-added chemicals (Figure 1.1). 3-HP is a promising platform chemical that can be used to create numerous compounds with industrial applications, such as acrylamide, acrylonitrile, acrylic acid, 1,3-propanediol, propiolactone, malonic acid, and ethyl 3-HP. These high value-added chemicals are used in various applications, such as making adhesives, polymers, plastic packaging, cosmetics, antistatic agents for textiles, a cross-linking agent for metal lubricants and polymer coatings (Kumar *et al.*, 2013).

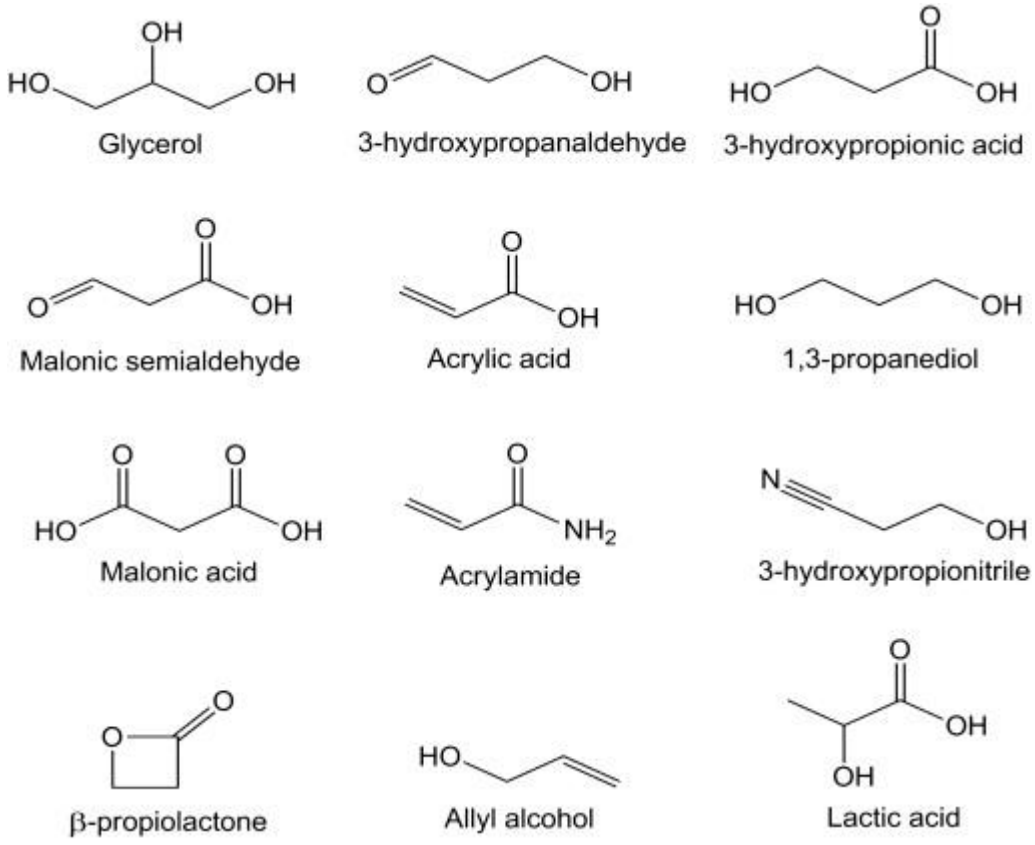


Figure 1.1: Chemical structures of 3-HP and some precursors and intermediates related to biological and chemical production of 3-HP. Adapted from Kumar *et al.* (2013).

3-HP can be produced by the petrochemical industry via several chemical synthetic routes (Jiang *et al.,* 2009). These routes include (1) hydrolysis of highly reactive β-propiolactone, (2) hydrolysis of 3-hydroxypropionitrile, derived from a chemical reaction between 2-chloroethanol and sodium cyanide, (3) catalysis of allyl alcohol as a precursor with a gold catalyst and (4) oxidation of 1-3-propanediol via a palladium-containing support catalyst (Kumar *et al.,* 2013). Despite four chemical synthetic routes have been suggested, none are commercially feasible at present due to the high cost of the raw materials and environmental and process incompatibilities (Jiang *et al.,* 2009;

Kumar *et al.*, 2013). Thus, the bioconversion of biomass into chemical building blocks or chemical intermediates could serve as an alternative to the petrochemical industry.

1.4.1 Native microbial metabolism involving 3-HP

Many microorganisms can produce 3-HP as either an intermediate or end product through a range of metabolic pathways (Kumar *et al.*, 2013). The natural producers include both prokaryotes and eukaryotes.

Lactobacillus sp. has been reported to co-produce 3-HP and 1-3-propanediol (PDO) from glycerol (Sobolov and Smiley, 1960; Talarico *et al.*, 1988; Garai-Ibabe *et al.*, 2008). Garai-Ibabe *et al.* (2008) isolated 67 lactic acid bacteria isolates from bitter-tasting ciders and found that 22 strains could convert glycerol to 3-HP and PDO, with *Lactobacillus collinoides* 17 producing the highest 3-HP (31.3 mM) and 37.1 mM PDO in 64 h. The co-production of 3-HP and PDO is through a two-step reaction (Figure 1.2). The first step converts glycerol to 3-hydroxypropionaldehyde (3-HPA) and is catalysed by glycerol dehydratase. Then, 3-HPA can be subsequently reduced to PDO or oxidised to 3-HP by an aldehydic dismutation of 3-HPA (Garai-Ibabe *et al.*, 2008), which occurred mainly in the absence of glucose when glycerol fermentation proceeded. In addition, 3-HP production was reduced, and PDO production was elevated in the presence of higher fructose concentrations or glucose (Kumar *et al.*, 2013).

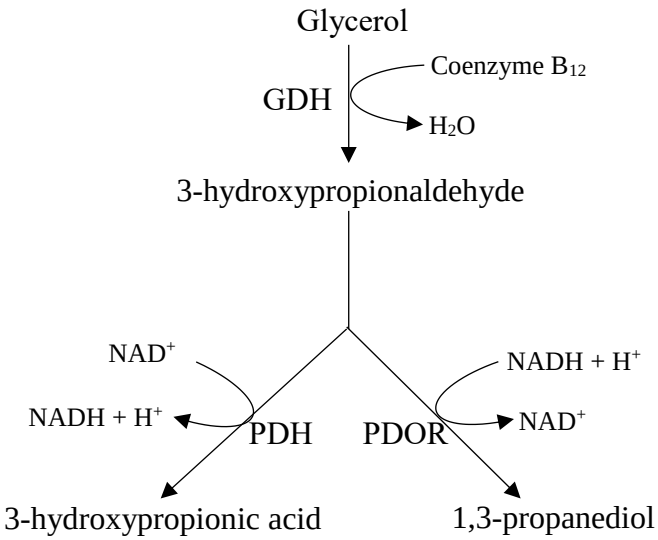


Figure 1.2: Pathway for co-production of 3-hydroxypropionic acid and 1,3-propanediol in *L. collioides*. Enzymes involved: GDH, B12-dependent glycerol dehydratase; PDH, Propionaldehyde dehydrogenase; PDOR, 1,3-propanediol oxidoreductase. Modified from Garai-Ibabe *et al.* (2008).

3-HP is a crucial intermediate of the 3-hydroxypropionate and 3-hydroxypropionate/4-hydroxybutyrate cycle pathways for autotrophic carbon dioxide fixation. *Chloroflexus aurantiacus,* a thermophilic photosynthetic bacterium, has been reported to secrete 3-HP during phototrophic growth and small amounts during autotrophic growth (Holo, 1989). The 3-HP was reported as an intermediate in the 3-HP cycle (Figure 1.3) in this microorganism (Ishii *et al.,* 2004) and in archaea such as *Sulfolobus* sp. strain VE6, *Sulfolobus metallicus*, *Acidianus ambivalens*, and *Acidianus brierleyi* (Ishii *et al.,* 2004).

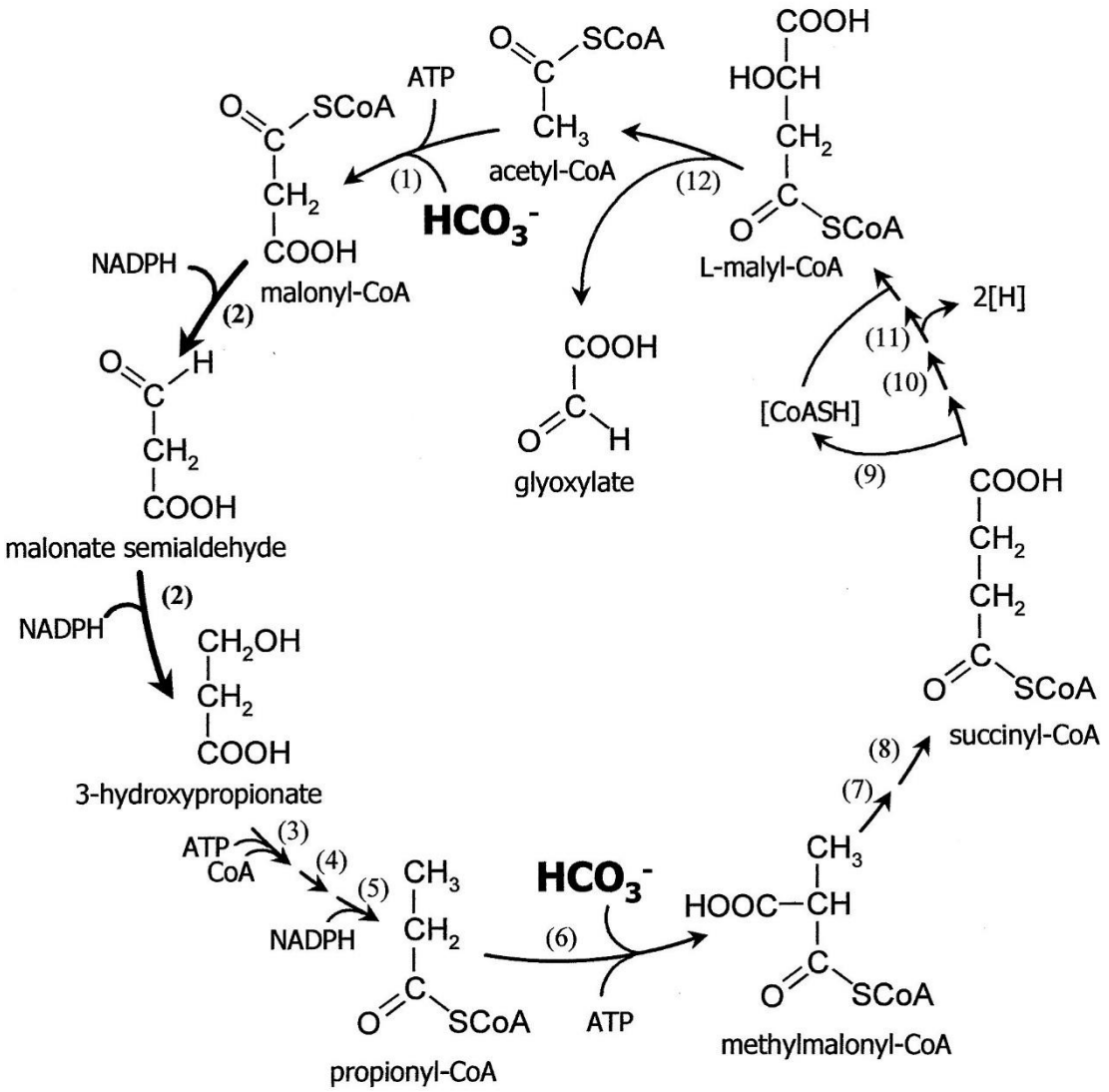


Figure 1.3: Production of 3-HP as an intermediate in 3-hydroxypropionate cycle for autotrophic CO_2 fixation in *C. aurantiacus*. Enzyme involved: 1, acetyl-CoA carboxylase; 2, malonyl-CoA reductase (NADPH); 3, 3-hydroxypropionyl-CoA synthetase; 4, 3-hydroxypropionyl-CoA dehydratase; 5, acryloyl-CoA reductase (NADPH); 6, propionyl-CoA carboxylase; 7, methylmalonyl-CoA epimerase; 8, methylmalonyl-CoA mutase; 9, succinyl-CoA:L-malate CoA transferase; 10, succinate dehydrogenase; 11, fumarate hydratase; 12, L-malyl-CoA lyase. Adapted from Hügler *et al.* (2002).

A new pathway for carbon dioxide assimilation known as the 3-hydroxypropionate/4-hydroxybutyrate cycle (3-HP/4-HB) (Figure 1.4) was also discovered in 2007 in *Metallosphaera sedula* (Berg *et al.*, 2007). In both pathways, acetyl-CoA and propionyl-CoA carboxylases are the enzymes that catalyse CO_2 fixation within the cycle. The pathways begin with the carboxylation of acetyl-CoA to malonyl-CoA by acetyl-CoA carboxylase. The malonyl-CoA is then reduced to 3-HP via malonate semialdehyde. The 3-HP is reduced to propionyl-CoA, then carboxylated to methylmalonyl-CoA, and finally isomerised to succinyl-CoA. After which, the two pathways adopt different routes; succinate and succinate semialdehyde (Berg *et al.*, 2007).

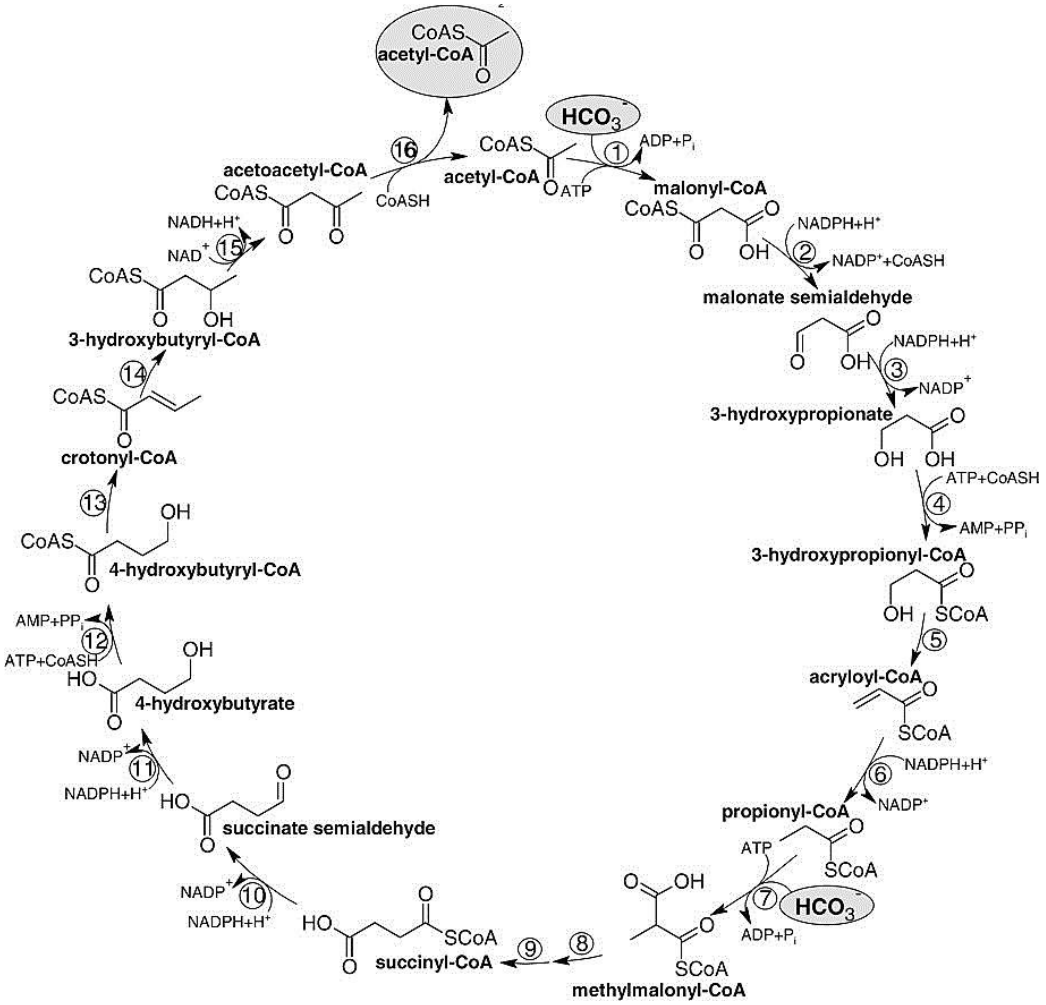


Figure 1.4: Production of 3-HP as an intermediate in 3-hydroxypropionate/4-hydroxybutyrate cycle in *M. sedula*. Enzyme involved: 1, acetyl-CoA carboxylase; 2, malonyl-CoA reductase; 3, malonate semialdehyde reductase; 4, 3-hydroxypropionyl-CoA synthetase; 5, 3-hydroxypropionyl-CoA dehydratase; 6, acryloyl-CoA reductase; 7, propionyl-CoA carboxylase; 8, methylmalonyl-CoA epimerase; 9, methylmalonyl-CoA mutase; 10, succinyl-CoA reductase; 11, succinate semialdehyde reductase; 12, 4-hydroxybutyryl-CoA synthetase; 13, 4-hydroxybutyryl-CoA dehydratase; 14, crotonyl-CoA hydratase; 15, 3-hydroxybutyryl-CoA dehydrogenase; 16, acetoacetyl-CoA β-ketothiolase. Adapted from Berg *et al.* (2007).

Fungi such as *Geotrichum*, *Trichoderma* sp. (Dave *et al.*, 1996), and *Byssochlamys* (Takamizawa *et al.*, 1993) have been reported to be able to produce 3-HP by metabolising acrylic acid. Furthermore, endophytic fungi such as *Phomopsis phaseoli* and *Melanconium betulinum* have also been reported to secrete 3-HP as a nematicide (Schwarz *et al.*, 2004). Despite the native production of 3-HP by microorganisms either as an intermediate of the 3-HP and 3-HP/4-HB cycle or from acrylic acid, the productivity is too low to be exploited on an industrial scale.

1.4.2 Recombinant production of 3-HP

To achieve large industrial scale biological production of 3-HP, the use of abundant and cheap substrates and the adoption of efficient production systems that result in a higher titer, yield and productivity are needed. Thus, the expression of pathways in genetically engineered hosts to produce 3-HP is the approach most widely investigated to achieve commercialisation (Kumar *et al.*, 2013). Production of 3-HP in this way has been implemented in different hosts such as *Escherichia coli, Corynebacterium glutamicum, Klebsiella pneumoniae, Bacillus subtilis, Pseudomonas denitrificans, Saccharomyces cerevisiae, Lactobacillus reuteri, Synechocystis sp., Schizosaccharomyces pombe*, and *Synechocystis elongates* using either glucose, glycerol or CO_2 as a substrate (Song *et al.*, 2016; Chen *et al.*, 2017; Ko *et al.*, 2017; Kalantari *et al.*, 2017; Zhou *et al.*, 2013; Kildegaard *et al.*, 2016; Zaushitsyna *et al.*, 2017; Wang *et al.*, 2016; Suyama *et al.*, 2017; Lan *et al.*, 2015).

1.4.2.1 Recombinant pathways for 3-HP production from glucose

Seven biosynthetic pathways via which 3-HP could be produced from glucose at 100% theoretical yield have been proposed and patented by Cargill, a US-based agricultural company (Jiang *et al.*, 2009; Kumar *et al.*, 2013). Ideally, the pathway should be redox-

balanced and generate ATP during 3-HP production to maintain cell growth. Otherwise, carbon substrates would be consumed to maintain the redox balance and ATP generation, thereby reducing product yield and increasing unwanted by-products (Straathof *et al.,* 2005). All the seven synthetic pathways are redox-balanced, but most are not ATP-producing, and some consume ATP (Kumar *et al.,* 2013). Pyruvate or phosphoenolpyruvate (PEP) are the key intermediates in all of the pathways for 3-HP production. Thus, two ATP and two NADH moles are produced when one mole of glucose is oxidised to two moles of pyruvate via glycolysis. The ATP yield for a pathway can vary and depends on the enzyme (s) involved in the pathway.

Pathway I (Lactate pathway): This pathway involves the production of 3-HP from glucose via lactate as an intermediate (Figure 1.5). One ATP molecule per 3-HP molecule is produced if the conversion of lactate to lactoyl-CoA is mediated by CoA-transferase, which is otherwise an ATP-consuming reaction. This pathway is redox balanced (Jiang *et al.,* 2009) but thermodynamically unfavourable (Kumar *et al.,* 2013), because 3-HP is an isomer of lactate, and the K_{eq} for the equilibrium between lactate and 3-HP is estimated to be 0.4 (Herrmann *et al.,* 2005), unless 3-HP is exported selectively while lactate is locked and accumulates inside the cell. Unfortunately, such mechanisms of 3-HP export have not been reported (Jiang *et al.,* 2009). The separation of lactate and 3-HP becomes very difficult, if lactate is secreted into the fermentation broth.

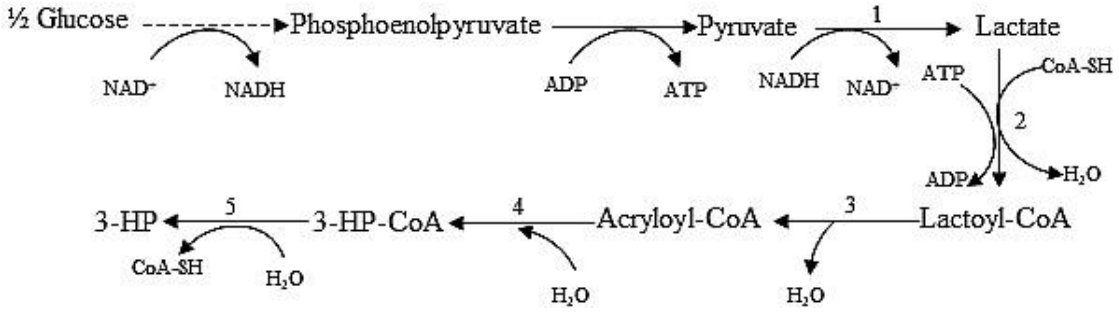


Figure 1.5: Lactate pathway for 3-HP production from glucose. 1: Lactate dehydrogenase; 2: acetate-CoA ligase/CoA transferase; 3: Lactoyl-CoA dehydratase; 4: 3-HP-CoA dehydratase; 5: 3-HP-CoA hydrolase/3-hydroxyisobutryl-CoA hydrolase. Modified from Kumar *et al.* (2013).

Pathway II (Malonyl-CoA pathway): In this pathway, 3-HP is produced via malonyl-CoA as an intermediate (Figure 1.6). The conversion of glucose into acetyl-CoA generates 2 NADH, 1 CO_2 and 1 ATP per acetyl-CoA produced. The conversion of acetyl-CoA to malonyl-CoA by acetyl-CoA carboxylase consumes 1 ATP, while 2 NADPH per 3-HP is consumed to reduce malonyl-CoA to 3-HP (Hügler *et al.*, 2002). Consequently, the pathway is redox neutral with no net ATP and is thermodynamically favourable (Jiang *et al.*, 2009). Besides the lack of net ATP generation, another drawback is the inability of malonyl-CoA reductase to utilise NADH as a cofactor. A possible solution is introducing of NAD(P) transhydrogenase into the host to maintain a balance of NADH and NADPH (Rathnasingh *et al.*, 2012).

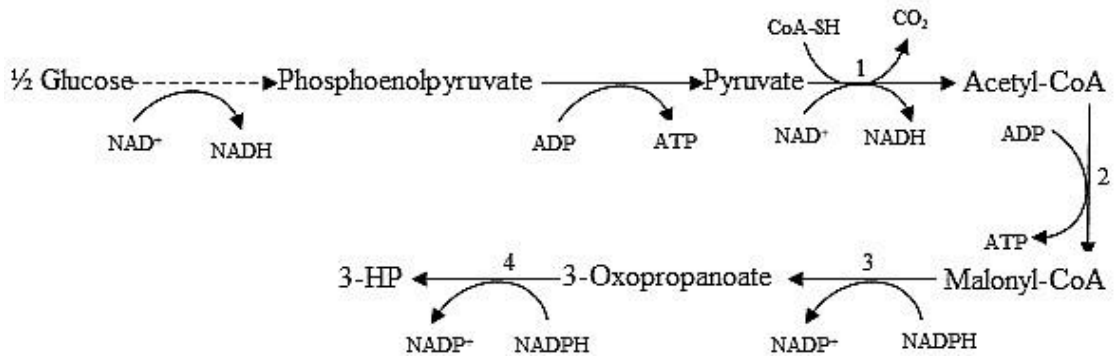


Figure 1.6: Malonyl-CoA pathway for 3-HP production from glucose. 1: Pyruvate dehydrogenase; 2: acetyl-CoA carboxylase; 3: Malonyl-CoA reductase; 4: Malonate semialdehyde reductase. Modified from Kumar *et al.* (2013).

Pathways III and IV (β-alanine pathway): 3-HP is produced from glucose via β-alanine as an intermediate in these pathways (Figure 1.7). The first three steps involving the conversion of pyruvate/ PEP to oxaloacetate (OAA) are common in both pathways. The carboxylation of PEP/pyruvate to oxaloacetate generates no ATP if catalysed by either PEP carboxylase or pyruvate carboxylase. One NADH per oxaloacetate is produced when glucose is converted to oxaloacetate. The 3-hydroxyisobutyrate dehydrogenase or glutamate dehydrogenase that catalyses the conversion of 3-oxopropanoate to 3-HP or oxaloacetate to aspartate, respectively, consumes 1 NADH. Thus, the pathways are redox neutral with no net ATP (Jiang *et al.,* 2009). The ATP yield can vary depending on the intermediate (pyruvate/PEP) and/or enzyme used (PEP carboxylase/carboxykinase).

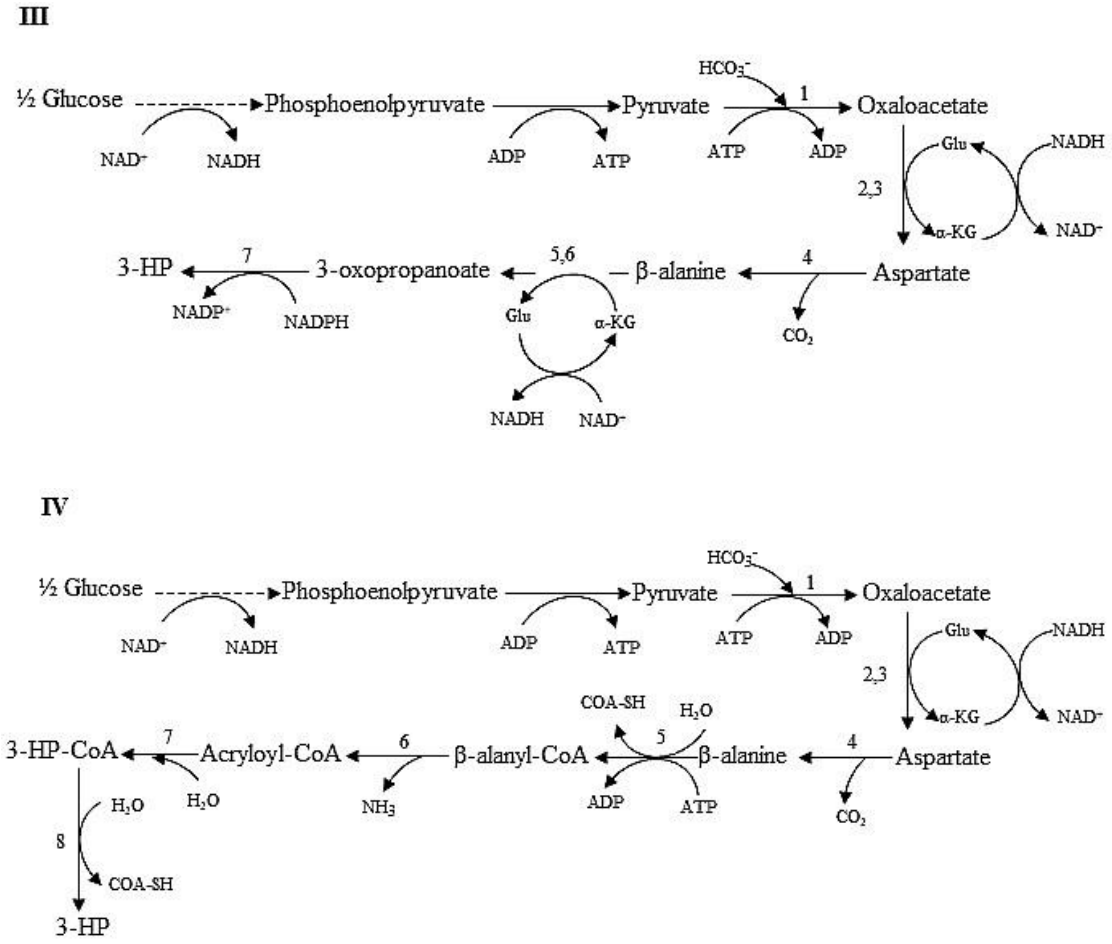


Figure 1.7: β-alanine pathways for 3-HP production from glucose. III) 1: Pyruvate carboxylase/phosphoenolpyruvate carboxylase; 2: Aspartate aminotransferase; 3: glutamate dehydrogenase; 4: Aspartate decarboxylase; 5: 4-aminobutyrate aminotransferase/β-alanine-2-oxoglutarate aminotransferase; 6: glutamate dehydrogenase; 7: 3-hydroxyisobutyrate dehydrogenase. **IV)** 1: Pyruvate carboxylase/phosphoenolpyruvate carboxylase; 2: Aspartate aminotransferase; 3: glutamate dehydrogenase; 4: Aspartate decarboxylase; 5: acetate-CoA ligase/CoA transferase; 6: β-alanine-CoA ammonium lyase; 7: 3-HP-CoA dehydratase; 8: 3-HP-CoA hydrolase/CoA-transferase/3-hydroxyisobutryl-CoA hydrolase. Glu: glutamate; α-KG: α-ketoglutarate. Modified from Kumar *et al.* (2013).

Pathways V and VI: These pathways were proposed to overcome the lack of energy production in pathways III and IV. They bypass the ATP-consuming carboxylation steps by using alanine-2,3-aminomutase to convert α-alanine to β-alanine (Figure 1.8). Aminomutases are the enzymes that catalyse the transfer of amino group from α- to β-carbon. Hence, they are redox balanced with one net ATP per 3-HP (Jiang *et al.*, 2009). Because a naturally occurring alanine-2,3-aminomutase has not been reported a lysine-2,3-aminomutase has been engineered to convert α-alanine to β-alanine (Liao *et al.*, 2005). The ATP yield per 3-HP for pathway V is +1, while it can be 1 or 0 in case of pathway VI, depending on how the conversion of β-alanine to β-alanyl-CoA is carried out.

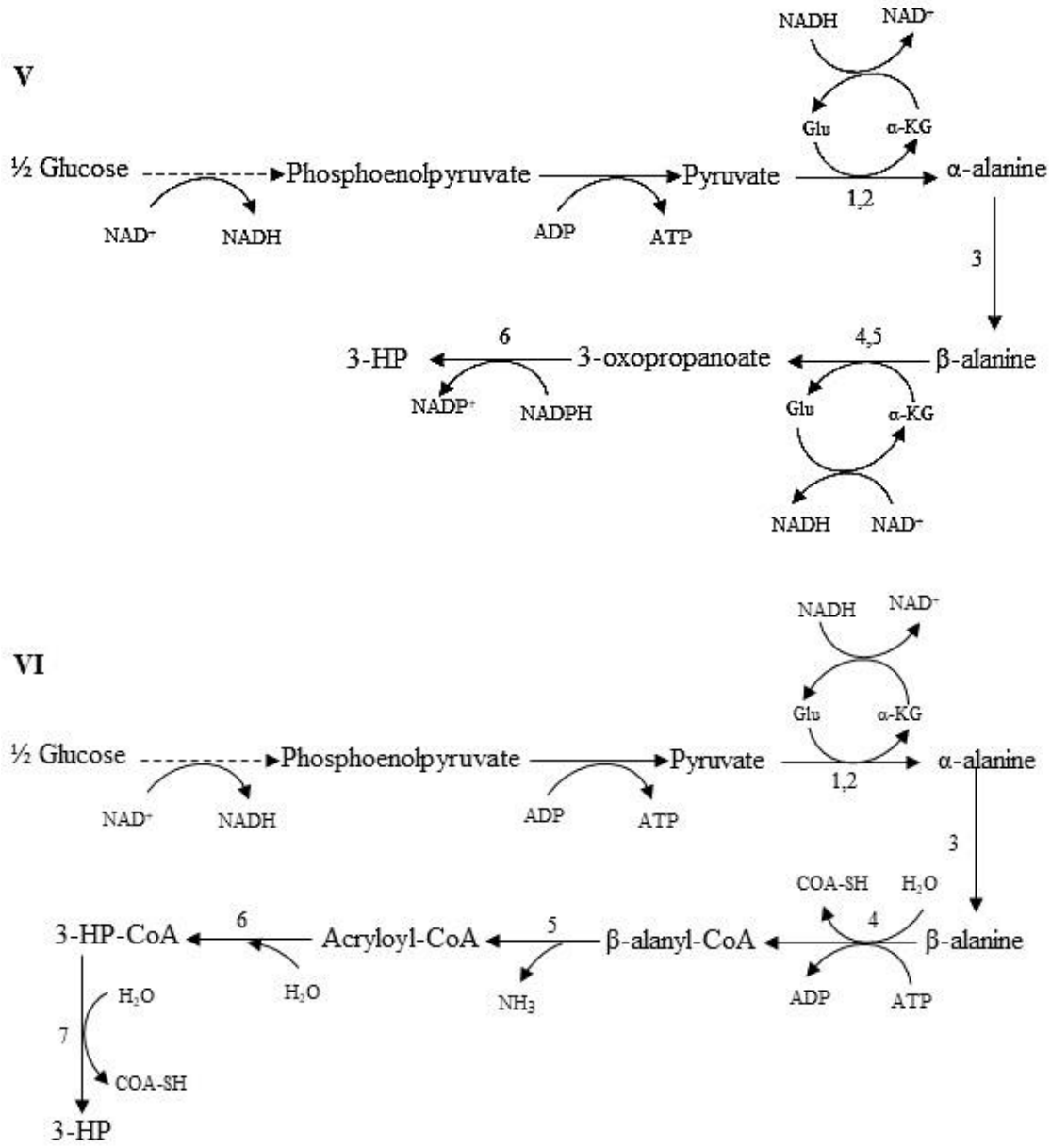


Figure 1.8: Pathways V and VII for 3-HP production via α-alanine and β-alanine intermediates. V) 1: Pyruvate-glutamate transaminase; 2: Alanine dehydrogenase; 3: Alanine-2,3-aminomutase; 4: 4-aminobutyrate aminotransferase/β-alanine-2-oxoglutarate aminotransferase; 5: glutamate dehydrogenase; 6: 3-hydroxyisobutyrate dehydrogenase. **VI)** 1: Pyruvate-glutamate transaminase; 2: Alanine dehydrogenase; 3: Alanine-2,3-aminomutase; 4: acetate-CoA ligase/CoA transferase; 5: β-alanine-CoA ammonium lyase; 6: 3-HP-CoA dehydratase; 7: 3-HP-CoA hydrolase/CoA-

transferase/3-hydroxyisobutryl-CoA hydrolase. Glu: glutamate; α-KG: α-ketoglutarate. Modified from Kumar *et al.* (2013).

Pathway VII: This pathway is based on the succinic acid fermentation pathway with propionate as an intermediate (Figure 1.9). The pathway was reported to exist in *Actinobacillus* species (Guettler *et al.,* 1999). Pyruvate or PEP is converted to succinate, which is then transformed to propionyl-CoA directly or via propionate. The dehydrogenation of propionyl-CoA removes two [H] and gives acryloyl-CoA, which is then hydrated to 3-hydroxypropionyl-CoA and followed by its hydrolysis to 3-HP. This pathway is redox neutral with a net ATP yield of -1, i.e., 1 ATP is consumed per 3-HP if propionyl-CoA is produced via propionate (Kumar *et al.*, 2013).

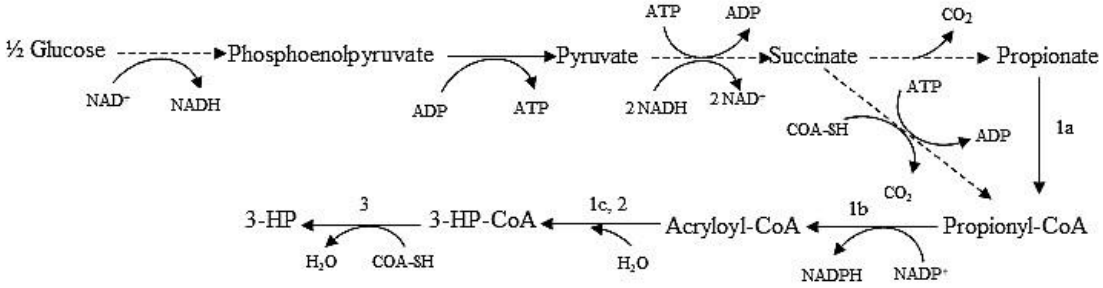


Figure 1.9: Pathway for 3-HP production via propionate/propionyl-CoA as intermediates. 1: OS17 enzyme (contains 3 functional domains: 1a: CoA-synthetase; 1b: dehydrogenase; 1c: 3-HP dehydratase); 2: 3-HP-CoA dehydratase; 3: 3-HP-CoA hydrolase/Co-transferase/3-hydroxyisobutryl-CoA hydrolase. Modified from Kumar *et al.* (2013).

For this study, pathway II, i.e., the malonyl-CoA pathway (Figure 1.6), was employed due to factors such as reducing power balance, thermodynamic favourability (Kumar *et al.*, 2013), and reliance on one or two heterologous enzymes for the reduction of malonyl-CoA intermediate.

1.5 Thermophilic Organisms for Biofuel and Chemical Production

1.5.1 Advantages of thermophiles in biofuel and chemical production

The production of fuels and chemicals from lignocellulosic biomass by thermophilic organisms has various potential advantages for the fermentation process. High-temperature bioprocessing minimises the chances of microbial contamination compared to mesophilic organisms and reduces cooling costs between successive fermentation cycles due to a reduction in energy input. The high temperature also facilitates the removal and recovery of volatile products such as ethanol, thereby minimising end-product inhibition (Kananaviciute and Citavicius, 2015). In addition, fermentation at high temperatures (50 to 60 °C) allows for simultaneous feedstock degradation and fermentation and can reduce the loading of hydrolytic enzymes (Lin *et al.,* 2014). Other advantages include reduced risk of phage attack, increased solubility of substrates such as lignocellulosic biomass, and allowing unfavourable reactions in mesophiles to be thermodynamically possible (Zeldes *et al.,* 2015). Thermophilic organisms have been of significant interest in industrial biotechnology due to their ability to produce thermostable enzymes. Moreover, given the usefulness of thermostable enzymes in molecular biology methods, they have been considered powerful tools for industrial catalysis (Zeldes *et al.*, 2015).

Thermophilic organisms such as *Thermoanaerobacterium saccharolyticum*, *Parageobacillus thermoglucosidasius*, and *Clostridium thermocellum* have been used for bioethanol production (Shaw *et al.,* 2008; Cripps *et al.,* 2009; Argyros *et al.,* 2011). In addition, other thermophiles, such as the *Thermus thermophillus*, *Thermotoga maritima*, and archaeon *Thermococcus kodakariensis*, could be employed in bioprocessing at high temperatures (Taylor *et al.,* 2011). However, *Parageobacillus*

spp. have some advantages, including growing on various substrates (C5 and C6 sugars) and a rapid growth rate that allows high cell densities and, thus, high production rate of the required products (Kananaviciute and Citavicius, 2015). In view of the advantages of thermophilic organisms, this study aimed to produce sustainable biofuel and chemicals using *P. thermoglucosidasius* NCIMB 11955 as a microbial chassis.

1.5.2 The genus *Parageobacillus*

Parageobacillus spp. (previously called *Geobacillus*) (Aliyu *et al.,* 2016), are Gram-positive, rod-shaped, endospore-forming, aerobic or facultative anaerobic organisms. The genus comprises thermophilic bacteria belonging to the family *Bacillaceae*, class *Bacilli* and phylum *Firmicutes*. Initially categorised as 'Group 5' organisms within the *Bacillus* genus, these bacteria were later grouped into a new genus *Geobacillus* in 2001 (Nazina *et al.,* 2001). They grow optimally at temperatures between 55 and 65 °C. However, growth temperature can vary from 37 to 75 °C in some strains or species. They are neutrophilic, growing in a pH range of 6.0-8.5, with optimum growth at 6.2–7.5. They also have a relatively high G-C content of DNA, 48.2-58 mol% (Nazina *et al.*, 2001). Most species are ubiquitous and widely distributed in nature. They can be isolated from many environmental samples, including subterranean oilfields, soils of moderate or hot temperatures, to even permafrost temperatures (Kananaviciute and Citavicius, 2015). *Parageobacillus* species are not fastidious and can metabolise diverse carbon sources. They carry out mixed acid fermentation to produce lactic acid, formic acid, acetic acid, and ethanol from pentose and hexose sugars (Cripps *et al.,* 2009). This study is based on the type strain *P. thermoglucosidasius* NCIMB 11955.

1.5.3 *Parageobacillus thermoglucosidasius* NCIMB 11955

P. thermoglucosidasius is a thermophilic, Gram-positive, rod-shaped, endospore-forming bacterium belonging to the phylum *Firmicutes*. First isolated from soil samples in Japan in 1983, *P. thermoglucosidasius* was initially classified as *Bacillus thermoglucosidasius* (Suzuki *et al.,* 1983). The genus *Bacillus* encompasses a diverse group of Gram-positive, aerobic or facultatively anaerobic, spore-forming and rod-shaped bacteria. Members of the genus are phenotypically heterogeneous, exhibiting a wide range of requirements and metabolic diversity (Ash *et al.,* 1991). To determine the phylogenetic structure of the genus, a comparison of the 16S ribosomal RNA (rRNA) sequences of 51 *Bacillus* species by reverse transcription was performed by Ash *et al.* (1991). Based on the sequence data, the genus was subdivided into five distinct phylogenetic groups, placing *B. thermoglucosidasius* into group five. Similar bacteria were also isolated at hot oilfields in China, Kazakhstan and Russia, and preliminary characterisation revealed there were identical to Ash's group five of the genus *Bacillus* (Nazina *et al.,* 2001). After further detailed characterisation of the isolates, including physiological and cultural features, fatty acid composition and DNA-DNA similarity to other thermophilic *Bacillus* species, a new genus called *Geobacillus* was proposed containing *B. thermoglucosidasius* among others (Nazina *et al.,* 2001). Phylogenomic re-assessment of the thermophilic genus *Geobacillus* using whole genome sequencing and phylogenomic metrics such as average amino acid identity (AAI), average nucleotide identity (ANI), and digital DNA-DNA hybridisation (dDDH) revealed that the genus is made up of clades I and II (Aliyu *et al.,* 2016). The authors suggest that the *Geobacillus* species residing within clade II, which comprised fourteen strains including *P. thermoglucosidasius,* be considered a new genus called *Parageobacillus* (Aliyu *et al.,* 2016).

P. thermoglucosidasius NCIMB 11955 has an optimum growth temperature of 50-60 $^{\circ}$C, pH 6.0-8.0, and produces lactate (210 mM), acetate (30 mM), formate (26 mM), and ethanol (69 mM) under fermentative conditions as products (Cripps *et al.*, 2009). *P. thermoglucosidasius* can ferment a wide range of substrates, including pentoses and hexoses, and can tolerate up to 10% (v/v) ethanol concentrations (Tang *et al.*, 2009), making it an ideal process organism for biofuel production.

In an effort to exploit the inherent advantages of thermophiles, Cripps *et al.* (2009) genetically engineered *P. thermoglucosidasius* NCIMB 11955 to produce ethanol as the sole product during fermentation by diverting the carbon flux from a typically mixed acid pathway to an ethanol-producing pathway. The metabolic engineering strategy involved the disruption of *ldhA* (encoding lactate dehydrogenase) and *pflB* (encoding pyruvate formate lyase) genes, together with placing the *pdhA* (encoding pyruvate dehydrogenase) gene under the control of the strong P_{ldh} promoter from *Geobacillus stearothermophilus*, thereby producing a triple mutant strain designated as TM242 (Δ*ldh*, Δ*pfl*, pdh^{up}). Production of ethanol from the strain TM242 was over 90% of the theoretical yield at 60 $^{\circ}$C (Cripps *et al.*, 2009). Strain TM242 was also demonstrated to ferment cellobiose and a mixture of hexose and pentose efficiently. Recently, the TM242 strain was also shown to produce ethanol at yields equating to 92% of theoretical yields (9.9 ± 0.4 g·l^{-1}) during fermentation with palm kernel cake (PKC) as the substrate, a mannan-rich waste product from palm oil processing (Raita *et al.*, 2016).

1.6 Genetic Tools for Metabolic Engineering of *P. thermoglucosidasius*

Sheng *et al.* (2016) sequenced and made available in GenBank the complete genome sequence of *P. thermoglucosidasius* NCIMB 11955 under the accession numbers CP016622-CP016624. Using the allele-coupled exchange (ACE) technique of genome editing, these authors produced a *P. thermoglucosidasius* LS242 strain (Sheng *et al.*, 2017) equivalent to the TM242. More recently, a novel and efficient gene-editing tool based on the CRISPR/Cas9 system was developed for *P. thermoglucosidasius* NCIMB 11955 (Lau *et al.,* 2021). Thus, the availability of genome sequence and genetic engineering tools for *P. thermoglucosidasius* NCIM 11955 further highlights the suitability of this organism as a suitable chassis for the production of biofuels and bio-based chemicals production.

In general, the metabolic engineering of microorganisms towards higher production of biofuels and chemicals can be achieved either by modification of their endogenous pathways or through the implementation of heterologous pathways. The former involves knocking out (KO) competing pathways, thereby eliminating the formation of by-products and potentially overexpression of endogenous pathway (Kumar and Prasad, 2011). In contrast, the latter requires knock-in (KI) of the genes of interest either integrated into the chromosome or introduced as an operon on a plasmid. However, to achieve higher titres, yield and productivity for the cost-effective production of biofuels and chemicals, more metabolic engineering of the microorganisms or pathways are desirable.

Efficient genome modification required to implement the changes in the metabolic pathways of a microbial host strain is largely dependent on the availability of genetic tools. Limited tools constrain the genetic manipulation of *Parageobacillus* compared to

other bacteria such as *B. subtilis* and *E. coli*. However, significant progress is being made in enhancing the genetic tools for this group of microbes (Kananaviciute and Citavicius, 2015; Sheng *et al.,* 2017; Lau *et al.,* 2021). In the case of the *P. thermoglucosidasius* TM242 production strain, genetic manipulations were made by sequential deletion and insertion methods through the introduction of a centrally located NotI site in the coding sequence (Cripps *et al.,* 2009). However, this method is highly labour intensive and time-consuming because it requires extensive screening to select double-crossover mutants.

1.6.1 Allele-Coupled Exchange

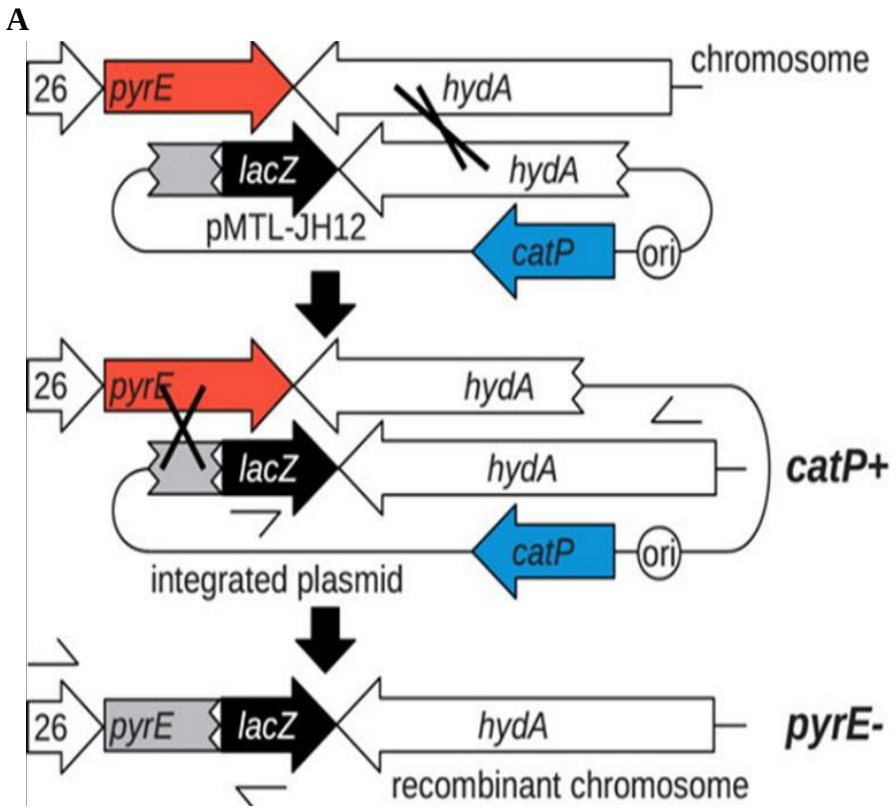
Allele-coupled Exchange (ACE) is a technique that eases the rapid insertion of heterologous DNA into a microbe genome. Heap *et al*. (2012) initially demonstrated the method in *Clostridium acetobutylicum*. The method exploits both the inactivation and reactivation of *pyrE* to produce a positively or negatively selectable allele. ACE is dependent on homologous recombination events and requires a two-step process to create a mutant either by gene knockout or by knock-in. The allelic exchange plasmid is integrated into the chromosome in the first recombination step, known as the single-crossover event. The resultant cells (single cross-over integrants) are selected via the acquisition of antibiotic resistance conferred by a plasmid-encoded antibiotic resistance gene. In the second step, the plasmid backbone is excised from the chromosome via a second homologous recombination step termed double-crossover event, and the cells (double cross-over mutants) are selected through resistance to a selective agent conferred by a counter-selection marker encoded by the plasmid (Heap *et al.,* 2012). ACE is part of the roadmap for gene system development in *Clostridium*, developed at the SBRC, University of Nottingham (Minton *et al.,* 2016).

Sheng *et al.* (2017) deployed ACE in *P. thermoglucosidasius* NCIMB 11955 based on *pyrE* as a counter-selection marker. The *pyrE* gene encodes orotate phosphoribosyltransferase, which converts orotic acid, an intermediate of the pyrimidine biosynthesis pathway, into orotidine 5-monophosphate (OMP) which is subsequently converted to uridine 5-monophosphate (UMP) by the *pyrF*-encoded orotidine 5′-phosphate decarboxylase. Orotate phosphoribosyltransferase also converts an analogue of orotic acid, 5-fluoroorotic acid (FOA), into 5-fluoroorotidine monophosphate (5-FOMP), which is later converted to 5-fluorouridine monophosphate (5-FUMP) by orotidine 5′-phosphate decarboxylase, rather than UMP. Accumulation of 5-FUMP is highly toxic to the cell and leads to cell death. Therefore, the deletion of *pyrE* in cells leads to FOA-resistant phenotype (FOA^R) and uracil auxotrophy (Minton *et al.*, 2016).

The first step in the ACE technique is the generation of the *pyrE* mutant (*pyrE* $^-$) strain of *P. thermoglucosidasius*. A *pyrE* KO vector, a unique allele exchange cassette containing two regions of homology of different lengths, was designed to control the recombination events' order. The homology arm consists of a long right homology arm (RHA) corresponding to the 1200 bp immediately downstream of *pyrE* and a short, left homology arm (LHA) corresponding to a 300 bp internal portion of *pyrE*. The long RHA directs plasmid integration through the first recombination event without inactivating the *pyrE*. The short LHA directs the second recombination event leading to plasmid excision with resultant double-crossover cells (FAO-resistant). The *pyrE* gene is inactivated, creating a *pyrE* negative strain (Figure 1.10a).

Heap *et al.* (2012) developed a simple system for gene knock-in with the *pyrE* based allelic exchange in *C. acetobutylicum.* Briefly, the DNA of interest to be integrated into the genome is flanked by the same 1200 bp RHA and the LHA being the corresponding

portion of partial *pyrE* gene. Thus, double homologous recombination results in full-length and functional *pyrE* gene recreation. The additional heterologous DNA sequences carried by the plasmid become integrated immediately downstream of the *pyrE* gene. Cells in which this recombination event has occurred are selected on media devoid of uracil, where only the recombinants with a restored, full length *pyrE* gene can synthesise uracil and are thus, able to grow (Figure 1.10b).



B

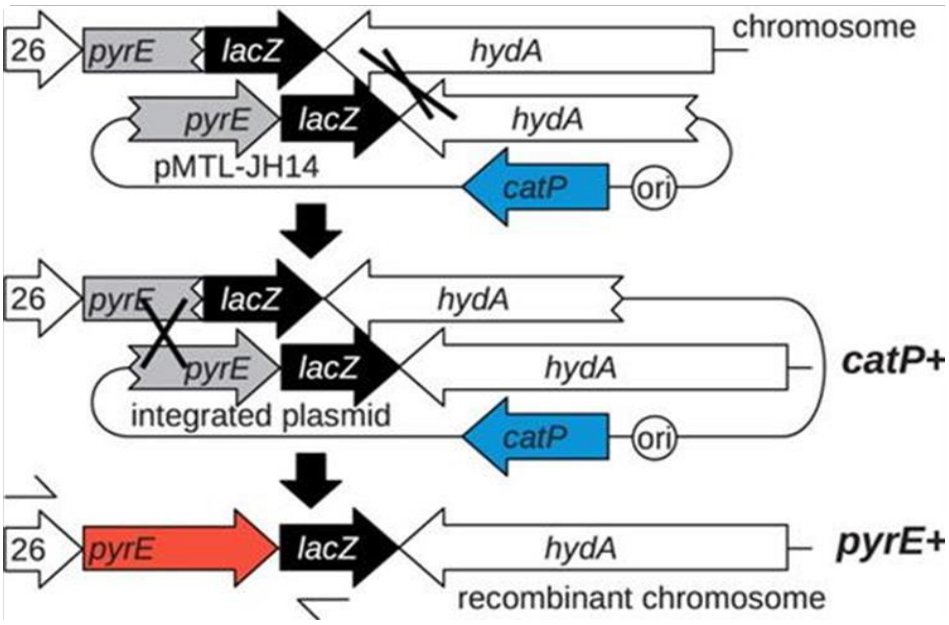


Figure 1.10: Schematic representation of DNA integration at the *pyrE* locus of *C. acetobutylicum*. (A) Generation of a *pyrE* mutant. Selection of stable double-crossover clones using pMTL-JH12, which has homology arms on either side of *lacZα*. The extended homology region mediates the first recombination event (plasmid integration) between pMTL-JH12 and *hydA*. Single-crossover clones are obtained on a medium containing thiamphenicol. The short homology region mediates the second recombination event (plasmid excision) between pMTL-JH12 and an inner portion of *pyrE*. Double-crossover clones are selected using 5-FOA. **(B)** Repair of *pyrE* with co-integration of DNA. Selection of stable double-crossover clones using pMTL-JH14. The extended homology region mediates the first recombination event (plasmid integration) between pMTL-JH14 and *hydA* / *lacZ*. Single-crossover clones are obtained on a medium containing thiamphenicol. The short homology region mediates the second recombination event (plasmid excision) between pMTL-JH14 and the corresponding portion of *pyrE*. Double-crossover clones restore the full-length *pyrE* gene with the co-

integration of DNA of interest in place of *lacZα*. They are selected using a growth medium lacking uracil. Adapted from Heap *et al.* (2012).

1.6.2 CRISPR/Cas9

1.6.2.1 CRISPR/Cas systems in adaptive bacterial immunity

CRISPR/Cas refers to Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated proteins (Cas). CRISPR are clusters of short-interval palindromic repeat sequences followed by a DNA spacer, while Cas are proteins associated with the CRISPR gene, having helicase and nuclease activities (Chylinski *et al.*, 2014). Ishino *et al.* (1987) first discovered CRISPRs in the genome of *E. coli*. Nearly 40% and 90% of genomes of sequenced bacteria and archaea, respectively, have been shown to possess the CRISPR loci (Sorek *et al.*, 2008). The CRISPR/Cas provide adaptive immunity against phage infection and plasmids by identifying and selectively cutting foreign DNA (Barrangou *et al.*, 2007).

There are two major CRISPR/Cas systems labelled, class 1 and class 2 and, classified based on Cas proteins. Class 1 includes types I, III and IV, whereas class 2 comprises types II, V and VI. Types I, II, and III represent the three major types. While the CRISPR/Cas systems in types I and III are complex in their conferring immunity mechanism and have not been exploited in genetic engineering, the type II CRISPR-Cas system is the simplest and employs a single multi-functional Cas9 protein to cut with exogenous DNA (Makarova *et al.*, 2015).

The CRISPR/Cas9 system (Figure 1.11) requires three components to efficiently cut foreign DNA: CRISPR RNA (crRNA), trans-activating crRNA (tracrRNA), and Cas9 Hryhorowicz *et al.*, 2017). The CRISPR associated adaptive immunity of the type II

CRISPR/Cas9 system consists of three significant steps: namely acquisition of CRISPRs, crRNA biogenesis and interference with foreign DNA. Upon exposure to invading DNA, the Cas nuclease cleaves the heterologous DNA into small fragments, called protospacers. These short DNA sequences are incorporated into the CRISPR locus of the bacterial genome (Wiedenheft *et al.,* 2012). The CRISPR locus consists of palindromic repeats separated by spacers acquired from invading DNA. The selection of protospacers is through a specific recognition sequence present within the foreign DNA, called protospacer adjacent motifs (PAMs). However, the PAMs are not present within the protospacer incorporated, and the sequences are specific to the Cas9 protein from different species (Mojica *et al.,* 2009). Subsequently, the CRISPR locus is transcribed to form pre-crRNA (precursor CRISPR RNA) in the biogenesis phase. The tracrRNA then hybridises to the repeat sequences of the pre-crRNA, after which, RNase III cleaves this complex, producing complete crRNAs consisting of a spacer and repeat sequence (Deltcheva *et al.,* 2011). Finally, the mature crRNA complexes with Cas9 protein and guide it to the complementary foreign DNA. The Cas9 protein degrades the invading DNA sequences (Garneau *et al.,* 2010). The Cas9 protein cut the DNA complementary and non-complementary strands through two nuclease domains, HNH and RuvC, by instituting double-stranded breaks upstream of the PAM (Jinek *et al.,* 2012). The three stages of CRISPR/Cas adaptive immunity are schematically illustrated in Figure 1.11.

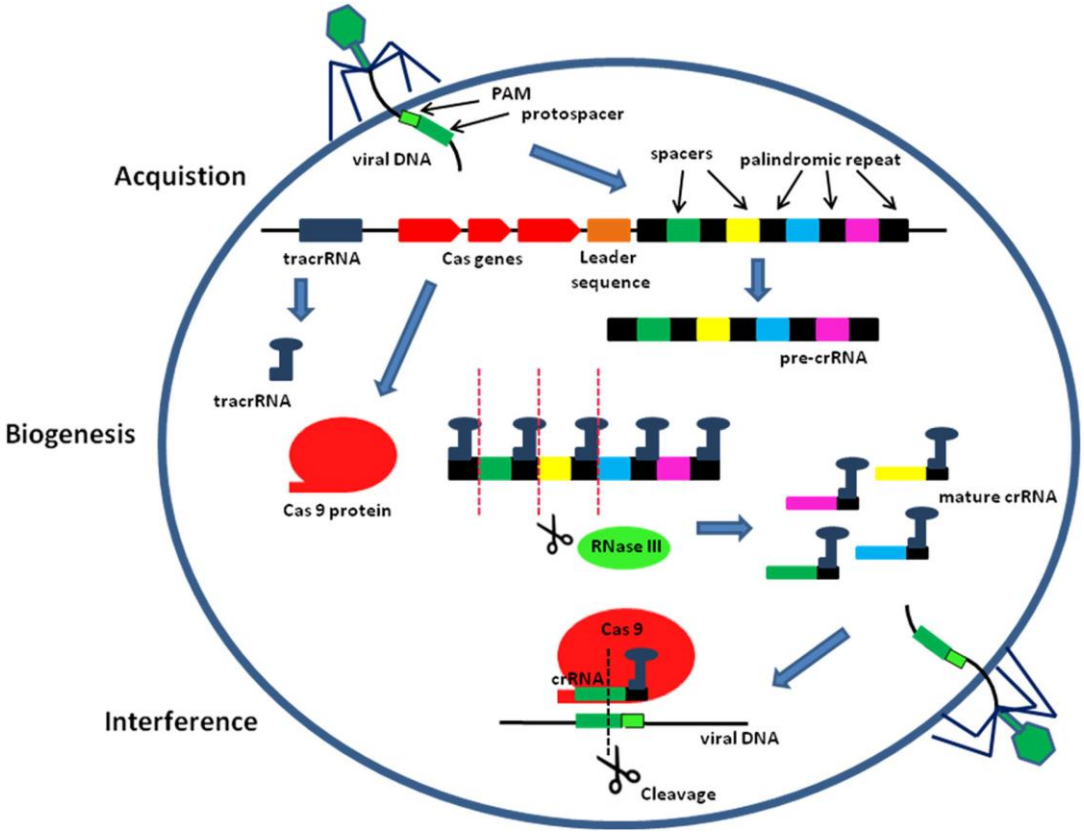


Figure 1.11: A schematic illustration of CRISPR/Cas bacterial adaptive immune system showing the three stages of acquisition, biogenesis of crRNA and interference of invading DNA. In the acquisition stage, the invading phage DNA is processed by a Cas nuclease into small DNA fragments, called protospacer sequences, and then incorporated into the CRISPR locus of the bacterial genome as a new spacer. Each CRISPR array encodes acquired spacers that are separated by repeat sequences. The selection of protospacers depends in part on the specific recognition of protospacer adjacent motifs (PAMs) present within the viral genome. However, protospacer sequences incorporated into the CRISPR locus do not contain PAM sites. Subsequently, in the biogenesis step, the CRISPR locus is transcribed into a precursor CRISPR RNA (pre-crRNA). TracrRNA hybridizes to the repeat sequences of the pre-crRNA and then endogenous RNase III cleaves this complex, yielding mature crRNAs, each containing one spacer and partial repeat sequence. Finally, in the interference step, mature crRNA

guides Cas9 protein to the complementary foreign DNA, triggering degradation of the DNA sequences of invading phages. Adapted from Hryhorowicz *et al.* (2017).

1.6.2.2 The CRISPR/Cas9 system in genome engineering

The CRISPR/Cas9 system of adaptive immunity has been exploited to manipulate the genome of microbes. This genome editing method is based on the heterologous expression of engineered Cas9 nuclease and a single guide RNA (sgRNA) that directs the nuclease to the targeted protospacer (Mougiakos *et al.,* 2016). The most widely used CRISPR-Cas9 system for genome modification is the *Streptococcus pyogenes* Cas9 (SpCas9). However, researchers have now focused on the CRISPR/Cas9 system of *Streptococcus thermophilus* (StCas9) due to advantages such as high specificity, small Cas9 fragment, and low off-target effects (Hao *et al.,* 2018). The sgRNA consists of a crRNA element and a tracrRNA element that facilitate the rapid application of the CRISPR/Cas9 system for genome editing. The recognition of the target DNA by Cas9 requires complementarity between the sgRNA and the targeted DNA upstream of the specific PAM. The Cas9-sgRNA complex introduces a double-stranded break (DSB) at the target sites. Through DNA repair processes of the DSB, the genome of a microbe can be manipulated. Homologous directed repair (HDR) and non-homologous end joining (NHEJ) are the two mechanisms by which DSBs are repaired. In HDR, the DSB is repaired via homologous recombination of exogenous DNA, and in the process, the genome is edited. On the other hand, NHEJ corrects DSBs through error-prone means leading to insertions and deletions (indels) at the target sites, thereby disrupting the target gene (Hryhorowicz *et al.,* 2017). The genome editing system is often delivered on a plasmid containing a homology editing template, making Cas9 an efficient counter-

selection marker for cells which have successfully undergone a two-step homologous recombination.

1.6.2.3 Previous studies on CRISPR/Cas9 genetic tool in *P. thermoglucosidasius*

Lau *et al.* (2021) developed a CRISPR/Cas9 tool for genome engineering in *P. thermoglucosidasius* NCIMB 11955 based on *S. thermophilus* CRISPR/Cas9 (StCas9). The successful implementation of this genetic tool was initially exemplified by the knockout of the endogenous acetate kinase (*ack*) gene in the organism. To further demonstrate the efficiency of the tool in genome editing, integrating exogenous DNA (eCGP123- a reporter gene) into both the chromosome and the megaplasmids present in *P. thermoglucosidasius* was accomplished (Lau *et al.*, 2021). Additionally, validation of the Cas9 genome-editing tool was further attained via the deletion of restriction-modification systems, leading to an increase in transformation efficiency. The CRISPR/Cas9 was also tested to cure plasmids by removing pNCI001 and pNCI002 megaplasmids present in *P. thermoglucosidasius.* Finally, the applicability of the CRISPR/Cas9 as a genetic tool for metabolic engineering in *P. thermoglucosidasius* was demonstrated through increased succinate production achieved by the deletion of succinate dehydrogenase. This deletion results in a 3.5-fold increase in succinate production compared to wild type (Lau, 2018). Thus, this successful implementation has paved the way for utilising this genetic tool to engineer *P. thermoglucosidasius* to produce biofuels and value-added products.

1.7 Lignocellulosic Biomass as Feedstock for Biofuel and Chemical Production

1.7.1 Composition of lignocellulosic biomass

Lignocellulosic biomass has gained increased attention as feedstock for biofuel and bio-based chemical production due to its abundance in nature and non-competition with food (Jiang *et al.*, 2017). The main constituents of lignocellulosic biomass are cellulose (40-55%), hemicellulose (25-50%), and lignin (10-40%) (Figure 1.12), depending on the source of the biomass, whether grasses, softwood or hardwood (Grange *et al.*, 2010). Cellulose is a linear, crystalline polymer made up of repeating D-glucose units. This cellulose polymer is water-insoluble and difficult to hydrolyse. On the other hand, hemicellulose comprises branched heteropolymers consisting of D-glucose, D-galactose, D-mannose, L-arabinose, and D-xylose as well as D-glucuronic and D-galacturonic acids. This structure is amorphous and hence, comparatively easy to hydrolyse. Lignin is a three-dimensional polymer synthesised from phenylpropanoid precursors, P-coumaryl, sinapyl and coniferyl alcohol-containing aromatic and aliphatic parts (Singh and Satapathy, 2018).

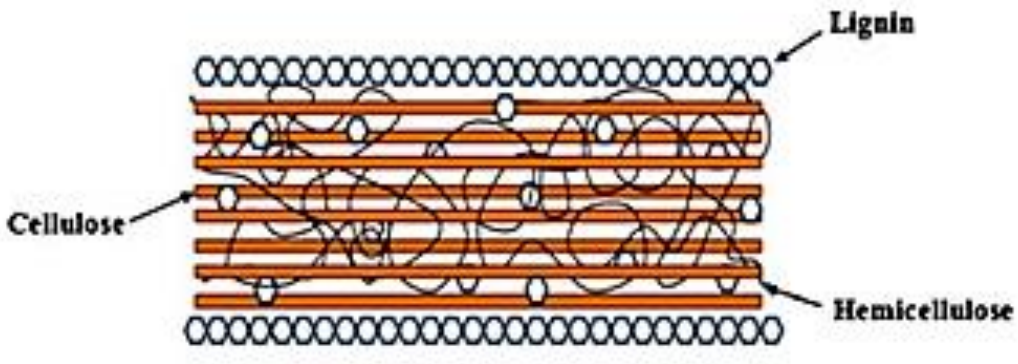


Figure 1.12: Structure of lignocellulosic biomass showing organisations of cellulose, hemicellulose and lignin. Adapted from Haghighi Mood *et al.* (2013).

1.7.2 Conversion process of lignocellulosic biomass to biofuels and chemicals

The overall conversion process of lignocellulosic biomass to biofuels and chemicals consists of three main steps: pre-treatment of the lignocellulosic biomass, enzymatic hydrolysis of cellulosic biomass and removal of lignin portion as solid residue, and fermentation of glucose and xylose to biofuel or chemical.

1.7.2.1 Pre-treatment of lignocellulosic biomass

Lignocellulosic biomass is recalcitrant due to the extreme complexity of the cell wall matrix, which causes difficulty in biomass degradation. Hence, the biomass is first pre-treated to break apart the lignocellulosic material. The primary purpose of pre-treatment is to separate the biomass into its components to make it accessible to hydrolytic enzymes (Figure 1.13). Pre-treatment methods are classified into different categories, namely, (a) physical methods such as grinding, sonication, microwave, and milling or, (b) chemical methods such as acidic pretreatment, alkaline pretreatment, ozonolysis, and organosolv process (use of organic or aqueous organic solvent mixtures with inorganic acid catalysts to extract lignin from lignocellulosic biomass), (c) physicochemical methods, such as ammonia fibre explosion (AFEX), steam explosion (autohydrolysis), hot water, wet oxidation, and CO_2 explosion, and (d) biological methods using microbes such as brown and white rot fungi which degrade hemicellulose and lignin, or enzymes such as laccase and peroxidase which degrade lignin (Maurya *et al.*, 2015; Bhatia *et al.*, 2017).

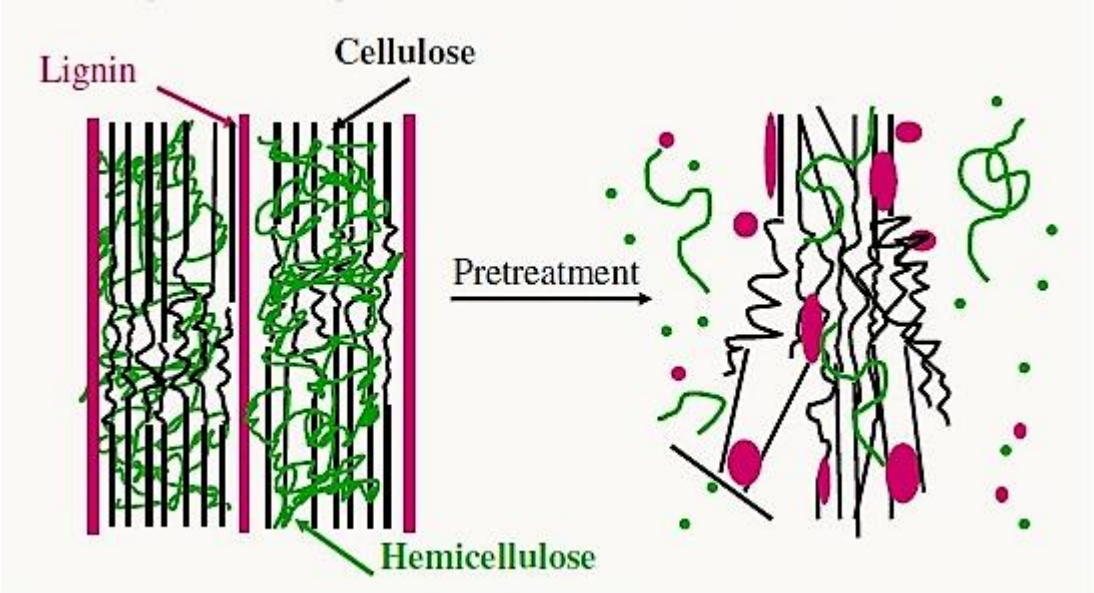


Figure 1.13: Effect of pretreatment on lignocellulosic biomass. Adapted from Liu and Fei (2013).

1.7.2.2 Enzymatic hydrolysis of cellulosic biomass

After pre-treatment, the cellulosic biomass is subjected to the action of hydrolytic enzymes. The complete enzymatic hydrolysis of cellulosic biomass requires the complementary activity of three enzyme classes: endoglucanases, exoglucanases or cellobiohydrolases, and β-glucosidases (Zhang and Lynd, 2004). Endoglucanases act on amorphous components of lignocellulose (hemicelluloses) and produce cellobiose and cello-oligosaccharides as hydrolysis products. Cellobiohydrolases yield cellobiose solely as the product of hydrolysis from the insoluble cellulose polymers. Finally, β-glucosidases convert the cellobiose and cello-oligosaccharides to produce glucose (Grange *et al.*, 2010).

1.7.2.3 Fermentation

The final process in the bioconversion of lignocellulosic biomass is the fermentation of simple hexose and pentose monomers/dimers sugars produced via enzymatic hydrolysis to biofuels or chemicals. Different approaches are used to perform out the fermentation by microbes, namely, separate hydrolysis and fermentation (SHF), simultaneous saccharification and fermentation (SSF), and simultaneous saccharification and co-fermentation (SScF) (Singh and Satapathy, 2018).

SHF is a method by which enzymatic hydrolysis and fermentation are performed sequentially. In the SHF process, enzymatic saccharification of pretreated lignocellulosic biomass is carried out first at the optimal temperature of the saccharifying enzyme. Subsequently, appropriate microorganisms are used to ferment the saccharified solution (Szambelan *et al.,* 2018).

SSF is a method by which enzymatic hydrolysis and fermentation are performed simultaneously in the same reactor by a single microbe. SSF is advantageous over the SHF process because it requires less equipment and fermentation time, thereby reducing costs. However, in the SSF process, the optimum temperature of enzymatic hydrolysis is typically greater than the yeast fermentation temperature (Miah *et al.,* 2022). Therefore, thermotolerant microorganisms such as *P. thermoglucosidasius* represent an ideal host.

SScF is a process in which enzymatic hydrolysis of the pretreated biomass occurs simultaneously with the co-fermentation of hexose and pentose sugars (mainly glucose and xylose) in a single unit. SScF offers the potential of streamlined processing while reducing capital costs compared to SHF (Qin *et al.,* 2018).

1.7.3 Consolidated bioprocessing

Consolidated bioprocessing (CBP) is a process in which microbes carry out cellulose hydrolysis and fermentation of simple sugars into biofuels and chemicals in a one-step process without the addition of extrageneous cellulases (Higashide *et al.,* 2011) (Figure 1.14). Due to its resistance to degradation, lignocellulosic biomass must be hydrolysed first, using pretreatment and hydrolysis methods. Consequently, the industrial-scale production of biofuels is hampered due to the cost of hydrolytic enzymes (Jiang *et al.,* 2017). The CBP approach is a promising solution for the low-cost production of biofuels and chemicals from lignocellulosic biomass. At present, two strategies are employed to achieve CBP. The first strategy referred to as the "native cellulolytic strategy", involves improving or engineering biofuel production ability of naturally cellulolytic microbes. Several promising cellulolytic organisms, such as *Clostridium cellulolyticum* (Higashide *et al.,* 2011) and *C. thermocellum* (Lin *et al.,* 2015), have been manipulated to produce biofuel directly from cellulose. The second strategy referred to as the "recombinant cellulolytic strategy", involves engineering non-native cellulolytic microbes with biofuel production ability to utilise cellulose directly (Grange *et al.,* 2010). *P. thermoglucosidasius* is a promising CBP organism employing the recombinant cellulolytic strategy (Bashir *et al.,* 2019).

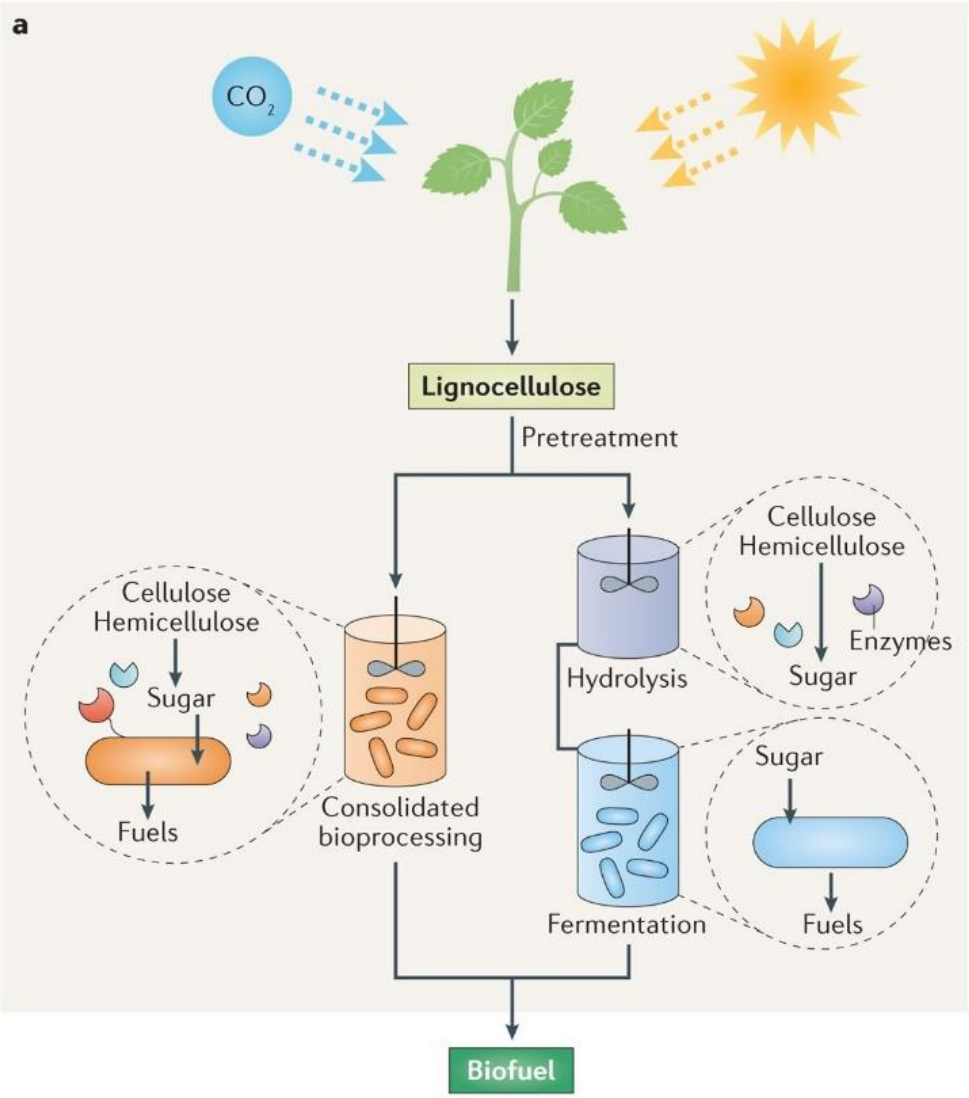


Figure 1.14: Consolidated bioprocessing. The lignocellulose of plant biomass can be converted to fuels through hydrolysis followed by fermentation, or consolidated bioprocessing, which combines the two processes in one reactor. Adapted from Liao *et al.* (2016).

1.8 Hypotheses

1) By systematically re-engineering *P. thermoglucosidasius* TM242 using a theophylline-responsive CRISPR/Cas9, we can understand how and why this strain produces ethanol from glucose with efficiencies that approach the theoretical maximum for a fermentative organism.

2) The native acetyl-CoA carboxylase in *P. thermoglucosidasius* can be used to produce 3-HP via the malonyl-CoA pathway.

3) By identifying genes responsible for 3-HP degradation, a strain incapable of 3-HP breakdown can be engineered.

4) *P. thermoglucosidasius* can be engineered to produce isobutanol via amino acid (valine) biosynthetic pathway.

1.9 Aims of the Study

The main aim of the present study was to engineer *P. thermoglucosidasius* NCIMB 11955 for sustainable fuels and chemical production. The study initially focused on adding to and characterising the pMTL60000 modular vector series developed for *P. thermoglucosidasius* NCIMB 11955 (Sheng *et al.*, 2017; Spencer, 2018).

The overall study was divided into two major parts. The first part was implementing the theophylline-responsive CRISPR/Cas9 gene-editing tool based on riboswitches (RiboCas) (Cañadas *et al.*, 2019) to metabolically engineer *P. thermoglucosidasius* NCIMB 11955 for bioethanol production. The second part was to engineer the metabolic pathways in *P. thermoglucosidasius* NCIMB 11955 to the produce platform chemical 3-hydroxypropionic acid via the malonyl-CoA pathway and investigate the native 3-hydroxypropionic acid catabolism. Finally, this study attempted to engineer *P.*

thermoglucosidasius NCIMB 11955 to produce isobutanol with the ultimate aim of cellulosic isobutanol production via an amino acid (valine) biosynthetic pathway.

CHAPTER 2

2.0 MATERIALS AND METHODS

2.1 Bioinformatics Methods

2.1.1 DNA database search

Searches for DNA sequences were performed using the Basic Local Alignment Search Tool (BLAST) algorithm through the National Centre for Biotechnology Information (NCBI) (U.S. National Library of Medicine Rockville Pike, Bethesda MD, USA). Available at https://www.ncbi.nlm.nih.gov/

2.1.2 ExPASy

Nucleotide (DNA) sequences obtained from Sanger sequencing were translated into protein sequences using the ExPASy translation tool. Available at https://web.expasy.org/translate/

2.1.3 Protein database search

Searches for translated protein sequences from protein databases were performed using the Blastp algorithm of the Basic Local Alignment Search Tool (Blastp) through the National Centre for Biotechnology Information (NCBI) (U.S. National Library of Medicine Rockville Pike, Bethesda MD, USA). Available at http://www.ncbi.nlm.nih.gov

2.1.4 Handling of DNA sequence data and plasmid map design

DNA sequence data were routinely handled using ApE (A plasmid Editor), SnapGene, and Benchling software. Plasmid maps were designed using ApE software (available at https://jorgensen.biology.utah.edu/wayned/ape/) and then visualised using SnapGene® software (from GSL Biotech; available at snapgene.com). Alignments of two or more

DNA sequences were predominantly performed using Benchling (https://benchling.com/) software, and to a lesser extent using ApE.

2.1.5 Oligonucleotides design, analysis and synthesiss

Oligonucleotides for the polymerase chain reaction (PCR) and sequencing of DNA were designed manually. For cloning-based-PCR, approximately 20-25 bp of forward and reverse primers identical to the target gene's desired 5′ and 3′ ends were designed, respectively. Appropriate restriction site(s) were added at the desired end of the primer sequence, and a 1-3 bp additional non-coding DNA sequence was incorporated, flanking the restriction site to improve cleavage using restriction enzymes. Oligonucleotide sequences were analysed using the OligoAnalyzer Tool of Integrated DNA Technologies (IDT) (Coralville, Iowa, USA). Available at https://www.idtdna.com/calc/analyzer. NEBuilder Assembly Tool was used to design primers for NEBuilder HiFi DNA Assembly reactions (https://nebuilder.neb.com/). Synthesis of oligonucleotide was performed by either Sigma Aldrich (St. Louis, MO, USA) or Eurofins MWG Operon (Ebersberg, Germany). All primers used in this study are listed in Table 2.1.

Table 2.1: List of primers used in this study

Name	Sequence (5′-3′)
ColE1_F2	CCGCCTTTGAGTGAGCTGATA
RepB_R1	CAGCAACTAAAATAAAAATGACGTTATTTC
BST1_F_AscI	tctaggcgcgccTCGCGCGTTTCGGTGAT
BST1_R_FseI	cttaggccggccTCCCTTTTCAGATAATTTTAGATTTGCTTTTC
sgRNA_Pdh	CGCGAATATGAAACGTTTTAGAGC
Pdh^{up}_HA_R	CCGTATCTTCGTTCAACCGC
sgRNA_Pfl	GAACAAAACGGCGGTTTTAGAG
Pfl_HA_R	GTTTCATAACAGTTTCCCTCCC
Cas3_Pfl_F	CAGTTTCCGGTGTTTTTTCTCATCG
Cas3_Pfl_R	GTTCCGCAAACGCTTAAGCCC
Cas3_Pdh^{up}_F	GCAGCGGTTTATCTGGTTGAC
Cas3_Pdh^{up}_R	GATGCCAGGAATTCCCGCT
SNP_G>A_F	CAGTGGAAACAACAATTGCGATG
SNP_G>A_R	CGTTTTCACATTCGAAGAATCGAG
FC_mmsA1	CCTTTCGGGATTTTAATGACTTCTTTAATG
RC_mmsA1	CGATACCGGGAAATAATCTTATATCG
FC_mmsA2	GTATTGTCGCTACGATTGCTTATAATTG
RC_mmsA2	GAACGGCTGGTGTTTGTGTATAAAC
ALS1_F_XhoI	gaatcctcgagTAAGGAGGAACTACTATGACGAAAGCGACG AAAGAACAAAAAAG
ALS2_F_XhoI	ttcactcgagTAAGGAGGAACTACTATGGGAGTAGGCACAG TCGAG
ALS3_F_XhoI	ttcactcgagTAAGGAGGAACTACTATGAATATTGCGGAAG AAATGGTTCAATC
ALS1_R_NheI	ttcagctagcTTATAACGCTTTTGTTTTCATTAATTCGCCAAA TT
ALS2_R_NheI	ttcagctagcTTAAACTAATTGATTAGGTAAAAGTTTTTTCG CAAG
ALS3_R_NheI	ttcagctagcTTACCGTGTTTCATCTCCACAGC
pLDH_F_SfbI	ttcacctgcaggCCAGGCATCAAATAAAACGAAAGGC
pLDH_R_LLKF	**AATAAATAATCGCCTACCGTATACAT**TGCATTCATCC TCCCTCAATATAATGC
LLKF_F_pLDH	**CATTATATTGAGGGAGGATGAATGCA**ATGTATACGG TAGGCGATTATTTATTAGATC
LLKF_R_XhoI	ttcactcgagTTAGCTTTTATTTTGTTCCGCAAATAATTTGC
pGPAD_F_SbfI	ttcacctgcaggCAAATAAAACGAAAGGCTCAGTCG
pGPAD_R_LLKF3	**CTAATAGGTAATCTCCTACTGTATACAT**AGTTCTACC TCCTTTATCTAATGCTATAC
LLKF3_F_pGpad	**GTATAGCATTAGATAAAGGAGGTAGAACT**ATG TATACAGTAGGAGATTACCTATTAGAC

LLKF3_R_XhoI	ttcactcgagTTATGATTTATTTTGTTCAGCAAATAGTTTACCC
pLDH_Native_F_SbfI	ttcacctgcaggCCAGGCATCAAATAAAACGAAAGGC
pLDH_Native_R_LLKF3	**CTAATAGGTAATCTCCTACTGTATACA**TGCATTCATCCTCCCTCAATATAATGC
LLKF3_Native_F_pLDH	**CGCATTATATTGAGGGAGGATGAATGCA**ATGTATACAGTAGGAGATTACCTATTAGAC
LLKF3_Native_R_XhoI	ttcactcgagTTATGATTTATTTTGTTCAGCAAATAGTTTACCC
pGPAD_F_NotI	tcagcggccgcCGAAGAAACGAAACAACAGCTTTG
pGPAD_R_LLKF(Mut.)	**CTAATAAATAATCGCCTACCGTATACAT**AGTTCTACCTCCTTTATCTAATGCTATAC
LLKF_Mut_F_pGPAD	**AGCATTAGATAAAGGAGGTAGAACT**ATGTATACGGTAGGCGATTATTTATTAGATC
LLKF_Mut_R_XhoI	ttcactcgagTTAGCTTTTATTTTGTTCCGCAAATAATTTGC
pGPAD_Seq	GAAACAACAGCTTTGGCAAATTTTCAAAAG
pLDH_Seq	GTCTGTCATGAAATGGACAAACAATAG
WT_LLM4_R_XhoI	ttcactcgagTTAGCTTTTATTTTGTTCAGCAAATAGTTTGC
LLKF_Seq1	GATCGCTTACATGAATTAGGCATTGAA
LLKF_Seq2	GTTGATGAAACATTACCGTCGTTTTTAG
LLKF_Seq3	CGATTTGTTTTATTATTAATAATGATGGCTATAC
B_ALS_Seq2	GCGACGGGATTATTAACAGCGAA
B_ALS_Seq3	GTCGACGATTAATCATATTGAACATGATG
LLKF_Native_Seq1	GACCGATTACACGAGTTAGGAATTGAA
LLKF_Native_Seq2	GTTGATGAAACTCTCCCTTCATTTTTAG
LLKF_Native_Seq3	CAATTTGCTTTATTATCAATAATGATGGTTATAC
LLM4_Seq1	GACCGCTTACACGAGTTGGGA
LLM4_Seq2	GGACGAAACGTTGCCGAGCTT
LLM4_Seq3	CTGCTTCATCATCAACAACGACG
ALS3_Seq 1	CTTGAAGAAAAAGGCGTACGGTATTTG
ALS3_Seq 2	CGGCGTTGTCCGAAAAACTTTG
ALS3_Seq 3	AGATGGAGGCTATGGCCTTATTG
ilvD_F_alsS1	**ACAAAAGCGTTATAA**ATGGAGGGATTGAGATTGG
ilvD_F_alsS2	**TAATCAATTAGTTTA**ATGGAGGGATTGAGATTGG
ilvD_F_alsS3	**GATGAAACACGGTAA**ATGGAGGGATTGAGATTGG
ilvD_R_Cpa	**GGCTTCTTATTTTTAT**TTAAATTTTCATAATCCCGCCCG
ilvD_R_adhA	**CTAAGTTCCTCCCTG**TTAAATTTTCATAATCCCGCCC
adhA_F_ilvD	**GGATTATGAAAATTTAA**CAGGGAGGAACTTAGATGAA
adhA_R_Cpa	**GGCTTCTTATTTTTAT**TTAACTGTTGGAAATAATGACTTTTAA

ilvD_Seq1	ATGGAGGGATTGAGATTGGGAAAAC
ilvD_Seq2	GAACGATATGGCTGCCCGAC
ilvD_Seq3	TATTCTGAAACGGGCGGTCTTG
adhA_Seq1	CAGGGAGGAACTTAGATGAAAGC
adhA_Seq2	CCACTTAATTGATGGAACTCAAGCAG
adhA_Seq3	ATTATGGAATTAACCGGCGGGAAAG
lepA1_F	CGCGATGAACGATGAAAGAAAG
lepA1_R	AGGGTATGCGTCGTTTGATTAT
pUB110_F	AACAAAATCGTGAAACAGGCG
pUB110_R	CCTTTTGTGACTGAATGCCATG

Restriction sites are represented in red and non-annealing sequences are indicated by lower case. Overhangs are represented in bold.

2.1.6 Synthesis of DNA fragments

DNA fragments and genes were synthesised by Integrated DNA Technologies (IDT) (Coralville, Iowa, USA) and Azenta Life Sciences Ltd (formerly GENEWIZ Ltd) (Burlington, MA, USA).

2.2 Reagents

Unless stated otherwise, all bacteriological media and chemicals were sourced from Sigma-Aldrich Company Ltd (Poole, UK) or Fisher Scientific Ltd (Loughborough, UK). DNA polymerases, restriction enzymes, DNA markers, reaction buffers were supplied by New England Biolabs Ltd. (Hitchin, UK) or Thermo Fischer Scientific (UK), T4 DNA ligase from Promega (Southampton, UK).

2.3 Bacterial Growth Media

Bacterial growth media were prepared in deionised water and autoclaved at 121 $^{\circ}$C for 20 min at 15 psi (pound-force per square inch). The media used during this study are described below

2.3.1 Luria-Bertani medium

Luria-Bertani broth consists of 5 $g \cdot l^{-1}$ yeast extract (Oxoid, UK), 10 $g \cdot l^{-1}$ tryptone (Difco Laboratories, UK) and 5 $g \cdot l^{-1}$ NaCl. Before autoclaving, the pH of the medium was adjusted to 7.0 using HCl or NaOH. To prepare Luria-Bertani agar, 10 $g \cdot l^{-1}$ of No. 1 Bacteriological agar (Oxoid, UK) was added to the Luria-Bertani broth before autoclaving.

2.3.2. Super optimal catabolite repression medium (SOC)

SOC medium was purchased from Invitrogen, UK.

2.3.3 Soy peptone yeast extract no glycerol (2SPYNG)

2SPYNG contained 16 g soy peptone (Sigma-Aldrich, St. Louis, MO, USA), 10 g yeast extract (Oxoid, UK) and 5 g NaCl per litre of deionised water. The pH was adjusted to 7.0 using 5 M KOH before autoclaving (Sheng *et al.*, 2017).

2.3.4 Soy peptone yeast extract (2SPY)

2SPY contained the same constituents as in 2SPYNG in addition to 10 g glycerol. The pH was adjusted to 7.0 using 5 M KOH before autoclaving (Sheng *et al.*, 2017).

2.3.5 Tryptic soy agar (TSA)

TSA agar was prepared using 40 g TSA (Sigma-Aldrich) per litre of deionised water and autoclaved (Sheng *et al.*, 2017).

2.3.6 *Clostridium* basal medium (CBM)

CBM was prepared according to O'Brien & Morris (1971) and consisted of 200 mg $MgSO_4 \cdot 7H_2O$, 7.58 mg $MnSO_4 \cdot H_2O$, 10 mg $FeSO_4 \cdot 7H_2O$, 1 mg p-aminobenzoic acid,

2 µg biotin, 1 mg Thiamine·HCl, 5 g Casamino acid, 0.5 g K_2HPO_4 and 0.5 g KH_2PO_4 per litre of deionised water. Minerals and vitamins were prepared as stock solutions of 1 mg/ml. K_2HPO_4 and KH_2PO_4 were made as 50 mM stock solution. The pH of the medium was adjusted to 7.0 using 5 M KOH before autoclaving. Appropriate sugars were added from sterile 10% solutions before autoclaving as required.

2.3.7 ASYE medium

ASYE medium was prepared as described by Cripps *et al.* (2009) comprising, 10 mM $NaH_2PO_4 \cdot 2H_2O$, 10 mM K_2SO_4, 2 mM citric acid, 1.25 mM $MgSO_4 \cdot 7H_2O$, 0.02 mM $CaCl_2 \cdot 2H_2O$ 1.65 mM $Na_2MoO_4 \cdot 2H_2O$, 20 mM $(NH_4)_2SO_4$, 25 µM $ZnSO_4 \cdot 7H_2O$, 100 µM $FeSO_4 \cdot 7H_2O$, 50 µM $MnSO_4 \cdot H_2O$, 5 µM $CuSO_4 \cdot 5H_2O$, 10 µM $CoSO_4 \cdot 7H_2O$, 16.85 µM $NiSO_4 \cdot 6H_2O$, 6.5 µM H_3BO_3 and 12.5 µM biotin, per litre of deionised water. Mineral and vitamins were prepared as 1 mg/ml stock solution. 10 $g \cdot l^{-1}$ Yeast extract and 20 $g \cdot l^{-1}$ glucose were added before filter sterilisation. Also added were 40 mM each (final concentrations) of Bis-Tris, PIPES, and HEPES at pH 7.0.

2.3.8 Modified ASYE medium (M-ASYE)

Modified ASYE medium was used for fermentation experiments and consists per litre of deionised water: M9 minimal salts-5X (Sigma), Trace Metal Mix A5-1000X dilution (Sigma), 0.24 g $MgSO_4$, 0.011 g $CaCl_2$, 0.01 g Thiamine), 0.384 g Citric acid, 0.0278 g $FeSO_4 \cdot 7H_2O$), 4.6 mg $NiCl_3 \cdot 6H_2O$, 47.6 g HEPES, 3.05 mg Biotin, 20 g Glucose, 10 g Yeast extract (Lin *et al.*, 2014).

2.3.9 Semi-synthetic medium

The defined semi-synthetic medium contained per litre of the following; 2.0 g $(NH_4)_2SO_4$, 12.5 g K_2HPO_4, 3.9 g KH_2PO_4, 0.25 g $MgSO_4 \cdot 7H_2O$, 0.02 g $CaCl_2$

·$2H_2O$, 8.0 g yeast extract, 10 g glucose, 2 ml vitamin stock solution, 2 ml trace metals solution, 2 ml Fe solution. The salts and appropriate sugars added were dissolved in water, and the vitamins, trace elements, Fe-solution were added. The pH of the solution was adjusted to 7.0 using 5 M KOH and then filtere sterilised.

Trace metals solution contained per liter: 70 mg $ZnCl_2$, 100 mg $MnCl_2 \cdot 4H_2O$, 60 mg H_3BO_3, 200 mg $CoCl_2 \cdot 4H_2O$, 20 mg $CuCl_2 \cdot 2H_2O$, 25 mg $NiCl_2 \cdot 6H_2O$, 35 mg $Na_2MoO_4 \cdot 2H_2O$ and 4 ml 37%, w/v HCl. The solution was prepared by dissolving all the above components, and the solution volume was adjusted to 1 litre and filter sterilised before storage at 4 °C.

Vitamins solution was composed of 50 mg biotin, 100 mg p-aminobenzoic acid, 5 g nicotinamide, 1 g pyridoxal phosphate, 1 g thiamine-HCl, and 2.5 g riboflavin per litre of solution. The vitamin solution was filter sterilised before storage at 4 °C.

Fe-solution contained 5 g $FeSO_4 \cdot 7H_2O$ and 4 ml 37%, w/v HCl per litre of solution.

2.4 Antibiotics and Supplements

Antibiotics and supplements were added to the growth medium at the following final concentrations, as presented in Table 2.2. According to the manufacturers' specifications, stock solutions of kanamycin and spectinomycin antibiotics were prepared. Theophylline was prepared from stock solution at 266 mM (48 mg/ml in DMSO). Theophylline was prepared fresh before each use.

Table 2.2: Concentrations of antibiotics and supplements in media

	E. coli	*P. thermoglucosidasius*
Kanamycin (Kan)	50 $\mu g \cdot ml^{-1}$	12.5 $\mu g \cdot ml^{-1}$
Spectinomycin	50 $\mu g \cdot ml^{-1}$	12.5 $\mu g \cdot ml^{-1}$
2-Ketoisovalerate (2-KIV)	-	5 $g \cdot l^{-1}$
Theophylline	-	8 mM

2.5 Bacterial Growth Conditions

All the bacterial strains used are listed in Table 2.3.

2.5.1 *E. coli*

E. coli Top10 (Invitrogen) and *E. coli* NEB-stable (New England Biolabs, UK) strains were grown in 5 ml Luria-Bertani media containing 50 $\mu g \cdot ml^{-1}$ kanamycin or 50 $\mu g \cdot ml^{-1}$ spectinomycin at 37 $^{\circ}$C or 30 $^{\circ}$C, respectively on a rotary shaker at 200 rpm overnight. Alternatively, they were grown overnight on Luria-Bertani agar plates with 50 $\mu g \cdot ml^{-1}$ kanamycin or 50 $\mu g \cdot ml^{-1}$ spectinomycin at 37 $^{\circ}$C or 30 $^{\circ}$C. *E. coli* strains were only used for cloning purposes throughout the study.

2.5.2 *P. thermoglucosidasius*

P. thermoglucosidasius wild-type and recombinant strains were grown in 2SPYNG or ASYE containing appropriate antibiotics in either 50 ml falcon tubes or 250 ml baffled conical flasks. Cultures were grown overnight at 52 or 60 $^{\circ}$C, shaking at 250 rpm, or alternatively, on TSA agar plates with the appropriate antibiotics at 52 $^{\circ}$C.

2.5.3 Measurement of bacterial growth

Bacterial growth in a liquid medium was monitored by measuring optical density at 600 nm (OD_{600}) using a Jenway 6300 (Jenway Ltd, UK). Samples were first diluted to 10^{-1} before measurement.

2.6 Bacterial Storage

E. coli strains were stored using Microbank™ cryogenic bead tubes (Pro-Lab Diagnostics, Richmond Hill, ON, Canada) following the manufacturer's instructions. *P. thermoglucosidasius* strains were stored as 1 ml glycerol stock in 2 ml sterile crew cap microtubes. Culture stocks containing 800 µl of bacterial culture and 200 µl of 50% glycerol were prepared. The stocks were stored at -80 ˚C.

2.7 Bacterial Strains

All strains used in this study are listed in Table 2.3.

Table 2.3: List of bacterial strains used in this study

Strain	Description	Reference/Source
E. coli Top10	F- *mcrA Δ(mrr-hsdRMS-mcrBC) Φ801acZM15 ΔlacX74 deoR recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (Str^R) endAl nupG*	Invitrogen, Ltd
E. coli NEB-stable	F' *proA+B+ lacI^q Δ(lacZ)M15 Tn10 (*Tet^R*)/Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- Φ80dlacZΔM15 recA1 endA1 nupG rpsL* (Str^R) *rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	NEB
P. thermoglucosidasius NCIMB 11955	Wild-type strain	TMO Renewables Ltd
P. thermoglucosidasius DSMZ 2542	Wild-type strain	DSMZ
P. thermoglucosidasius TM89	Δ*ldh*	TMO Renewables Ltd
P. thermoglucosidasius TM242	Δ*ldh, pdh^{up}*, Δ*pfl*	TMO Renewables Ltd
P. thermoglucosidasius LS242	Δ*ldh, pdh^{up}*, Δ*pfl*	Sheng *et al.*, 2017
*P. thermoglucosidasius*ΔpNCI001	pNCI001 megaplasmid deficient strain	Lau *et al.*, 2021
*P. thermoglucosidasius*ΔpNCI002	pNCI002 megaplasmid deficient strain	Lau *et al.*, 2021
P. thermoglucosidasius AM180	Δ*ldh, pdh^{up}*	This study

P. thermoglucosidasius AM236	Δ*ldh*, Δ*pfl*	This study
P. thermoglucosidasius AM242	Δ*ldh*, pdh^{up}, Δ*pfl*	This study
P. thermoglucosidasius TM242*::*ecGP123	Δ*ldh*, pdh^{up}, Δ*pfl*: 2005857-2006296 region swapped with eCGP123	This study
P. thermoglucosidasius AM242*::*ecGP123	Δ*ldh*, pdh^{up}, Δ*pfl*: 2005851-2006290 region swapped with eCGP123	This study
P. thermoglucosidasius TM242_A>G	Δ*ldh*, pdh^{up}, Δ*pfl*: SNP correction A>G	This study
P. thermoglucosidasius AM242_G>A	Δ*ldh*, pdh^{up}, Δ*pfl*: SNP creation G>A	This study
P. thermoglucosidasius KT1	P_{ldh}_mcr_{ca}; mcr_{ca} expressed constitutively on plasmid; Kan^R	This study
P. thermoglucosidasius KT2	P_{ldh}_mcr_{ms}-msr_{ss} ; mcr_{ms}-msr_{ss} operon expressed constitutively on plasmid; Kan^R	This study
P. thermoglucosidasius KT3	P_{ldh}_mcr_{ms}-msr_{ms} ; mcr_{ms}-msr_{ms} operon expressed constitutively on plasmid; Kan^R	This study
P. thermoglucosidasius KT4	P_{ldh}_mcr_{st}-msr_{ss} ; mcr_{st}-msr_{ss} operon expressed constitutively on plasmid; Kan^R	This study
P. thermoglucosidasius KT5	P_{ldh}_mcr_{st}-msr_{ms} ; mcr_{st}-msr_{ms} operon expressed constitutively on plasmid; Kan^R	This study
P. thermoglucosidasius Δ*ldhA*	lactate dehydrogenase deficient	Sheng e*t al.*, 2017
P. thermoglucosidasius KT401	P_{ldh}_mcr_{st}-msr_{ss} ; mcr_{st}-msr_{ss} operon expressed constitutively on plasmid in Δ*ldh* background; Kan^R	This study

P. thermoglucosidasius KT501	P_{ldh}_*mcr*$_{st}$-*msr*$_{ms}$; *mcr*$_{st}$-*msr*$_{ms}$ operon expressed constitutively on plasmid in Δ*ldh* background; Kan^R	This study
P. thermoglucosidasius Δ*mmsA2*	*mmsA2* deficient	Lili Sheng
P. thermoglucosidasius Δ*mmsA2*Δ*mmsA1*	*mmsA2* and *mmsA1* deficient	Lili Sheng
P. thermoglucosidasius::ilvC	Wild-type with *ilvC* integrated at *pyrE* locus	Lili Sheng
Lactooccus lactis	Wild-type Environmental isolate	Sarah Chapman
P. thermoglucosidasius AK01	P_{gapd}_*kivD*[LLKF(mut.)]; k*ivD*[LLKF(mut.)] expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AK02	P_{gapd}_*kivD*(LLKF_1386); k*ivD*(LLKF_1386) expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AK03	P_{gapd}_*kivD*; *kivD* expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AK04	P_{ldh}_*kivD*[LLKF(mut.)]; k*ivD*[LLKF(mut.)] expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AK05	P_{ldh}_*kivD*(LLKF_1386); k*ivD*(LLKF_1386) expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AK06	P_{ldh}_*kivD*; *kivD* expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AK07	P_{ldh}_*kivD*(LLM4); k*ivD*(LLM4) expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AK08	P_{ldh}_*kivD*(WT_LLM4); k*ivD*(WT_LLM4) expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AM001	P_{gapd}_*kivD*[LLKF(mut.)]_*alsS*$_{bs}$; k*ivD*[LLKF(mut.)]_*alsS*$_{bs}$ operon	This study

	expressed on autonomous plasmid; Kan^R	
P. thermoglucosidasius AM002	P_{gapd}_*kivD*[LLKF(mut.)]_$alsS_{bc}$; k*ivD*[LLKF(mut.)]_$alsS_{bc}$ operon expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AM003	P_{gapd}_*kivD*[LLKF(mut.)]_$alsS_{td}$; k*ivD*[LLKF(mut.)]_$alsS_{td}$ operon expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AM004	P_{gapd}_*kivD*(LLKF_1386)_$alsS_{bs}$; k*ivD*(LLKF_1386)_$alsS_{bs}$ operon expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AM005	P_{gapd}_*kivD*(LLKF_1386)_$alsS_{bc}$; k*ivD*(LLKF_1386)_$alsS_{bc}$ operon expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AM006	P_{gapd}_*kivD*(LLKF_1386)_$alsS_{td}$; k*ivD*(LLKF_1386)_$alsS_{td}$ operon expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AM007	P_{ldh}_*kivD*(LLKF_1386)_$alsS_{bs}$; k*ivD*(LLKF_1386)_$alsS_{bs}$ operon expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AM008	P_{ldh}_*kivD*(LLKF_1386)_$alsS_{bc}$; k*ivD*(LLKF_1386)_$alsS_{bc}$ operon expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AM009	P_{ldh}_*kivD*(LLKF_1386)_$alsS_{td}$; k*ivD*(LLKF_1386)_$alsS_{td}$ operon expressed on autonomous plasmid; Kan^R	This study
P. thermoglucosidasius AM007b	*ilvC* expressed in the chromosome at *pyrE* locus; P_{ldh}_*kivD*(LLKF_1386)_$alsS_{bs}$; *kivD*(LLKF_1386)	This study

	_*alsS*$_{bs}$ operon expressed on autonomous plasmid; Kan^{R}	
P. thermoglucosidasius AM008b	*ilvC* expressed in the chromosome at *pyrE* locus; P$_{ldh}$_*kivD*(LLKF_1386)_*alsS*$_{bc}$; *kivD*(LLKF_1386) _*alsS*$_{bc}$ operon expressed on autonomous plasmid; Kan^{R}	This study
P. thermoglucosidasius AM009b	*ilvC* expressed in the chromosome at *pyrE* locus; P$_{ldh}$_*kivD*(LLKF_1386)_*alsS*$_{td}$; *kivD*(LLKF_1386) _*alsS*$_{td}$ operon expressed on autonomous plasmid; Kan^{R}	This study
P. thermoglucosidasius AM010	*ilvC* expressed in the chromosome at *pyrE* locus; P$_{ldh}$_*kivD*(LLKF_1386)_*alsS*$_{bs}$_*ilvD*; *kivD*(LLKF _1386)_*alsS*$_{bs}$_*ilvD* operon expressed on autonomous plasmid; Kan^{R}	This study
P. thermoglucosidasius AM011	*ilvC* expressed in the chromosome at *pyrE* locus; P$_{ldh}$_*kivD*(LLKF_1386)_*alsS*$_{bc}$_*ilvD*; *kivD*(LLKF _1386)_*alsS*$_{bc}$_*ilvD* operon expressed on autonomous plasmid; Kan^{R}	This study
P. thermoglucosidasius AM012	*ilvC* expressed in the chromosome at *pyrE* locus; P$_{ldh}$_*kivD*(LLKF_1386)_*alsS*$_{td}$_*ilvD*; *kivD*(LLKF _1386)_*alsS*$_{td}$_*ilvD* operon expressed on autonomous plasmid; Kan^{R}	This study
P. thermoglucosidasius AM013	*ilvC* expressed in the chromosome at *pyrE* locus; P$_{ldh}$_*kivD*(LLKF_1386)_*alsS*$_{bs}$_*ilvD_adhA*; *kivD* LLKF_1386)_*alsS*$_{bs}$_*ilvD_adhA* operon expressed on autonomous plasmid; Kan^{R}	This study
P. thermoglucosidasius AM014	*ilvC* expressed in the chromosome at *pyrE* locus; P$_{ldh}$_*kivD*(LLKF_1386)_*alsS*$_{bc}$_*ilvD_adhA*; *kivD* LLKF_1386)_*alsS*$_{bc}$_*ilvD_adhA* operon expressed on autonomous plasmid; Kan^{R}	This study

P. thermoglucosidasius AM015	*ilvC* expressed in the chromosome at *pyrE* locus; P_{ldh}*_kivD*(LLKF_ 1386)*_alsS*$_{td}$*_ilvD_adhA*; *kivD* LLKF_1386)*_alsS*$_{td}$*_ilvD_adhA* operon expressed on autonomous plasmid; Kan^R	This study

2.8 Plasmids

All plasmids used or created in this study are listed in Table 2.4.

Table 2.4: List of plasmids used in this study

Plasmid	Description	Reference/Source
pMTL61321	*P. thermoglucosidasius* Shuttle vector pUB110.1, *aad9*, ColE1, MCS	This study
pMTL62321	*P. thermoglucosidasius* Shuttle vector pUB110.2, *aad9*, ColE1, MCS	This study
pMTL63321	*P. thermoglucosidasius* Shuttle vector pUB110.3, *aad9*, ColE1, MCS	This study
pMTL64321	*P. thermoglucosidasius* Shuttle vector pNCI001, *aad9*, ColE1, MCS	Spencer, 2018
pMTL65321	*P. thermoglucosidasius* Shuttle vector pNCI002, *aad9*, ColE1, MCS	Spencer, 2018
pMTL66331	*P. thermoglucosidasius* Shuttle vector pBST1, *aad9*, p15a, MCS	This study
pMTL66321	*P. thermoglucosidasius* Shuttle vector pBST1, *aad9*, ColE1, MCS	This study
pMTL67321	*P. thermoglucosidasius* Shuttle vector pGEOTH02, *aad9*, ColE1, MCS	This study
pTMO372	*G. stearothermophilus* DSMZ 6285 *pheB* based reporter vector. *pheB* flanked	TMO Renewables Ltd

	upstream by the *P. thermoglucosidasius idh* promoter. pBM1, repBST1, Kan^R, Amp^R	
pMTL_RbxE_P_{gapd}	pMTL61110_pldhRBS containing T1T2 terminator, P_{gapd}, Riboswitch E, sGFP gene	Lau, 2018
pUC57-Kan-Cas3-Pfl	Vector harbouring *pflB* gene deletion cassette	Azenta Life Science, Ltd
pUC57-Kan-Cas3-Pdh^{up}	Vector harbouring *pdhA* promoter upregulation cassette	Azenta Life Science, Ltd
pMTL-AM180	*pdhA* promoter upregulation vector	This study
pMTL-AM236	*pflB* gene knockout vector	This study
pMTL-AM731	TM242 & AM242 strains SNP region replacement vector with P_{gapd}_eCGP123 sequence	This study
pMTL-AM732	P_{gapd}_eCGP123 replacement vector with corrected SNP sequence of TM242 strain	This study
pMTL-AM733	P_{gapd}_eCGP123 replacement vector with created SNP sequence of AM242 strain	This study
pMTLgSlimS-ZV1b	Expression vector for bifunctional mcr_{ca} under P_{ldh} promoter	Bashir, 2018
pMTLgSlimS-ZV2b	Expression vector for mcr_{ms}-msr_{ss} operon under P_{ldh} promoter	Bashir, 2018
pMTLgSlimS-ZV3b	Expression vector for mcr_{ms}-msr_{ms} operon under P_{ldh} promoter	Bashir, 2018
pMTLgSlimS-ZV4b	Expression vector for mcr_{st}-msr_{ss} operon under P_{ldh} promoter	Bashir, 2018
pMTLgSlimS-ZV5b	Expression vector for mcr_{st}-msr_{ms} operon under P_{ldh} promoter	Bashir, 2018
pMTL-LLKF1	Expression vector for *kivD* (LLKF_1386) gene under P_{gapd} promoter	This study
pMTL-LLKF2	Expression vector for *kivD* [LLKF(mut.)] gene under P_{gapd} promoter	This study

pMTL-KIVD1	Expression vector for *kivD* gene under P_{gapd} promoter	This study
pMTL-LLKF3	Expression vector for *kivD* (LLKF_1386) gene under P_{ldh} promoter	This study
pMTL-LLKF4	Expression vector for *kivD* [LLKF(mut.)] gene under P_{ldh} promoter	This study
pMTL-KIVD2	Expression vector for *kivD* gene under P_{ldh} promoter	This study
pMTL-LLM4	Expression vector for *kivD* (LLM4) gene under P_{ldh} promoter	This study
pMTL-WT_LLM4	Expression vector for *kivD* (WT_LLM4) gene under P_{ldh} promoter	This study
pMTL-AM1	Expression vector for *kivD* [LLKF (mut.)] _$alsS_{bs}$ operon under P_{gapd} promoter	This study
pMTL-AM2	Expression vector for *kivD* [LLKF (mut.)]_ $alsS_{bc}$ operon under P_{gapd} promoter	This study
pMTL-AM3	Expression vector for *kivD* [LLKF (mut.)]_ $alsS_{td}$ operon under P_{gapd} promoter	This study
pMTL-AM4	Expression vector for *kivD* (LLKF_1386)_ $alsS_{bs}$ operon under P_{gapd} promoter	This study
pMTL-AM5	Expression vector for *kivD* (LLKF_1386)_ $alsS_{bc}$ operon under P_{gapd} promoter	This study
pMTL-AM6	Expression vector for *kivD* (LLKF_1386)_ $alsS_{td}$ operon under P_{gapd} promoter	This study
pMTL-AM7	Expression vector for *kivD* (LLKF_1386)_ $alsS_{bs}$ operon under P_{ldh} promoter	This study
pMTL-AM8	Expression vector for *kivD* (LLKF_1386)_ $alsS_{bc}$ operon under P_{ldh} promoter	This study
pMTL-AM9	Expression vector for *kivD* (LLKF_1386)_ $alsS_{td}$ operon under P_{ldh} promoter	This study
pMTL-AM10	Expression vector for *kivD* (LLKF_1386)_ $alsS_{bs}$_*ilvD* operon under P_{ldh} promoter	This study

pMTL-AM11	Expression vector for *kivD* (LLKF_1386)_ *alsS*$_{bc}$*_ilvD* operon under P$_{ldh}$ promoter	This study
pMTL-AM12	Expression vector for *kivD* (LLKF_1386)_ *alsS*$_{td}$*_ilvD* operon under P$_{ldh}$ promoter	This study
pMTL-AM13	Expression vector for *kivD* (LLKF_1386)_ *alsS*$_{bs}$*_ilvD_adhA* operon under P$_{ldh}$	This study
pMTL-AM14	Expression vector for *kivD* (LLKF_1386)_ *alsS*$_{bc}$*_ilvD_adhA* operon under P$_{ldh}$	This study
pMTL-AM15	Expression vector for *kivD* (LLKF_1386)_ *alsS*$_{td}$*_ilvD_adhA* operon under P$_{ldh}$ promoter	This study

2.9 DNA Manipulation

2.9.1 Extraction and purification of *P. thermoglucosidasius* chromosomal DNA

The chromosomal DNA of *P. thermoglucosidasius* was extracted and purified using the GenElute^{TM} Bacterial genomic DNA Kit (Sigma-Aldrich, UK) according to the manufacturer's instructions.

Phenol: chloroform DNA extraction was performed for high purity genomic DNA. Culture (2 ml) was pelleted, and the supernatant was discarded. The cell pellet was re-suspended in 180 µl lysozyme containing phosphate-buffered saline (PBS) and incubated at 37 ˚C for 30 min with periodic gentle agitation. 25 µl Proteinase K (Sigma), 85 µl ddH_2O and 110 µl 10 % w/v SDS solution were added to the mixture and mixed by inversion followed by incubation at 65 ˚C for 30 min. 20 µl RNase A (17,500 U) (Qiagen, Germany) was added to the solution and incubated for 10 min at room temperature. The solution was transferred to a phase-lock tube (VWR, USA). 400 µl Phenol: chloroform: isoamyl alcohol (Sigma-Aldrich) was added to the tube and

inverted to mix. 20 µl RNase A (17,500 U) (Qiagen, Germany) was added to the solution and incubated for a further 10 min at room temperature, followed by centrifugation at 16,000 x *g* for 5 min. The extraction was repeated twice by adding 400 µl phenol: chloroform: isoamyl alcohol to the top layer of the solution in the same tube, followed by centrifugation at 16,000 x *g* for 5 min. The top layer of the solution was transferred to an ice-cold Eppendorf tube containing 40 µl 3 M sodium acetate and 800 µl absolute ethanol, mixed thoroughly but gently by inversion and held at -80 ˚C for 30 min. The sample was centrifuged at 4 ˚C at full speed for 15 min. The supernatant was discarded, and the pellet was washed with 1 ml 70 % ethanol and centrifuged for 3 min at full speed. The supernatant was removed and further spun for 3 min before air drying the pellet for 45 min and re-suspending in 50 µl ddH_2O. The procedure was performed within a fume hood. The concentration of extracted DNA was measured using Qubit Flex Fluorometer (ThermoFisher Scientific, UK) and run on 1% agarose gel to confirm no shearing of the DNA.

2.9.2 Extraction and purification of plasmid DNA

Plasmid DNA from *E. coli* was extracted and purified using the Monarch® Plasmid Miniprep Kit (New England Biolabs, UK) according to the manufacturer's instructions.

2.9.3 Amplification of DNA fragments by polymerase chain reaction (PCR)

Amplification of DNA fragments was carried out using either Phusion High-Fidelity PCR Master Mix with HF Buffer (2X) (New England Biolabs, UK), Q5 High-Fidelity 2X Master Mix (New England Biolabs, UK), or DreamTaq Green PCR Master Mix (2X) (Thermo Scientific, UK).

According to the manufacturer's protocol, Phusion High-Fidelity PCR Master Mix with HF Buffer and Q5 High-Fidelity 2X Master Mix was used for all PCR reactions. The

reaction mixture consisted of (final concentration) 1X Phusion Master Mix or 1X Q5 High-Fidelity 2X Master Mix, 0.5 µM forward and reverse primers, and DNA template (1 pg-10 ng for plasmid DNA, and 50-250 ng or 1 ng-1 µg for genomic DNA). The reaction volume was made up to either 25 or 50 µl using nuclease-free water in a thin-walled PCR tube and placed in a Mastercycler nexus (Eppendorf, UK).

2.9.4 Screening of transformants by colony PCR

Colony PCR was used for screening purposes. Discrete colonies from agar plates were picked using sterile toothpicks and suspended in 5 µl nuclease-free water. Further, a 16 µl master solution containing appropriate quantities of DreamTaq Green PCR Master Mix (2X) (Thermo Scientific, UK), forward/reverse primers and water were added into the PCR tubes. The PCR reactions were performed following the manufacturer's instructions, in addition to a 6 min initial denaturation step to release the genomic DNA from the bacterial cells.

2.9.5 Amplification of DNA using SOEing PCR

Complete or partial genetic elements were joined together without restriction endonucleases by gene splicing by overlap extension (SOEing) PCR using primers with appropriate overlapping overhangs (Horton *et al.,* 1989).

2.9.6 Restriction endonuclease digestion of DNA

Restriction enzymes were sourced from either New England Biolabs (UK) or FastDigest™ enzymes from Thermo Scientific™ (UK) and used to cleave DNA sequences at specific locations following the manufacturers' instructions.

2.9.7 DNA analysis by agarose gel electrophoresis

Agarose gel electrophoresis was performed to visualise DNA and purification of restriction enzyme digestions and PCR products. Gels were prepared using 1% agarose (Sigma, UK) (1g in 100 mL) dissolved in 1X TAE buffer (40 mM Tris, 1 mM EDTA and 0.1% (v/v) glacial acetic acid). Ten microliters (10 µl) of 10,000X SYBR Safe stain (Invitrogen, UK) was added to the molten agarose and poured into the gel casting tray fitted with a comb to make wells and allowed to solidify. DNA reaction mixtures were obtained by mixing a DNA sample with 6X loading buffer (New England Biolabs, UK) in a 5:1 ratio before adding to the agarose gel wells covered with 1X TAE buffer in a gel tank. The DNA samples and the Quick-Load Purple 2-Log DNA ladder (New England Biolabs, UK) were separated by running the gel at 100-120V for 30-60 min. The separated DNA along the ladder was visualised using BioRad Imaging Documentation systems (BioRad, UK).

2.9.8 Agarose gel DNA extraction and purification

DNA fragments in agarose gels were excised under blue-light transilluminator (Biometra, UK), using a sterile scalpel. The DNA fragments were purified using a Monarch® DNA Gel Extraction Kit (New England Biolabs, UK), following the manufacturer's instructions.

2.9.9 Quantification of DNA

The concentrations of purified DNA was measured using a benchtop NanoDrop Lite Spectrophotometer (Thermo Scientific™, UK). A 1 µl DNA sample was loaded on the spectrophotometer, and absorption at 260 nm (A_{260}) was measured, allowing calculation of DNA concentration.

2.9.10 Ligation of DNA fragments

Digested and purified plasmid DNA fragments and insert fragments with compatible ends were ligated in a molar ratio of 1:3 for vector: insert. Ligation reactions were performed using either LigaFast™ Rapid DNA ligation Systems (Promega, UK) or T4 DNA ligase (Promega, UK) for overnight ligations following the manufacturers' instructions.

2.9.11 NEBuilder-HiFi DNA assembly

NEBuilder-HiFi DNA assembly was used for cloning multiple DNA fragments in a single step. Specific primers were designed using the NEBuilder Assembly Tool or the Assembly Wizard of Benchling software to amplify the DNA fragments to be assembled. A final volume of 20 µl reaction containing 10 µl of 2X NEBuilder-HiFi DNA Assembly Master Mix (New England Biolabs, UK) and purified PCR fragments were assembled and used according to the manufacturer's instructions.

2.9.12 Sequencing of DNA

PCR products and plasmid DNA after purification were sent for Sanger sequencing at Eurofins Genomics or Source Bioscience UK Limited, Nottingham, UK. Samples were prepared according to their specifications.

2.10 Transfer of Plasmid DNA into Bacterial Cells

2.10.1 Preparation of chemically competent *E. coli* cells

A 10 µl of *E. coli* cells was used to inoculate a 5 ml Luria-Bertani medium and incubated overnight at 37 °C with shaking at 200 rpm. The overnight culture (1 ml) was used to inoculate 100 ml Luria-Bertani medium in a 500 ml conical flask, followed by

incubation at 37 °C with shaking at 200 rpm until OD_{600} 0.3-0.4 (approximately 2-2.5 h). The culture was divided into two 50 ml falcon tubes and chilled on ice for 30 min, followed by centrifugation to harvest the cells at 3,500 x *g* for 10 min at 4 °C. The supernatant was discarded, and the pellet was resuspended in 10 ml chilled 100 mM $CaCl_2 \cdot 2H_2O$ and incubated on ice for 30 min and then centrifuged at 3000 x *g* for 10 min at 4 °C. The cell pellet was resuspended in 1.5 ml (per 100 ml) of 100 mM $CaCl_2 \cdot 2H_2O$ and incubated on ice (in a cold room) for 3 h. Finally, 1 ml of 50% (v/v) filter-sterilized glycerol was added to the cell suspension, and aliquots of 100 µl were transferred into pre-chilled Eppendorf tubes and stored at -80 °C.

2.10.2 Transformation of chemically competent *E. coli* cells using heat-shock

An aliquot of chemically competent *E. coli* cells (100 µl) was placed on ice to thaw, and the ligation reaction or plasmid was added to the cells, followed by incubation on ice for 30 min. The mixture was heat-shocked by incubation at 42 °C for 30 s and further incubated immediately on ice for 5 min. An aliquot (900 µl) of SOC medium was added to the cell suspension, and the cells were incubated at 37 °C with shaking at 200 rpm for 1 h. 100-200 µl of the cells were plated onto pre-warmed Luria-Bertani agar plates at 37 °C, supplemented with the appropriate antibiotic selection, and incubated at 37 °C overnight.

2.10.3 Preparation of electro-competent *P. thermoglucosidasius* cells

A 100 µl of *P. thermoglucosidasius* cells was plated on TSA and incubated overnight at 52 °C. Cells were taken from the TSA plate using a 10 µl loop to seed pre-warmed 50 ml 2SPYNG media at 52 °C in a 250 ml baffled conical flask and incubated at 52 °C with shaking at 250 rpm. The cells were incubated for 3-4.5 h to reach an OD_{600} 2.0-2.5. The cells were evenly transferred to two 50 ml pre-chilled falcon tubes, incubated

on ice for 10 min, and then centrifuged at 5000 x *g* for 15 min at 4 °C. The supernatant was discarded and the cells were then resuspended using 30 ml, 20 ml, 10 ml, and 10 ml pre-chilled electroporation buffer (EPB) (0.5 M sorbitol, 0.5 M mannitol, 10% v/v glycerol) and centrifuged after each resuspension at 5000 x *g* for 15 min at 4 °C. Finally, the cells were resuspended in 2 ml EPB. Aliquots of 60 µl of the electro-competent cells were transferred into pre-chilled Eppendorf tubes and stored at -80 °C.

2.10.4 Transformation of *P. thermoglucosidasius* by electroporation

Aliquots of electro-competent *P. thermoglucosidasius* cells (60 µl) were thawed on ice. Plasmid (100-500 ng/µl) was added to an aliquot of pre-thawed electro-competent cells, and the mixture was transferred to a pre-chilled electroporation cuvette (1 mm gap) and placed in a BioRad Genepulser electroporator with settings of 2500 V voltage, 10 µF capacitance, and 600 Ω resistance. Immediately after the pulse, 1 ml of pre-warmed 2-SPY medium at 52 °C was added to the cuvette, and the suspension transferred to a 50 ml falcon tube. The cells were then recovered at 52 °C, 250 rpm for 2-4.5 h. The cells were harvested by centrifugation at 5,000 x *g* for 5 min and resuspended in 150 µl of the 2-SPYNG medium. The cells were then plated onto pre-warmed TSA agar plates at 52 °C supplemented with the appropriate antibiotic and incubated at 52 °C overnight.

2.11 Gene knock-out in *P. thermoglucosidasius* using CRISPR/Cas9

Electro-competent *P. thermoglucosidasius* cells were transformed with 1 µg of the appropriate CRISPR/Cas9 plasmid as described in section 2.10.4. Transformants were spread on TSA plates containing 12.5 $\mu g \cdot ml^{-1}$ Kan and incubated overnight at 52 °C. Single colonies from these plates were picked and cultured in 10 ml 2SPYNG containing 12.5 $\mu g \cdot ml^{-1}$ Kan and 8 mM theophylline and incubated overnight at 52 °C to induce Cas9 production. The colonies were further cultured in

2SPYNG+Kan+theophylline at 52 °C for three passages (3 x 12 h), sequentially transferring the cells to fresh media each time to achieve clean knock-out mutants. Cultures were diluted to 10^{-5} and plated on TSA+Kan and then incubated overnight at 52 °C. Successful gene knock-outs were confirmed by colony-PCR screening as described in section 2.9.4.

2.12 Gene knock-in in *P. thermoglucosidasius* using CRISPR/Cas9

Electro-competent *P. thermoglucosidasius* cells were transformed with 1 µg of the appropriate CRISPR/Cas9 plasmid as described in section 2.10.4. Transformants were spread on TSA plates containing 12.5 µg·ml^{-1} Kan and incubated overnight at 52 °C. Single colonies from these plates were picked and cultured in 10 ml 2SPYNG containing 12.5 µg·ml^{-1} Kan and 8 mM theophylline and incubated overnight at 52 °C to induce Cas9 production. The colonies were further cultured in 2SPYNG+Kan +theophylline at 52 °C for three passages (3 x 12 h), sequentially transferring the cells to fresh media each time to obtain pure integrants. Cultures were diluted to 10^{-5} and plated on TSA+Kan and then incubated overnight at 52 °C. Successful gene integrations were confirmed by colony-PCR screening as described in section 2.9.4.

2.13 Plasmid Segregational Stability Determination in *P. thermoglucosidasius*

Plasmid segregational stability was assessed using a modified method previously described by Pennington (2006). *P. thermoglucosidasius* NCIMB 11955 were transformed with plasmids containing different replicons and selected on TSA plate supplemented with spectinomycin. The following day, single colonies were picked and inoculated at 52 °C for 16 h in 10 ml 2SPYNG with antibiotics. Then, 100 µl of the cultures were used to inoculate 10 ml 2SPYNG again with antibiotics at 52 °C for 12 h. This ensured maintenance of the plasmids while the cells attained the same state of

growth. From this point on, cells were inoculated as previously but into 10 ml of non-selective 2SPYNG for 12 h at 60 °C or 52 °C. The action was repeated for 72 h at every 12-h interval. After every 12 h of growth without selection pressure, serial dilutions were performed for each cultures from 10^{-1} – 10^{-6} in fresh 2SPYNG pre-warmed for 30 min at 52 °C. A 100 µl aliquot of each dilution was plated out on a non-selective TSA plate pre-dried for 1 h at 37 °C. On the following day, one hundred single colonies from the TSA plates were replica-plated using a sterile inoculation loop (1 µl) on TSA plates with and without antibiotics. Placing plates against a printed 100-grid aided the process of replica plating. After 24 h, all colonies were counted against the grid, and the percentage plasmid loss was calculated using the difference between the number of non-resistant and resistant colonies. Plasmid retained per generation was calculated with the equation $\sqrt[n]{R}$ and plasmid lost per generation as $1-\sqrt[n]{R}$, where n is the number of generations and R is the percentage of cell population retaining the plasmid.

2.14 Determination of Plasmid Copy Number

Plasmid copy numbers were estimated using real-time quantitative PCR (qPCR) by comparing the quantification signal from the plasmid to those from the chromosome, as described by Reeve *et al.* (2016). Primer pairs lepA1_F/lepA1_R were designed to target the chromosomal *lepA1* gene (encoding elongation factor 4), and pUB110_F/pUB110_R primers (detailed primer sequences in Table 2.1) targeting the pUB110.2 and pUB110.3 replicons of the plasmids. The predicted sizes of the amplified DNA fragments were 120 bp for *lepA1* and 117 bp for pUB1110.2 and pUB1110.3. Cultures of *P. thermoglucosidasius* harbouring plasmids containing the pUB1110.2 and pUB1110.3 replicons were grown in 2SPYNG in triplicate at 52 °C for 24 h. Genomic DNA extraction was performed using a DNeasy kit (Qiagen, Germany) according to the

manufacturer's instructions and used as the template in qPCR. PCR reactions were performed in a 20 µl reaction volume with LightCycler 480 II (Roche, Switzerland) using LuminoCt SYBR Green qPCR ReadyMix (Sigma, UK) according to the manufacturer's instructions. Cycle threshold (Ct) values were calculated automatically by the Lightcycler 480 software. The amplification efficiency of the primers was calculated from a 10-fold serial dilution of plasmid and genomic DNA. Amplification efficiency values for the chromosomal and plasmid amplicons were 1.9454 (*lepA1*) and 1.9905 (pUB110), respectively. Plasmid copy numbers per chromosome from triplicate experiments were calculated based on the equation:

Plasmid copy number= $(Ec^{Ctc})/(Ep^{Ctp})$

Ec and Ctc are the amplification efficiency and cycle threshold for the amplification from the chromosome. Ep and Ctp are the amplification efficiency and cycle threshold for the amplification from the plasmid.

2.15 Whole Genome Sequencing

2.15.1 Illumina sequencing of strains

Illumina sequencing of genomic DNA was performed by MicrobesNG (http://www.microbesng.uk), at the University of Birmingham, and Deep-Seq (Next Generation Sequencing Facility) (http://www.nottingham.ac.uk/deepseq), at the University of Nottingham. High throughput Illumina MiSeq using 2x250 bp paired-end reads was used. Raw data were used for mapping and variant calling.

2.15.2 Read mapping

Paired-end reads were mapped against the published *P. thermoglucosidasius* NCIMB 11955 genome (Sheng *et al.*, 2016) in CLC Genomics Workbench version 20.0.4 (Qiagen, DK) using the resequencing analysis tool.

The following parameters were used for mapping:

Match cost = 1

Mismatch cost = 2

Cost of insertions and deletions = linear gap cost

Insertion cost = 3

Deletion cost = 3

Insertion open cost = 6

Insertion extend cost = 1

Deletion open cost = 6

Deletion extend cost = 1

Length fraction = 0.5

Similarity fraction = 0.8

Global alignment = No

Auto-detect paired distances = Yes

Non-specific match handling = map randomly

Masking mode = no masking

2.15.3 Variant calling

CLC Genomics Workbench version 20.0.4 (Qiagen, DK) was used for Single Nucleotide Polymorphism (SNP) calling based on mapped reads (Section 2.14.2).

The following parameters were used for the SNP calling:

Ploidy = 1

Required variant probability (%) = 90.0

Minimum coverage = 10

Minimum count = 5

Minimum frequency (%) = 35.0

Read direction filter = Yes

Direction frequency (%) = 25.0

Relative read direction filter = Yes

Significance (%) = 1.0

Ignore positions with coverage above = 100,000

Restrict calling to target regions = not set

Ignore broken pairs = Yes

Ignore non-specific matches = reads

Minimum read length = 20

Base quality filter = No

Neighbourhood radius = 5

Minimum central quality = 20

Minimum neighbourhood quality = 15

Read position filter = No

Significance (%) = 1.0

Remove pyro-error variants = No

In homopolymer regions with minimum length = 3

With frequency below = 0.8

2.16 Detection of Metabolites

2.16.1 High-performance liquid chromatography analysis

Samples from fermentation experiments were centrifuged at 14,000 x *g* for 15 min, and cell-free supernatant used to quantify metabolites concentration using high-performance liquid chromatography (HPLC). The cell-free supernatant was mixed in a ratio of 1:1 with HPLC diluent (5 mM H_2SO_4 and 50 mM valeric acid). The mixture was then filtered through a 0.2 µm syringe filter (Whatman® Spartan®, GE Healthcare Life Sciences, UK) into HPLC vials (Restek, USA) containing 300 µL inserts with split caps. The samples were run on the Dionex UltiMate 3000 HPLC system (Thermo Scientific, UK) equipped with a 300 mm × 7.8 mm x 9 µm column Aminex® HPX-87H (Bio-Rad, UK) column set at 35 °C for 55 min at a flow rate of 0.5 mL/min, and monitored using a refractive index (RI) and Diode Array UV-Vis-detector.

2.17 Statistical Analysis

All data are plotted as mean ± SD and One-way ANOVA followed by Tukey's and Dunnett's multiple comparisons tests were performed using GraphPad Prism version 9.4.1 for Windows, GraphPad Software, San Diego, California, USA, www.graphpad.com.

CHAPTER 3

3.0 CHARACTERISATION OF *P. THERMOGLUCOSIDASIUS* MODULAR SHUTTLE VECTORS

3.1 Introduction

Thermophilic bacteria, including those belonging to the genus *Parageobacillus,* represent an attractive option as a chassis for producing chemicals and fuels from biomass derived feedstocks. The full exploitation of their potential would be facilitated by the availability of effective genetic tools for heterologous expression of biosynthetic genes and construction of strains producing different value-added products. In other Gram-positive chassis, such as Clostridia, this need has been met by the development of a standardised, pMTL80000 modular vector system (Heap *et al.,* 2009) that allowed the assembly of plasmids with standardised modules in multiple configurations. The system met the four desired specifications: being modular, rapid and facile, reversible, and extensible. Accordingly, the modular plasmid format has been adopted in *Cupriavidus necator* H16 with the publication (Ehsaan *et al.,* 2021) of the pMTL70000 vector system and a preliminary description of the basis of an equivalent system for *P. thermoglucosidasius* NCIMB 11955, the pMTL60000 plasmid series, has been described (Sheng *et al.,* 2017).

There have been a number of studies that describe parts and the development of genetic tools for *P. thermoglucosidasius*, including shuttle and expression vectors, recombination methods, and gene knock-out and knock-in techniques (Taylor *et al.,* 2008, Cripps *et al.,* 2009, Bartosiak-Jentys *et al.,* 2013, Lin *et al.,* 2014, Reeve *et al.,* 2016, Sheng *et al.,* 2017, Lau *et al.,* 2021). However, in all cases, the shuttle vectors made were all based on heterologous replicons. This included the study of Sheng *et al.* (2017), where all of the replicons used were based on the replication region of plasmid pUB110 derived from *Staphyloccous aureus*. Because the initial pMTL60000 vector series have been shown to work (Sheng *et al.,* 2017, Lau *et al.,* 2021), it was therefore of value to expand the series to include additional replicons native to thermophiles, add

additional application modules, undertake a full characterisaton of the vectors made and provide a web-based resource where they and their sequences can be sourced.

3.1.1 *The P. thermoglucosidasius* modular vector system

In keeping with the pMTL70000 and pMTL80000 plasmids, the pMTL60000 vector series is divided into four modules flanked by the unique 8 bp recognition sites of the type II restriction endonucleases *AscI*, *FseI*, *PmeI* and *SbfI* (Figure 3.1). The modules, arranged in the same order in all vectors, comprise a Gram-positive replicon (*AscI*/*FseI*), an antibiotic selection marker (*FseI*/*PmeI*), a Gram-negative replicon (*PmeI*/*SbfI*), and an application-specific module (*SbfI/AscI*). Respective modules were amplified using the PCR, digested with appropriate restriction endonucleases, and ligated to form a specific plasmid vector.

Figure 3.1: Schematic of pMTL60000 series modular plasmids

Each module was allocated a unique number, and their combination defines a specific vector's name (Table 3.1). A standardised nomenclature in the form of pMTL60000 is used, where the first 0 designates a Gram-positive replicon, allowing replication of the plasmid in its thermophilic Gram-positive hosts. The second 0 represents a positive selectable marker. The third 0 designates a Gram-negative replicon that supports plasmid maintenance in the *E. coli* shuttle host. The last 0 is an application-specific module, such as a multiple cloning site (MCS), allowing the insertion of additional genetic elements into the plasmid. Furthermore, some of the replicons include origins of transfer that support the mobilisation of the vector by conjugation. The base plasmid of the pMTL60000 series, pMTL61110, was made (Sheng *et al.*, 2017) using the

pUB110.1 replicon and the pUB110 *kan* (encoding kanamycin adenyltransferase) gene conferring resistance to kanamycin at temperatures no higher than 60 ˚C, as a selectable marker from pTMO31 (Cripps *et al.,* 2009). The pUB110.1 replicon and *kan* gene were both originally derived from the *S. aureus* cryptic plasmid pUB110. Two variants of the pUB110.1 replicon were generated to derive a more unstable plasmid, in which the 412-bp region preceding *repB* present in plasmid pMTL61110 was reduced to 362 and 189 bp. The plasmids generated were denoted pMTL62110 (pUB110.2) and pMTL63110 (pUB110.3), respectively (Sheng *et al.,* 2017). The Gram-negative replicon (ColE1 + *tra*) and an MCS were obtained from pMTL85151 (Heap *et al.,* 2009). Other plasmids in the series were then constructed by interchanging individual modules with the available parts using appropriate flanking restriction sites. The pUB110-derived replicons made by Sheng *et al.* (2017) were added to (Spencer, 2018) through the inclusion of those of the pNCI001 and pNCI002 megaplasmids of *P. thermoglucosidasius* and the replicon of plasmid pGEOTH02 from *P. thermoglucosidasius* C56-YS93. The replicon of the cryptic plasmid pBST1 from *B. stearothermophilus* was also added, but the vector made was not stored and is, therefore, no longer available, thus, needing reconstruction. Other modules included were the *aad9* selectable marker (provides resistance to spectinomycin), a *kan* mutant marker encoding a *KanHT* variant that is functional up to 70 ˚C (Liao and Kanikula, 1990), a Gram-negative replicon p15a and two fluorescent reporters sGFP (super folder green fluorescent protein) and eCGP123 (enhanced consensus green protein variant 123) (Lau *et al.,* 2021).

The components of the pMTL60000 modular system as available and added to in this study are listed in Table 3.1.

Table 3.1: Numbering scheme for pMTL60000 series modular plasmids

pMTL60000 Modular Series			
Gram +ve Rep	**Marker**	**Gram –ve Rep**	**Application-specific module**
0. Spacer	0. -	0. -	0. Spacer
1. pUB110.1	1. *kan*	1. ColE1+*oriT*	1. MCS
2. pUB110.2	2. *kanHT*	2. ColE1	2. P_{ldh} + MCS
3. pUB110.3	3. *aad9*	3. p15a	3. MCS + sGFP reporter
4. pNCI001		4. p15a+*oriT*	4. MCS + eCGP123 reporter
5. pNCI002			
6. pBST1			
7. pGEOTH02			

3.1.2 Chapter aims

This chapter involves adding to the envisaged pMTL60000 vector series by constructing a pMTL66331 modular vector based on the pBST1 replicon and thereafter undertaking its comparative characterisation in terms of transmissibility and segregational stability alongside the other available replicons. Having completed full characterisation of the available replicon options, the nucleotide sequences of system modules were made available in the vector generator facility at the web www.plasmidvectors.com to allow users to download the desired plasmid sequence.

3.2 Results

3.2.1 Construction of pMTL66331 modular shuttle vector for *P. thermoglucosidasius*

The replication region encompassing *repBST1*, originally from the cryptic *B. stearothermophilus* plasmid pBST1, was amplified using the PCR from pTMO372 (provided by TMO Renewables Ltd) employing BST1_F_AscI and BST1_R_FseI

primers (detailed primer sequences in Table 2.1). The amplified ~ 3 kb r*epBST1* DNA fragment was digested with *AscI* and *FseI* enzymes, gel purified and ligated between the *AscI*/*FseI* restriction sites of the modular vector pMTL67331 (Spencer, 2018), which contained the *aad9* gene (encoding spectinomycin 9-adenyltransferase), Gram-negative replicon p15a, and a multiple cloning site, yielding plasmid pMTL66331 (Figure 3.2).

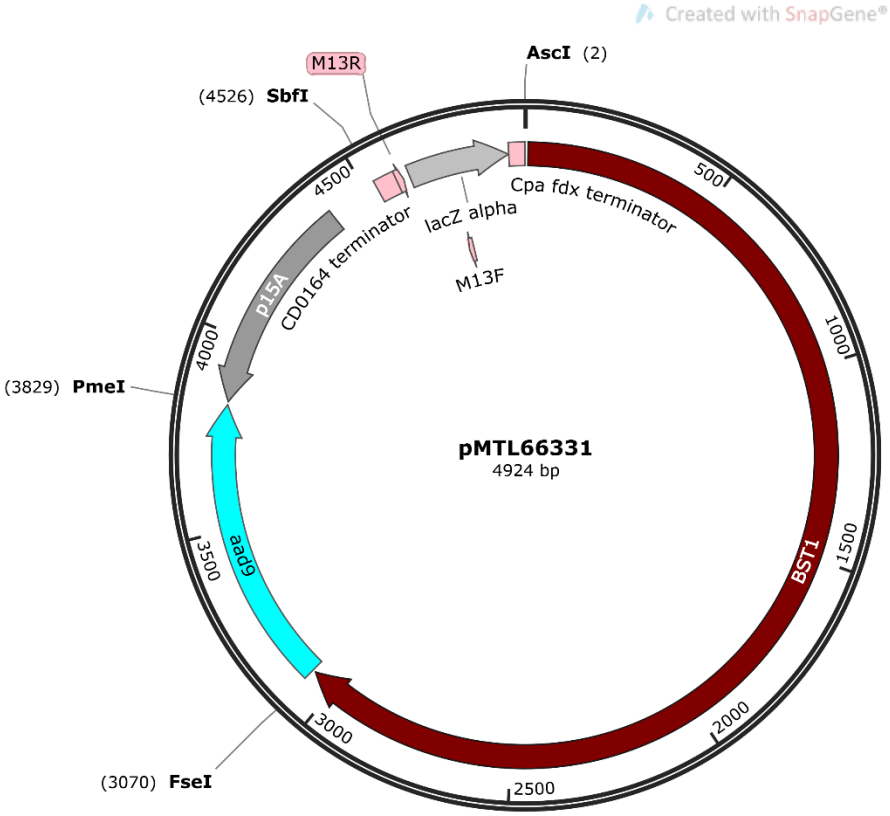


Figure 3.2: Vector map of pMTL66331. The vector consists of *repBST1* ligated at *AscI*/*FseI* restriction sites of the pMTL67331 plasmid backbone containing *aad9* gene (encoding spectinomycin 9-adenyltransferase), Gram-negative replicon p15a, and a multiple cloning site.

To select the transformed colonies from *E. coli* Top10 carrying the putative pMTL66331 plasmid, colony-PCR was carried out to screen for the presence of *repBST1* using BST1_F_AscI and BST1_R_FseI primers (detailed primer sequences in Table 2.1) as shown in Figure 3.3. Sanger sequencing confirmed the presence of the expected *repBST1* nucleotide sequences in the tested plasmids.

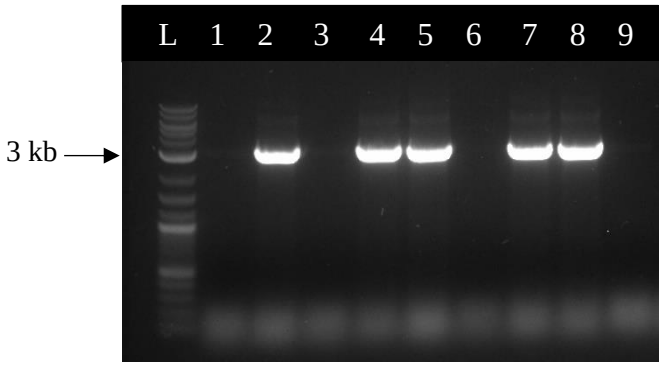


Figure 3.3: Colony-PCR screening of *repBST1* from *E. coli* Top10 transformed with pMTL66331 plasmid. L=A 2-log DNA Ladder (0.1-10 kb), lane 1=wild-type *E.coli* Top10 as negative control, lane 2=*repBST1* gene as positive control, lanes 3-9=screened *E. coli* Top10 transformants for *repBST1*. Colonies that had the *repBST1* in pMTL66331 yielded a band at ~ 3 kb.

3.2.2 Characterisation of the modular vectors

To investigate the suitability of the pMTL60000 modular plasmid vector series for strain engineering in *P. thermoglucosidasius* NCIMB 11955, the plasmids were characterised in terms of transformation efficiency and segregational stability.

3.2.2.1 Transformation efficiency of the modular plasmids

The modular vectors pMTL61321, pMTL62321, pMTL63321, pMTL66321 and pMTL67321 (Table 3.2) were transformed into wild-type *P. thermoglucosidasius*, and pMTL64321 and pMTL65321 (Table 3.2) into *P. thermoglucosidasius* ΔpNCI001 and *P. thermoglucosidasius* ΔpNCI002 (Lau *et al.*, 2021), respectively, as described in Methods section 2.10.4. Transformations were performed using 100 ng of plasmid DNA, added to a competent cells aliquot (60 μL). The time constants of the transformations were consistently between 5.1 – 5.4 ms. Cells were immediately added to pre-warmed soy peptone yeast extract (2SPY) and recovered at 52 °C at 250 rpm for 4 h. After recovery, the cells were plated onto TSA+Spectinomycin plates and left to grow overnight. Specifically, 100 μL of 1 x 10^{-1} dilution and an undiluted sample were grown for each transformation. The colonies of individual plates were then manually counted to determine the transformation efficiency of each sample, calculated by the number of colony-forming units per μg of DNA (CFUs per μg DNA). The transformation efficiencies of the various modular plasmids were in the order of 10^3–10^4 CFUs per μg DNA (Figure 3.4). The highest transformation efficiency of 2.4 x 10^4 CFU per μg DNA was observed with pMTL67321 and the lowest with pMTL62321 (8.0 x 10^3 CFU per μg DNA).

Table 3.2: Plasmids and their components used in comparing replicon properties

Plasmid	Gram +ve Rep	Marker	Gram –ve Rep	Application-specific
pMTL61321	1. pUB110.1	3. *aad9*	2. ColE1	1. MCS
pMTL62321	2. pUB110.2	3. *aad9*	2. ColE1	1. MCS
pMTL63321	3. pUB110.3	3. *aad9*	2. ColE1	1. MCS
pMTL64321	4. pNCI001	3. *aad9*	2. ColE1	1. MCS
pMTL65321	5. pNCI002	3. *aad9*	2. ColE1	1. MCS
pMTL66321	6. pBST1	3. *aad9*	2. ColE1	1. MCS
pMTL67321	7. pGEOTH02	3. *aad9*	2. ColE1	1. MCS

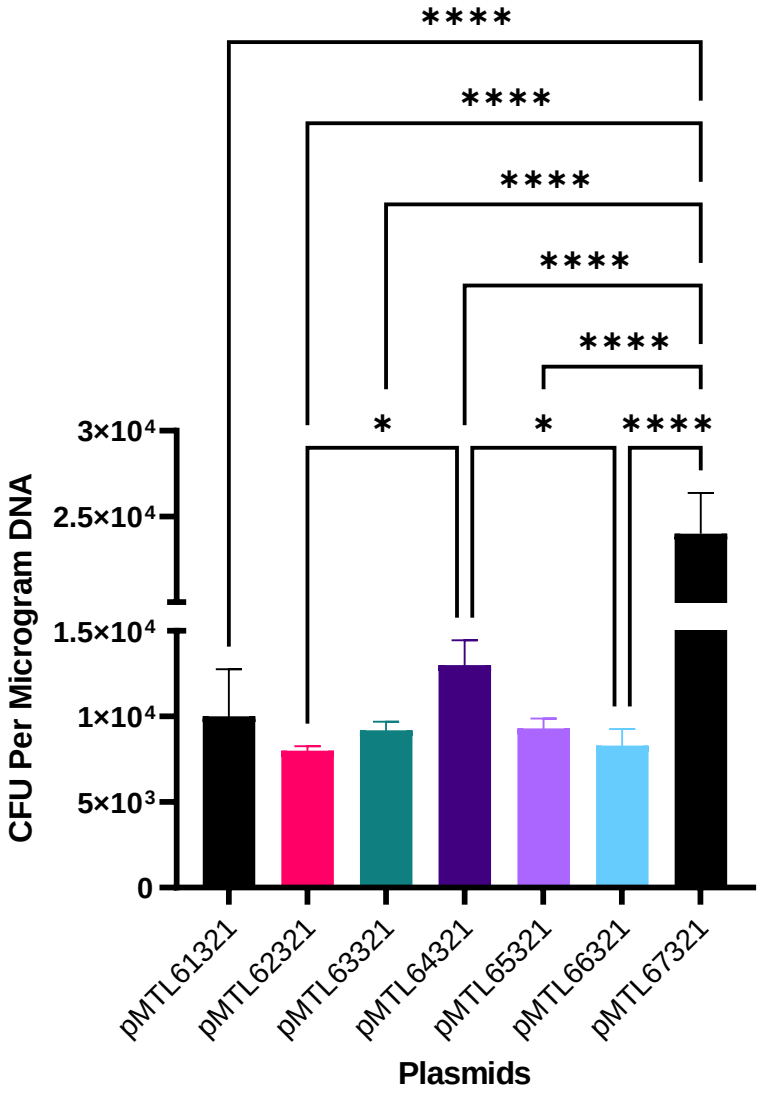


Figure 3.4: Transformation efficiency of *P. thermoglucosidasius* NCIMB 11955 with pMTL60000 modular plasmids. The hosts for plasmids pMTL64321 and pMTL65321 were *P. thermoglucosidasius* ΔpNCI001 and *P. thermoglucosidasius* ΔpNCI002, respectively. All other plasmids were transformed into the wild-type strain. Frequencies are expressed as CFU/μg DNA. Plasmids assessed and their replicons (in brackets) are pMTL61321 (pUB110.1), pMTL62321 (pUB110.2), pMTL63321 (pUB110.3), pMTL64321 (pNCI001), pMTL65321 (pNCI002), pMTL66321 (pBST1) and pMTL67321 (pGEOTH02). Results are shown as mean ± SD of three biological replicates (three cultures each derived from three different transformation events). *P***

≤ 0.01, *P***** < 0.0001 were calculated by One-way ANOVA followed by Tukey's multiple comparisons test.

3.2.2.2 Segregational stability of the modular plasmids

The segregational stability of the modular vectors in *P. thermoglucosidasius* was assessed by estimating the number of cells retaining resistance to plasmid-encoded spectinomycin after one to six, 12 h serial passages at 52 °C and 60 °C in 2SPYNG lacking antibiotic supplementation, as described in Methods section 2.13. The percentage of cells retaining the plasmid was estimated by plating appropriately diluted cell cultures on TSA media and replica plating 100 randomly picked single colonies on selective and non-selective plates, i.e., TSA media with and without spectinomycin (12.5 $\mu g \cdot ml^{-1}$). Figure 3.5 shows the percentage of cells retaining plasmids at 12 h intervals for 72 h at 52 °C, while Table 3.2 reveals the plasmid's retention and loss per generation. Plasmids based on the pUB110.1, pNCI001, and pGEOTH02 replicons exhibited high stability (Figure 3.5), with more than 90% of the cells retaining pMTL61321, pMTL64321, and pMTL67321 plasmids after 72 h and an average of 2.76 x 10^{-4} ± 3.25 ± 10^{-4}, 2.20 x 10^{-4} ± 2.52 ± 10^{-4}, and 5.84 x 10^{-4} ± 2.15 x 10^{-4} plasmid loss per generation (Table 3.2), respectively. In comparison, only 87% of the cells maintained pMTL65321 plasmids based on the pNCI002 replicon, after equivalent periods of time, with an average of 7.45 x 10^{-4} ± 7.34 ± 10^{-4} plasmid loss per generation. Plasmids pMTL62321 (pUB110.2), pMTL63321 (pUB110.3) and pMTL66321 (pBST1) showed less stability, with 100% of the cells losing the plasmid in 60 h and 72 h, respectively, with an average of 3.56 x 10^{-1} ± 4.99 x 10^{-1}, 3.74 x 10^{-1} ± 4.85 ± 10^{-1} and 3.88 x 10^{-2} ± 9.74 x 10^{-3} plasmid loss per generation, respectively (Table 3.2). The results of the plasmids based on the pUB110 replicons (pUB110.1, pUB110.2, and

pUB110.3) support previous findings (Sheng *et al.*, 2017). The pMTL62321 (pUB110.2), pMTL63321 (pUB110.3) and pMTL66321 (pBST1) could be utilised in applications where unstable plasmids are required, such as gene integration and knock-out.

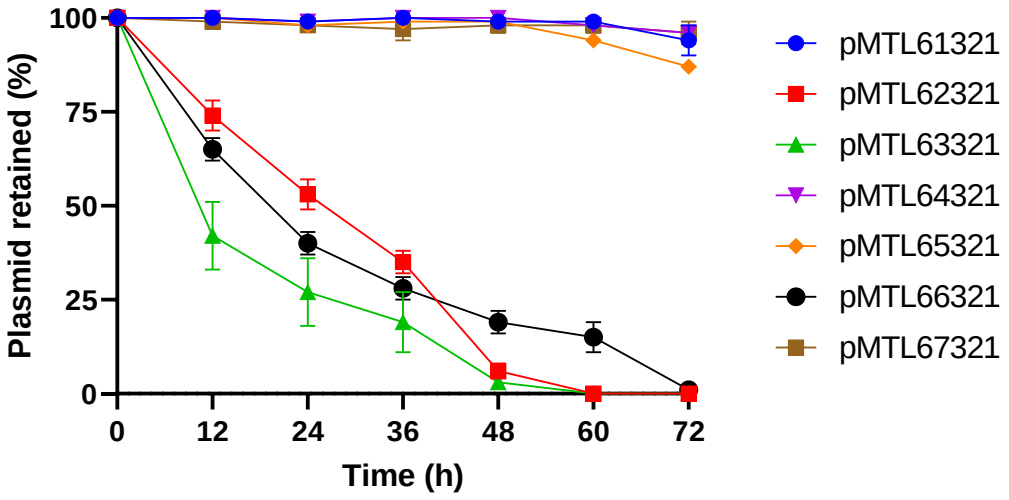


Figure 3.5: Percentage plasmid retention against time in non-selective media at 52 ˚C. The host of plasmids pMTL64321 and pMTL65321 were *P. thermoglucosidasius* ΔpNCI001 and *P. thermoglucosidasius* ΔpNCI002, respectively. The wild-type strain was the host for all the other plasmids. The percentage of plasmid retained was calculated using the difference between the number of non-resistant and resistant colonies. Plasmids assessed and their replicons (in brackets) are pMTL61321 (pUB110.1), pMTL62321 (pUB110.2), pMTL63321 (pUB110.3), pMTL64321 (pNCI001), pMTL65321 (pNCI002), pMTL66321 (pBST1) and pMTL67321 (pGEOTH02). Results are shown as mean ± SD of three biological replicates (three cultures each derived from three different transformation events).

Table 3.3: Plasmid segregational stability at 52 °C

Plasmid	Retention per generation	Loss per generation
pMTL61321	$0.999 \pm 3.25 \times 10^{-4}$	$2.76 \times 10^{-4} \pm 3.25 \pm 10^{-4}$
pMTL62321	$0.644 \pm 4.99 \times 10^{-1}$	$3.56 \times 10^{-1} \pm 4.99 \times 10^{-1}$
pMTL63321	$0.626 \pm 4.85 \times 10^{-1}$	$3.74 \times 10^{-1} \pm 4.85 \times 10^{-1}$
pMTL64321	$0.999 \pm 2.52 \times 10^{-4}$	$2.20 \times 10^{-4} \pm 2.52 \pm 10^{-4}$
pMTL65321	$0.999 \pm 7.34 \times 10^{-4}$	$7.45 \times 10^{-4} \pm 7.34 \times 10^{-4}$
pMTL66321	$0.961 \pm 9.74 \times 10^{-3}$	$3.88 \times 10^{-2} \pm 9.74 \times 10^{-3}$
pMTL67321	$0.999 \pm 2.15 \times 10^{-4}$	$5.84 \times 10^{-4} \pm 2.15 \times 10^{-4}$

Plasmid retained per generation was calculated with the equation $\sqrt[n]{R}$ and plasmid lost per generation as $1-\sqrt[n]{R}$ where n is the number of generations and R is the percentage of cell population retaining the plasmid.

The segregational stability of the modular plasmids at 60 °C was also determined using essentially the same procedure as used for 52 °C (Figure 3.6 and Table 3.3). Plasmids pMTL64321 (pNCI001), pMTL65321 (pNCI002), and pMTL67321 (pGEOTH02) were found to be highly stable (Figure 3.6), with 84%, 87%, and 91% of the cells maintaining the plasmids over 72 h and exhibiting an average of $1.51 \times 10^{-3} \pm 8.87 \pm 10^{-4}$, $1.27 \times 10^{-3} \pm 9.85 \times 10^{-4}$, and $6.83 \times 10^{-4} \pm 4.98 \times 10^{-4}$ plasmid loss per generation (Table 3.3), respectively. In contrast, only 2% of the cells retained pMTL66321 (pBST1) after the same period, with an average of $3.14 \times 10^{-2} \pm 1.13 \times 10^{-2}$ plasmid loss per generation. Cells harbouring pMTL61321 (pUB110.1), pMTL62321 (pUB110.2), and pMTL63321 (pUB110.3) lost 100% of the plasmids in just 12 h at 60 °C, suggesting that the pUB110 replicon is temperature sensitive and does not function at temperatures of 60 °C and above. This observation supports previous findings (Cripps

et al., 2009, Sheng *et al.*, 2017). This phenotype could prove helpful in plasmid loss selection.

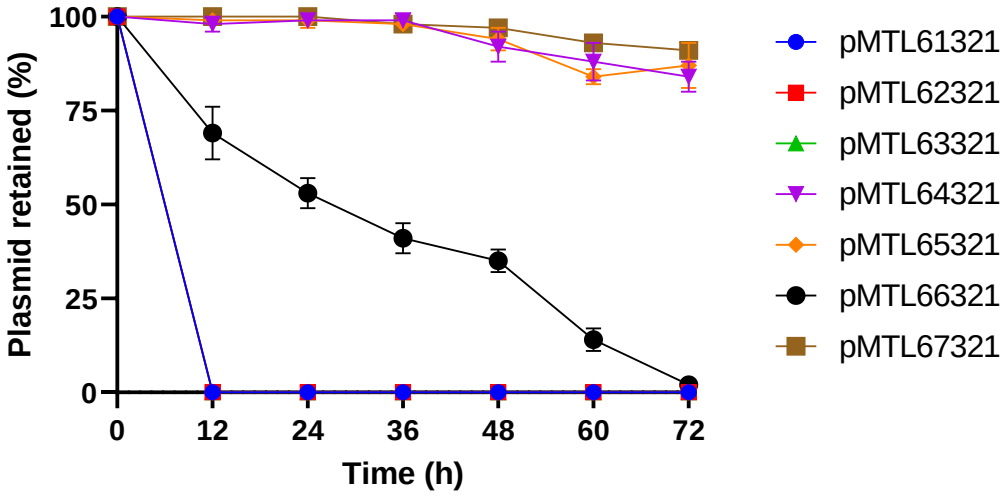


Figure 3.6: Percentage plasmid retention against time in non-selective media at 60 °C. The host of plasmids pMTL64321 and pMTL65321 were *P. thermoglucosidasius* ΔpNCI001 and *P. thermoglucosidasius* ΔpNCI002, respectively. The wild-type strain was the host for all the other plasmids. The percentage of plasmid retained was calculated using the difference between the number of non-resistant and resistant colonies. Plasmids assessed and their replicons (in brackets) are pMTL61321 (pUB110.1), pMTL62321 (pUB110.2), pMTL63321 (pUB110.3), pMTL64321 (pNCI001), pMTL65321 (pNCI002), pMTL66321 (pBST1) and pMTL67321 (pGEOTH02). Results are shown as mean ± SD of three biological replicates (three cultures each derived from three different transformation events).

Table 3.4: Plasmid Segregational Stability at 60 ˚C

Plasmid	Retention per generation	Loss per generation
pMTL61321	0	1
pMTL62321	0	1
pMTL63321	0	1
pMTL64321	$0.998 \pm 8.87 \times 10^{-4}$	$1.51 \times 10^{-3} \pm 8.87 \pm 10^{-4}$
pMTL65321	$0.999 \pm 9.85 \times 10^{-4}$	$1.27 \times 10^{-3} \pm 9.85 \times 10^{-4}$
pMTL66321	$0.969 \pm 1.13 \times 10^{-2}$	$3.14 \times 10^{-2} \pm 1.13 \times 10^{-2}$
pMTL67321	$0.999 \pm 4.98 \times 10^{-4}$	$6.83 \times 10^{-4} \pm 4.98 \times 10^{-4}$

Plasmid retained per generation was calculated with the equation $\sqrt[n]{R}$ and plasmid lost per generation as $1-\sqrt[n]{R}$ where n is the number of generations and R is the percentage of cell population retaining the plasmid.

3.2.2.3 Plasmid copy number

Spencer (2018) previously determined the plasmid copy numbers for all the replicons except pUB110.2 and pUB110.3, which are estimated in this study. Plasmid copy numbers were determined using qPCR by comparing the quantification signal from the plasmid to those from the chromosome as described in Methods section 2.14. Plasmid based on the pGEOTH02 replicon had the highest copy numbers of 297 ± 93.0 copies per chromosome, compared to only 36.99 ± 13.97 copies in the case of the pUB110.3 replicon (Figure 3.7).

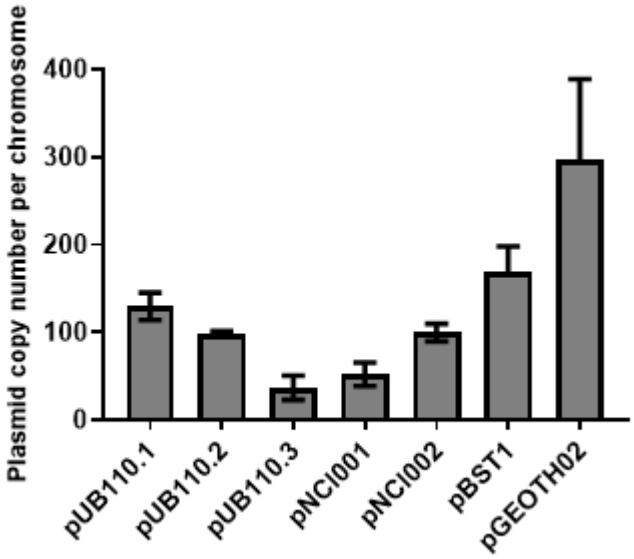


Figure 3.7: Estimation of plasmid copy number per chromosome. Data from Spencer (2018), except for pUB110.2 and pUB110.3 replicons determined in this study. The qPCR was used to estimate the relative copy number of plasmids containing the Gram-positive replicons. Results are shown as mean ± SD of triplicate experiments.

3.3 Discussion

The potential of *P. thermoglucosidasius* as a chassis for fuel and chemicals production would benefit from the availability of a greater selection of genetic tools for the construction of strains producing different metabolites. Work towards the development of a vector set composed of interchangeable, standardised modular parts was initiated by Sheng *et al*. (2017) and extended by Spencer (2018) to include a number of different thermophilic replicons. Here a further replicon (from *B. stearothermophilus*) was added to the series and all of the vectors (pMTL61321, pMTL62321, pMTL63321, pMTL64321, pMTL65321, pMTL66321, and pMTL67321) characterised in terms of efficiency of DNA transfer and segregational stability.

The transformation efficiency of all of the modular plasmids had transfer frequencies in the order of 10^3–10^4 CFUs per μg DNA. However, statistically significant differences ($P < 0.0001$) were observed, with the highest efficiency of 2.4 x 10^4 CFU per μg DNA obtained with pMTL67321 which was based on the pGEOTH02 replicon derived from *P. thermoglucosidasius* C56-YS93. Similar transformation efficiencies were reported by Reeve *et al.* (2016) in *P. thermoglucosidasius* with plasmids based on *repBST1* and *repB* replicons derived from *B. stearothermophilus* cryptic plasmid pBST1 and *S. aureus* cryptic plasmid pUB110, respectively.

In terms of the Gram-positive replicons included in the modular plasmid series to allow propagation in *P. thermoglucosidasius*, it was desirable to have several replicons exhibiting different properties such as transformation efficiency, copy number and segregational stability. Here the plasmids were tested for segregational stability at 52 °C and 60 °C over a 72 h period. Plasmids based on replicons pUB110.1, pNCI001, pNCI002, and pGEOTH02 were found to exhibit the highest segregational stability at 52 °C compared to those based on pUB110.2, pUB110.3 and pBST1. At 60 °C, plasmids based on the replicons of pNCI001, pNCI002 and pGEOTH02 were found to exhibit the highest stabilities. This high level of stability is consistent with the fact that the plasmids from which the replicons were derived were isolated from thermophiles. The pUB110 replicon is temperature-sensitive (Cripps *et al.*, 2009, Sheng *et al.*, 2017), and as observed in this study, cells harbouring plasmids containing the pUB110.1, pUB110.2, and pUB110.3 replicons, respectively, were able to maintained the plasmids at 52 °C and not 60 °C. This phenotype could prove useful in screening for plasmid loss in applications such as genome editing tools. Plasmids with high segregational stability will be more suited for gene expression purposes. Therefore, these highly stable replicons could be useful for gene expression studies in *P. thermoglucosidasius*, where

high protein production levels are desired. In contrast, low stability is desired for gene knock-out and knock-in, where plasmid curing is often a time-limiting step in genetic modification.

Plasmid copy number is an additional factor that can be important in maximising expression, as increased gene dosage invariably leads to high expression of cloned genes. These were previously determined for plasmids based on the various replicons by Spencer (2018). Plasmid copy numbers were estimated by real-time quantitative PCR (qPCR) using primers targeted at the chromosomal *lepA1* gene (encoding elongation factor 4) and the respective plasmid replicons. Plasmids based on the pGEOTH02, pBST1, pUB110.1 and pNCI002 replicons were shown to have the highest copy numbers of 297 ± 93.0, 168 ± 30.14, 130 ± 15.62, and 100 ± 10.0 copies per chromosome, respectively, compared to only 52 ± 13.65, 98.67 ± 2.70, and 36.99 ± 13.97 copies in the case of the plasmids based on the pNCI001, pUB110.2, and pUB110.3 replicons. On this basis, plasmids based on pGEOTH02 may represent the most appropriate to base an expression vector on.

The pMTL60000 plasmid series may be sourced from www.plasmidvectors.com. At this site is the facility to download the nucleotide sequence of any plasmid in *.gb format using the modular 'plasmid generator' function. This allows any of the interchangeable parts to be selected. Accordingly, the sequence of all of the modular parts were generated in *.gb format and uploaded to the web site.

3.4 Conclusions

In conclusion, additional pMTL60000 modular vector based on the pBST1 replicon was constructed. Further, the pMTL60000 modular vectors were characterised, showing differences in transformation efficiencies, segregational stability, and copy number in *P. thermoglucosidasius* NCIMB 11955. This will allow the selection of plasmids exhibiting desired properties for various applications. Plasmids and their sequences may be sourced from www.plasmidvectors.com.

CHAPTER 4

4.0 IMPLEMENTATION OF THEOPHYLLINE-RESPONSIVE CRISPR/CAS9 GENOME EDITING TOOL TO ENGINEER *P. THERMOGLUCOSIDASIUS* FOR BIOETHANOL PRODUCTION

4.1 Introduction

The ultimate aim of microbial metabolic engineering is to increase the production of desired metabolites, such as platform chemicals and biofuels. Metabolic engineering approaches generally involve expressing heterologous metabolic pathways either on plasmids or integrated into the chromosome or specific gene editing and redirecting carbon flux of native metabolic pathways (Mougiakos *et al.,* 2018). However, the effective engineering of microbial host strains is dependent on the availability of efficient genome-editing tools that could be used to achieve the desired changes to metabolic pathways *via* gene knock-out (KO) and knock-in (KI) (Sheng *et al.,* 2017). Despite the advantages of thermophiles, their use is hindered by inadequate gene-editing tools compared to mesophiles (Taylor *et al.,* 2011). Nonetheless, significant progress has been made in developing genetic tools for these group of bacteria. For example, a less laborious method for creating KO and KI that use the *pyrE* gene as both a positive and negative selection marker was described in *P. thermoglucosidasius* (Sheng *et al.,* 2017). However, the method does not allow for gene KIs independent of location or significantly reduce the timescale for mutant creation. A step forward has been the development and implementation of genome editing based on CRISPR in *P. thermoglucosidasius,* which is efficient and faster in mutant generation (Lau *et al.,* 2021).

4.1.2 CRISPR/Cas9 system

Clustered regularly interspaced short palindromic repeats and their associated proteins, commonly referred to as CRISPR and Cas, respectively, have been reviewed in detail in Section 1.5.2. CRISPRs were first observed in the genome of *E. coli* (Ishino *et al.,* 1987) and recognised to provide an adaptive bacterial immunity (Barrangou *et al.,*

2007). This system has been exploited in genome editing to manipulate the genome of microbes. Thus, an organism's genome is cleaved at a directed and precise location by transforming the Cas9 nuclease complex and a single guide RNA (sgRNA) into the organism. This allows the efficient KO and KI of specific genes at a defined location within an organism's genome. The genome editing is based on homologous recombination replacement of a wild-type sequence with the desired mutant allele, followed by eradicating the parent cell population via Cas9 cleavage guided by sgRNA. Therefore, it follows that in the perfect system, a firmly controlled promoter ought to be utilised to forestall the production of Cas9 until the allelic exchange required to produce the wanted mutant cell population has occurred. Likewise, limiting the time during which Cas9 is available decreases the chance for potential off-target effects and the probability of mutations that inactivate either the sgRNA or Cas9 (Cañadas *et al*., 2019). Recently, a genome-editing tool based on CRISPR/Cas9 was developed in our group for use in *P. thermoglucosidasius* NCIMB 11955 (Lau *et al.,* 2021). The system used a constitutive promoter to control the production of Cas9. Even though the method is efficient and speeds up mutant generation, it has been associated with low transformation and/or editing efficiencies. Therefore, one solution to increase the transformation efficiency and/or editing efficiencies would be to implement an inducible genome-editing system based on theophylline-responsive CRISPR/Cas9. Such an inducible CRISPR/Cas9 genome editing tool based on theophylline-responsive riboswitches was developed and successfully implemented in Clostridia to overcome these challenges (Cañadas *et al*., 2019).

4.1.3 Theophylline-responsive CRISPR/Cas9 genome editing tool (RiboCas)

Riboswitches are mRNA-encoded genetic control elements that regulate the expression of genes in an effector molecule-dependent style without the requirement of proteins

(Winkler and Breaker, 2005; Cui *et al*., 2016). These regulatory elements consist of an aptamer domain that recognises and allows binding of the effector molecule and an expression platform, whose conformation changes in response to ligand binding to the aptamer to regulate the initiation of translation or termination of transcription (Tucker and Breaker, 2005;Cui *et al*., 2016). Consequently, an mRNA that consists of a riboswitch is directly associated with modulating its activity in response to the concentration of its effector molecule.

The synthetic riboswitches utilised in this study are activated by theophylline molecule, which bears structural similarity with caffeine, and has been shown to work successfully in diverse bacterial species (Topp *et al.*, 2010; Cui *et al.*, 2016). In the absence of theophylline, a stem-loop structure is formed by the riboswitch, which sequesters the ribosome binding site in the mRNA transcript; hence, there is no gene expression. On the other hand, in the presence of theophylline, it binds to the aptamer resulting in changes in the riboswitch conformation, thereby releasing the ribosome binding site and thus, gene expression is observed (Tucker and Breaker, 2005).

The general genome editing process based on the inducible theophylline-responsive CRISPR/Cas9 genome editing tool (RiboCas) is illustrated in Figure 4.1.

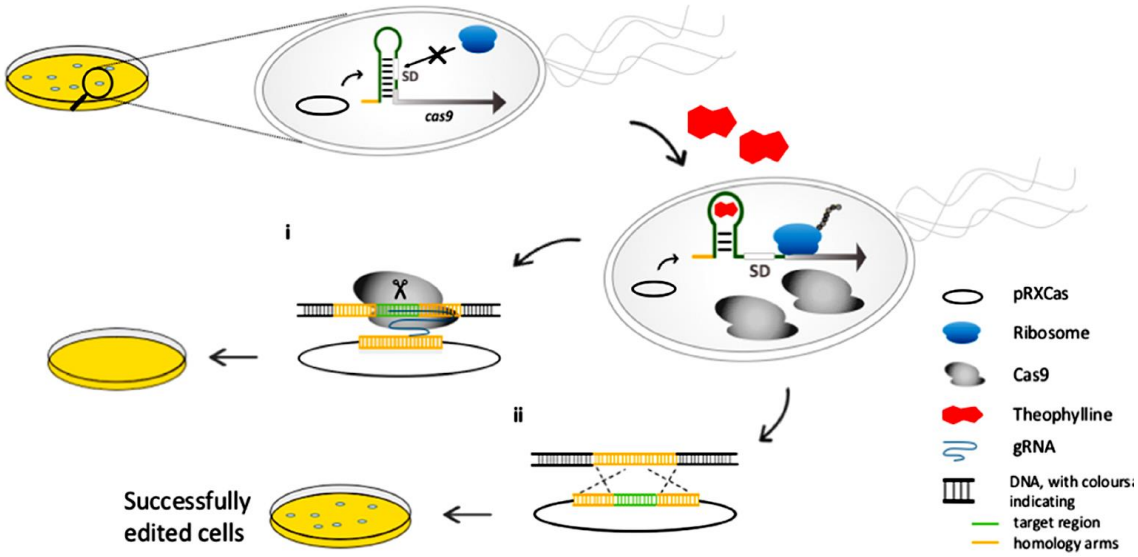


Figure 4.1: Schematic illustration of the theophylline-responsive CRISPR/Cas9 mediated genome editing. Without induction, i.e., in the absence of theophylline, transformed cells survive on selective media due to tight repression applied by the riboswitch, which blocks the translation of the Cas9 nuclease. Whereas, in the presence of theophylline and after induction, a Cas9-sgRNA complex is formed between the translated Cas9 and sgRNA, thereby introducing a double-stranded break on the target DNA. This results in two outcomes; **(i)** the unedited cells are killed during the process due to the lethal effect of the Cas9-sgRNA complex, and **(ii)** cells survive only if homologous recombination occurs between the gene-targeting plasmid and the genome. Adapted from Cañadas *et al.* (2019).

This novel theophylline-responsive CRISPR/Cas9 genome editing tool has been successfully applied to facilitate genome editing in several species of Clostridia, including the pathogens *C. botulinum* and *C. difficile*, the non-pathogen *C. sporogenes*, and the solventogenic *C. pasteurianum* with 100% mutant generation efficiency (Cañadas *et al*., 2019).

Preliminary studies in *P. thermoglucosidasius* NCIMB 11955 in SBRC Nottingham by Lau (2018) led to the testing of the effects of theophylline on the growth of *P. thermoglucosidasius* and the subsequent development of a riboswitch mediated inducible expression system. The growth of *P. thermoglucosidasius* was essentially unaffected up to a concentration of 10 mM theophylline. Furthermore, the relative strengths of three different riboswitches, termed E, F, and G, to drive the expression of the super-folder green fluorescent protein (sGFP) reporter gene under the control of P_{gapd}, P_{fbpa}, and P_{pfk} promoters were investigated. It was observed that all the three riboswitches, E, F and G, showed a significant increase in expression when induced with 8 mM theophylline under both promoters compared to the constitutive promoters alone. However, riboswitch G seems to have the strongest induction, followed by F and then E, both under the control of the P_{gapd} promoter. Furthermore, the riboswitch systems are leaky, with the least observed in E, followed by F and G (Lau, 2018). This data suggest that the system can be used as part of a genome-editing tool in *P. thermoglucosidasius* for genetic engineering.

4.1.4 Previous studies on TM242 equivalent recreation using allele-coupled exchange (ACE)

Sheng *et al.* (2017) developed the ACE technique of genome editing in *P. thermoglucosidasius* based on *pyrE* as a counter-selection marker. To validate the method's usefulness, an industrial ethanol production strain of TMO Renewables–TM242 (Δ*ldh, pdh*^{up}, Δ*pfl*) (Cripps *et al.*, 2009) was recreated from *P. thermoglucosidasius* NCIMB 11955, the parental strain. However, the production of ethanol by the created strain, LS242 (Δ*ldh, pdh*^{up}, Δ*pfl*), differed significantly from that of TM242, producing only 20-30 mM ethanol compared to 180-200 mM. But the addition of 0.1% acetic acid to the growth medium under micro-erobic conditions resulted in the LS242 having an ethanol yield similar to that of TM242. Comparisons of the whole genome sequences of LS242 and TM242 strains revealed single-nucleotide polymorphisms (SNPs) and insertions and deletions (Indels) occurring at every modification step. While 11 mutations occurred during the construction of TM242 by Cripps *et al.* (2009), the LS242 strain carried additional four mutations compared to the parent strain *P. thermoglucosidasius* NCIMB 11955 (Sheng *et al.*, 2017). Therefore, it was hypothesised that LS242 might not be producing high ethanol due to the SNPs/Indels it acquired. Alternatively, TM242 may be producing higher ethanol yields due to the acquisition of SNPs/Indels. This project aimed to answer this question by re-engineering strains equivalent to TM242 and LS242 using the theophylline-responsive CRISPR/Cas9 genome editing method based on the "If I understand it, I can build it" paradigm of synthetic biology.

4.1.5 Chapter aims

The main aim of this chapter was to implement the novel theophylline-responsive CRISPR/Cas9 genome editing tool (RiboCas) to metabolically engineer *P. thermoglucosidasius* NCIMB 11955 for ethanol production. In doing so, multiple independent strains equivalent to *P. thermoglucosidasius* TM242 were systematically re-engineered to investigate why this strain is capable of efficient ethanol production. Characterisation of the engineered strains in terms of ethanol yields was explored followed by genome analysis for unintended mutations. Ethanol production using barley straw hydrolysate was also performed.

4.2 Results

Thirty-six, independent strains equivalent to previously constructed ethanol production strains, *P. thermoglucosidasius* TM242 (Δ*ldh, pdh*^{up}, Δ*pfl*) (Cripps *et al.*, 2009) and LS242 (Δ*ldh, pdh*^{up}, Δ*pfl*) (Sheng *et al.*, 2017) were engineered to understand why the TM242 produces ethanol close to theoretical maximum. These were generated by KO of the *pflB* (pyruvate formate lyase) gene together with upregulation the of *pdhA* (pyruvate dehydrogenase) gene (Figure 4.2) by placing *pdhA* under the control of the P$_{ldh}$ promoter from *G. stearothermophilus* for overexpression. *P. thermoglucosidasius* TM89 (Δ*ldh*) was choosen as the starting strain (which TMO claimed was a forerunner of TM242) and making six independent versions of the first KO (that means making a CRISPR vector, doing six independent transformations, and choosing one clone from each experiment - that way they could not be siblings). This was followed by making six independent second gene KO in each of the six first stage KOs – thereby creating thirty-six independent strains. Following engineering and characterisation, whole-genome sequencing of the strains was performed for genome analysis.

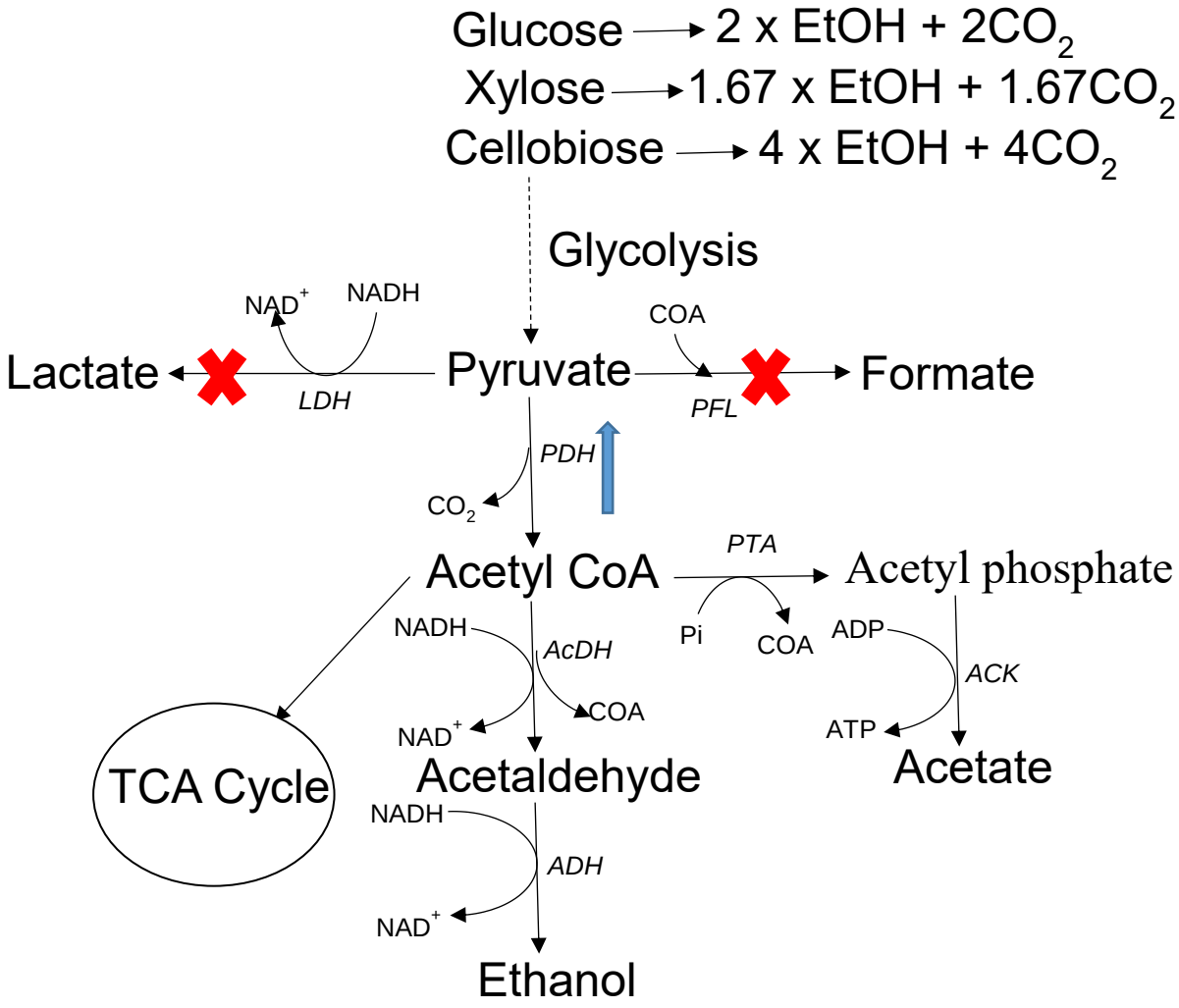


Figure 4.2: Schematic diagram of the modified mixed acid fermentation in *P. thermoglucosidasius* NCIMB 11955. Abbreviations are; *LDH*, lactate dehydrogenase; *PFL*, pyruvate formate lyase; *PDH*, pyruvate dehydrogenase; *PTA*, phosphotransacetylase; *ACK*, acetate kinase; *AcDH*, acetaldehyde dehydrogenase; *ADH*, alcohol dehydrogenase; TCA cycle, tricarboxylic acid cycle, EtOH; ethanol. *LDH* and *PFL* gene knockouts are indicated with red crosses, and the up-regulation of the *PDH* by the blue arrow. Theoretical maximum: Glucose: 2 moles of ethanol and 2 moles CO_2; Xylose: 1.67 moles of ethanol and 1.67 moles CO_2; Cellobiose: 4 moles of ethanol and 4 moles CO_2.

4.2.1 Construction of pMTL-AM180 and pMTL-AM236 plasmids

The vector pMTL-RbxE-P_{gapd} previously constructed by Lau (2018) was used as the backbone, which contained pMTL61110-pldhRBS made up of *Streptococcus thermophilus* Ca9-3 (*stCas9-3*), P_{gapd}, Riboswitch E, sGFP gene, and T1T2 terminator. The application-specific modules for both *pdhA* upregulation and *pflB* deletion were synthesised and located between BamHI and AscI restriction sites of a pUC57-Kan vector by Azenta Life Sciences Ltd (formerly GENEWIZ, Ltd), yielding plasmids pUC57-Kan-Cas3-pdh^{up} and pUC57-Kan-Cas3-pfl, respectively. Up-regulation of *pdhA* involved replacing its native 337 bp promoter fragment with the *G. stearothermophilus* 165 bp *ldhA* promoter, P_{ldh}. Therefore, for *pdhA* upregulation, the module contained sgRNA targeting P_{pdhA} (identified using Benchling CRISPR Guide Design software (www.benchling.com), T1T2 terminator, P_{ldh} from *G. stearothermophilus* flanked by an editing template comprising homology arms (450 bps) from up and downstream of the P_{pdhA}. On the other hand, the module for *pflB* deletion contained sgRNA targeting *pflB* gene (identified using Benchling CRISPR Guide Design software (www.benchling.com), T1T2 terminator, and an editing template comprising fused left and right homology arms (450 bps), corresponding to the regions up and downstream of the *pflB* gene.

Plasmids pMTL-RbxE-P_{gapd} and pUC57-Kan-Cas3-pdh^{up} containing the P_{pdhA} promoter replacement cassette with P_{ldh} were digested with BamHI and AscI and gel purified. The P_{pdhA} replacement cassette was ligated between the BamHI/AscI restriction recognition sites of the vector pMTL-RbxE-P_{gapd}, yielding plasmid pMTL-AM180 (Figure 4.3 A). The same procedure was carried out to construct plasmid pMTL-AM236

using the pMTL-RbxE-P_{gapd} vector backbone and pUC57-Kan-Cas3-pfl containing *pflB* deletion cassette, respectively (Figure 3.3 B).

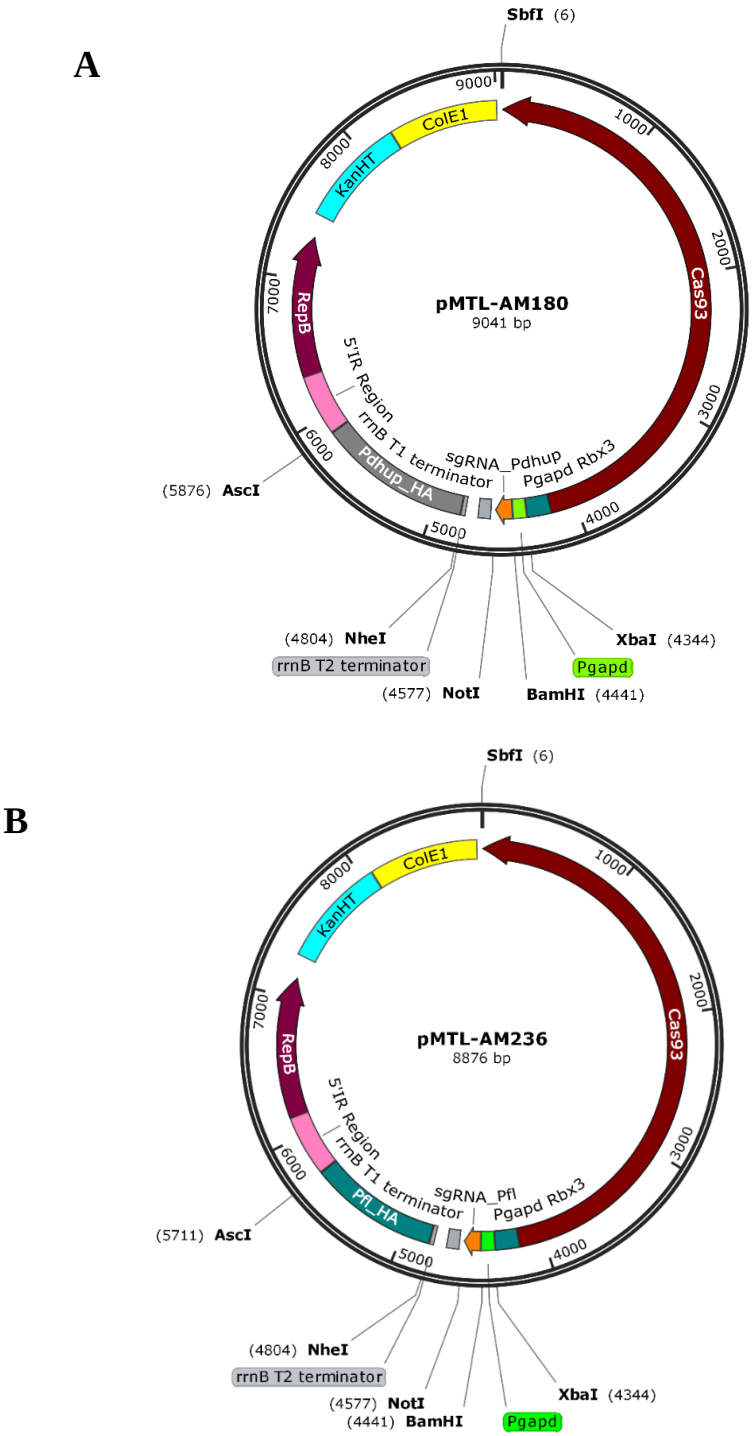


Figure 4.3: Vector map of plasmids. (A) pMTL-AM180 (B) pMTL-AM236. The maps consist of RepB, *kanHT*, ColE1, *stCas9-3* and P_{pdhA} promoter replacement cassette

(A) and *pflB* deletion cassette **(B)**, respectively, ligated between BamHI and AscI sites in the pMTL-RbxE-P_{gapd} vector.

Five transformed *E. coli* Top10 colonies were grown overnight to screen for the correct constructs, and then pMTL-AM180 and pMTL-AM236 plasmids were extracted. Purified pMTL-AM180 and pMTL-AM236 plasmids were screened by restriction digestion using BamHI and AscI. Additionally, Sanger sequencing was performed to confirm the presence of the correct nucleotide sequence. All the five pMTL-AM180 candidates screened contained the P_{pdhA} promoter replacement cassette (~1.4 kb). Similarly, all five pMTL-AM236 candidates screened were shown to contain the ~1.2 kb *pflB* deletion cassette (Figure 4.4).

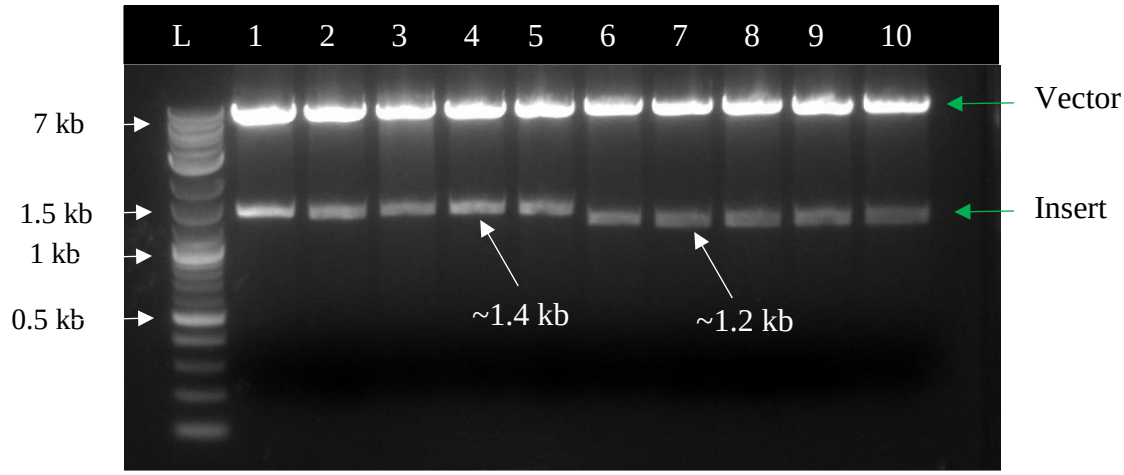


Figure 4.4: Screening of pMTL-AM180 and pMTL-AM236 plasmids from transformed *E. coli* Top10 by restriction digestion. L= DNA marker (0.1-10 kb); lanes 1-5 represent digested pMTL-AM180 plasmids with BamHI and AscI from five transformed *E. coli* Top10; lanes 6-10 illustrate digested pMTL-AM236 plasmids with BamHI and AscI from five transformed *E. coli* Top10. Expected band sizes of ~1.4 kb and ~1.2 kb correspond to P_{pdhA} promoter replacement and *pflB* deletion cassettes, respectively, while ~7.6 kb corresponds to the pMTL-RbxE-P_{gapd} backbone.

4.2.2 Engineering strains for bioethanol production

In implementing the theophylline-responsive CRISPR/Cas9 genome editing tool, the equivalent of the industrial bioethanol production strain TM242 (Cripps *et al.*, 2009) and LS242 (Sheng *et al.*, 2017) were enginereed from TM89 (Δ*ldh*), the progenitor strain of TM242. Stepwise, the creation of a *pdhA* (encoding pyruvate dehydrogenase) promoter up-regulation (replacement of its promoter with P_{ldhA} [encoding lactate dehydrogenase] from *G. stearothermophilus*) was first carried out in the Δ*ldh* background (Δ*ldh*, pdh^{up}), followed by the deletion of *pflB* (encoding pyruvate formate lyase) gene (Δ*ldh*, pdh^{up}, Δ*pfl*).

Because the CRISPR/Cas9 genome editing method allowed the precise deletion of genes (Lau *et al.*, 2021), the *pflB* gene was inactivated by deleting the entire coding region, similar to that in LS242, in which the *ldhA* and *pflB* genes were disrupted by creating clean in-frame deletions, leaving only the start and stop codons intact (Sheng *et al.*, 2017). This differs from the strategy used in the generation of TM242, where both *ldhA* and *pflB* genes were only centrally disrupted via the introduction of a *NotI* site (Cripps *et al.*, 2009).

4.2.2.1 Pyruvate dehydrogenase (*pdhA*) promoter replacement

After confirming the correct nucleotide sequence, the plasmid pMTL-AM180 harbouring the P_{pdhA} replacement with P_{ldhA} cassette was transformed into six different batches of electro-competent cells of *P. thermoglucosidasius* TM89 (Δ*ldh*) (strain obtained from TMO Renewables) by electroporation as described in the Methods section 2.10.4. After overnight incubation, colonies were picked from each batch of transformations, cultured in 10 ml 2SPYNG+Kan+8 mM theophylline, and incubated overnight at 52 $^{\circ}$C to induce Cas9 production. The overnight cultures were diluted to 1

x 10^{-5} using 2SPYNG and plated on TSA+Kan plates and then incubated overnight at 52 ˚C to obtain discrete colonies. Colony PCR was performed using Cas3_Pdh^{up}_F and Cas3_Pdh^{up}_R primers (primer sequences are shown in Table 2.1) to determine whether the P_{pdhA} replacement with the P_{ldhA} was successful. It was observed that the *pdhA* upregulation was successful; however, the colonies were composed of a mixture of mutant and wild-type, i.e., not pure. The colonies were further cultured in 10 ml 2SPYNG+Kan+8 mM theophylline and incubated at 52 ˚C for 12 h x 3 passages to obtain clean mutants. The cultures were diluted to 1 x 10^{-5} and plated on TSA+Kan plates and then incubated overnight at 52 ˚C and screened for possible clean mutants. Eight individual colonies from each batch of transformations were then screened by colony PCR using Cas3_Pdh^{up}_F and Cas3_Pdh^{up}_R primers (primer sequences are shown in Table 2.1) and ran on an agarose gel to identify clean mutants (Figure 4.5).

A high success rate was observed for batches 3 and 4, with 6/8 screened colonies showing successful promoter replacement. On the other hand, 5/8 and 4/8 of the screened colonies showed successful promoter replacement in batches 1, 2, 6, and 5, respectively. Sanger sequencing of the PCR products confirmed the desired replacement of the P_{pdhA} with the P_{ldhA} promoter. One clone was selected from each batch, cultured in 10 ml 2SPYNG and incubated at 60 ˚C for 12 h x 6 passages for plasmid loss. Colonies with successful plasmid loss were screened again for mutants by colony PCR.

A total of six independent, double mutant strains were created and designated as *P. thermoglucosidasius* AM180_1, AM180_2, AM180_3, AM180_4, AM180_5 and AM180_6 (Δ*ldh, pdh^{up}*) strains.

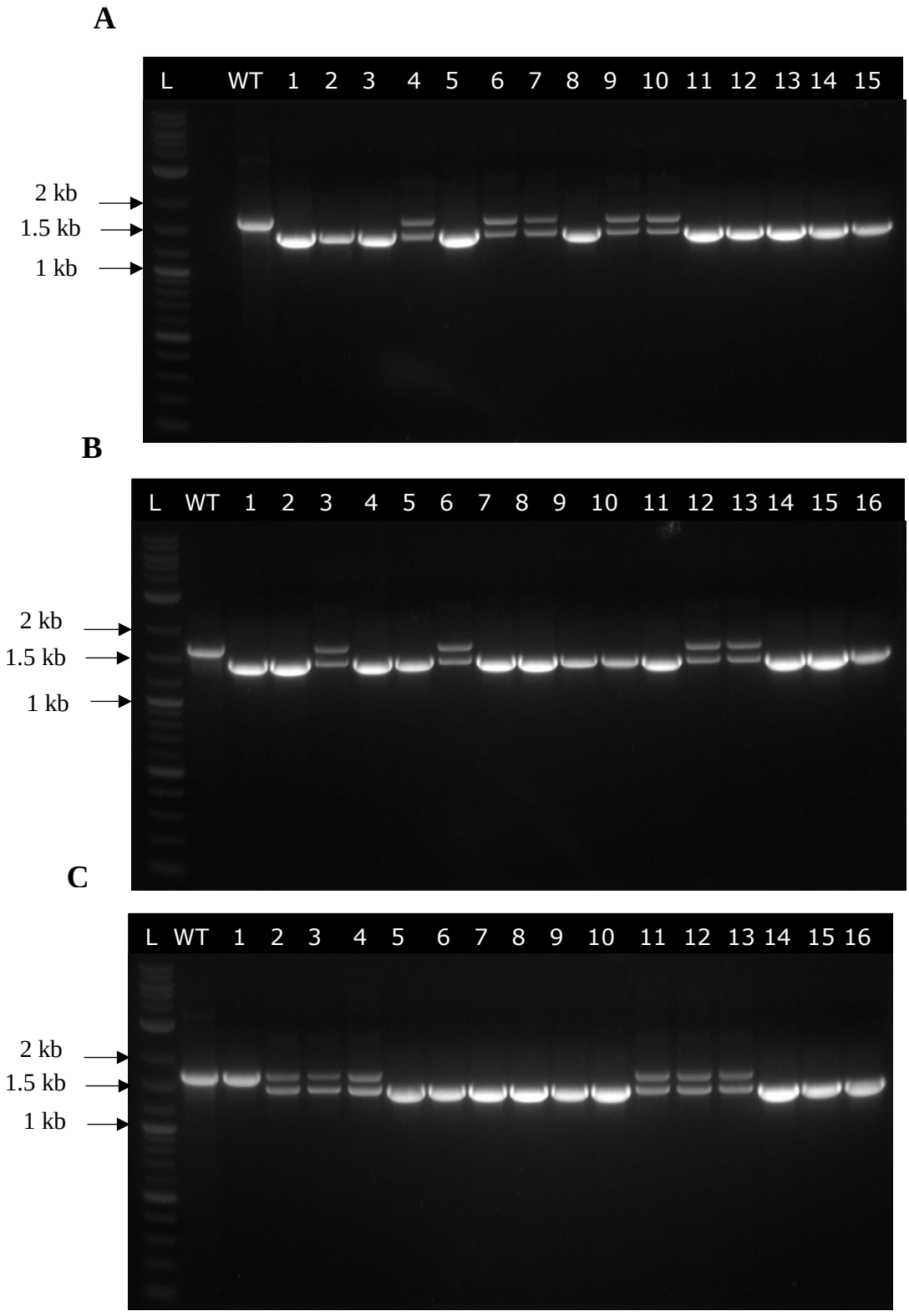


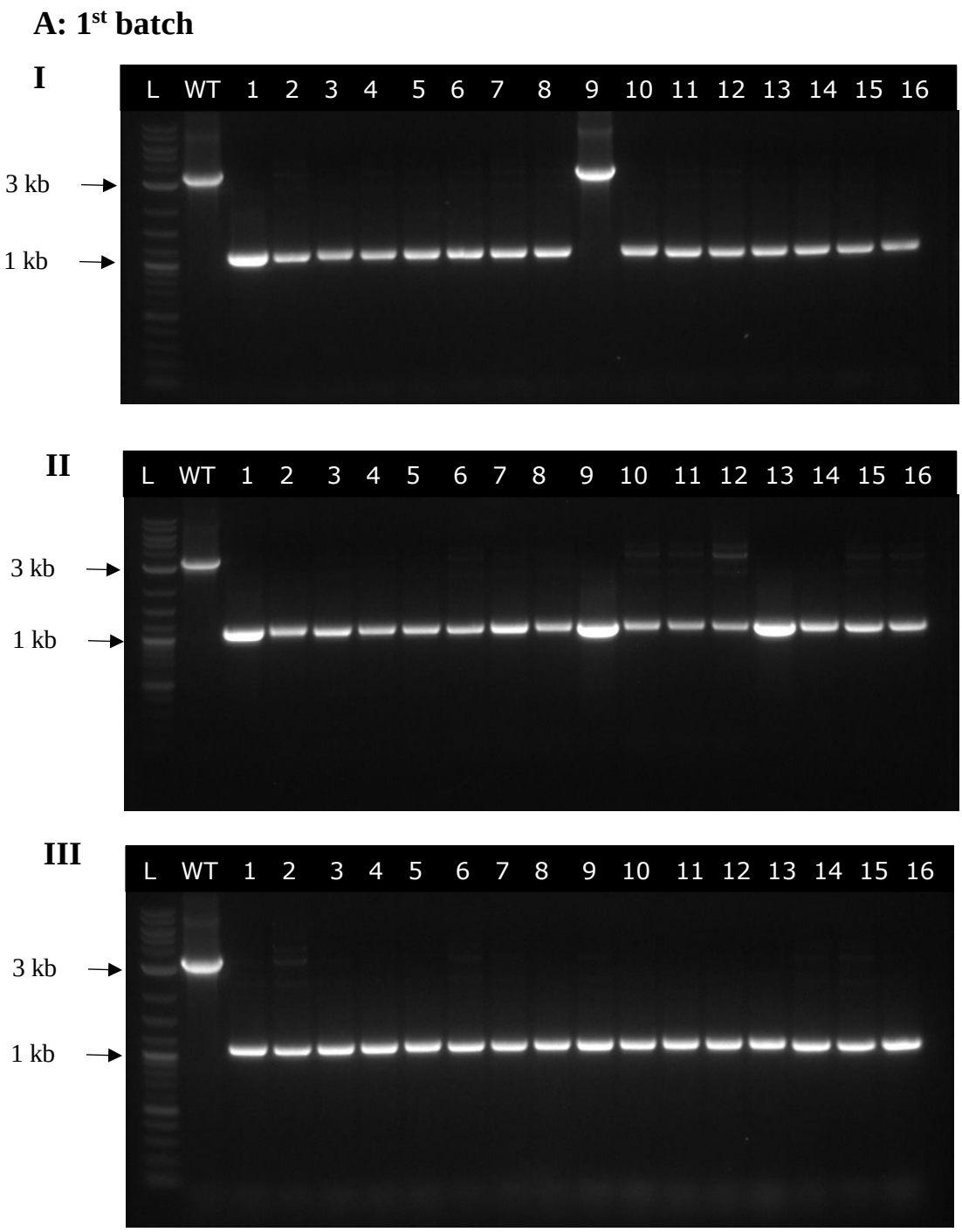
Figure 4.5: Screening of mutants for P_{pdhA} replacement with P_{ldhA} promoter using Cas3_Pdhup_F and Cas3_Pdhup_R primers. The wild-type (WT) generated a ~ 1.7 kb DNA fragment, whereas the DNA template from a Δ*ldh*, pdh^{up} mutant generated a ~

1.5 kb fragment. WT: PCR performed with *P. thermoglucosidasius* NCIMB 11955 wild-type. L = NEB 2-log DNA Ladder (0.1 -10 kb); **A**: Lanes 1-8 = 1st batch, and lanes 9-15 = 2nd batch; **B**: Lanes 1-8 = 3rd batch, and lanes 9-16 = 4th batch; **C**: Lanes 1-8 = 5th batch, and lanes 9-16 = 6th batch.

4.2.2.2 Pyruvate formate lyase (*pflB*) gene deletion in *P. thermoglucosidasius* AM180 strain (Δ*ldh, pdh*^{up})

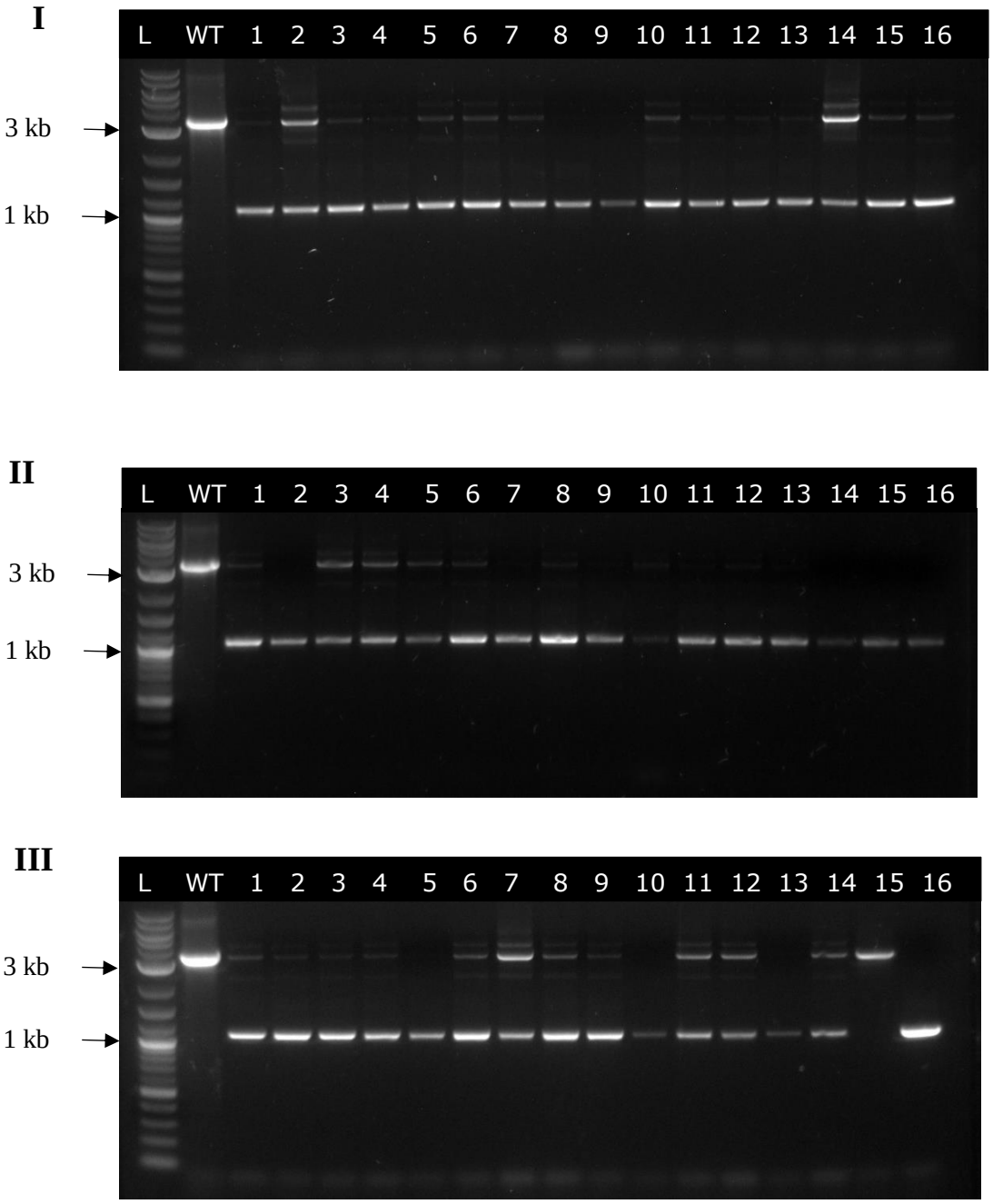
To achieve KO of the *pflB* gene, plasmid pMTL-AM236 containing the *pflB* deletion cassette was transformed into six different batches of electro-competent cells of *P. thermoglucosidasius* AM180_1, AM180_2, AM180_3, AM180_4, AM180_5, and AM180_6 strains by electroporation as described in Methods section 2.10.4. After overnight incubation, colonies were picked and cultured in 10 ml 2SPYNG+Kan+ 8 mM theophylline and incubated overnight at 52 °C to induce Cas9 production. After 3 x 12-h passages, the cultures were screened for clean mutants. The cultures were first diluted to 1 x 10^{-5} and plated on TSA+Kan plates and then incubated overnight at 52 °C. Eight individual colonies from each batch of transformations (one to six) of AM180_1, AM180_2, AM180_3, AM180_4, AM180_5, and AM180_6 were then screened by colony PCR using Cas3_Pfl_F and Cas3_Pfl_R primers (primer sequences are shown in Table 2.1) and electrophoresed on an agarose gel, to identify clean mutants (Figure 3.6). Sanger sequencing of the PCR products confirmed the desired deletion of the *pflB* gene. One clone was selected from each batch, cultured in 10 ml 2SPYNG and incubated at 60 °C for 12 h x 6 passages for plasmid loss. Colonies with successful plasmid loss were screened again for mutants by colony PCR using Cas3_Pfl_F and Cas3_Pfl_R primers (primer sequences are shown in Table 2.1).

A total of thirty-six independent, triple mutant strains, which are equivalent to TM242 and LS242 (Δ*ldh, pdh*^{up}, Δ*pfl*) strains, were enginereed and designated as *P. thermoglucosidasius* AM242 (Δ*ldh, pdh*^{up}, Δ*pfl*) strains.



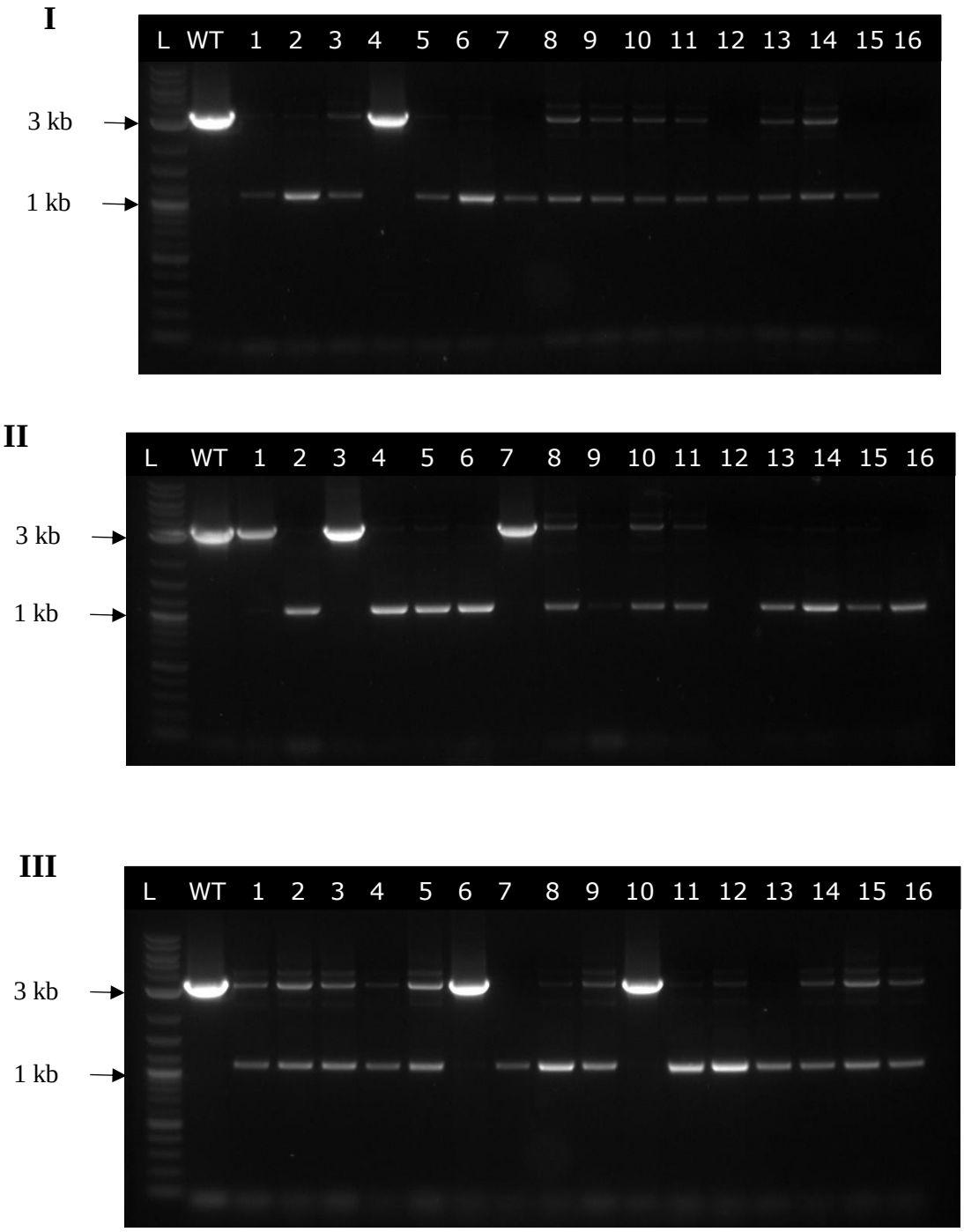
For legend see figure 4.6, p. 120-121

B: 2nd batch



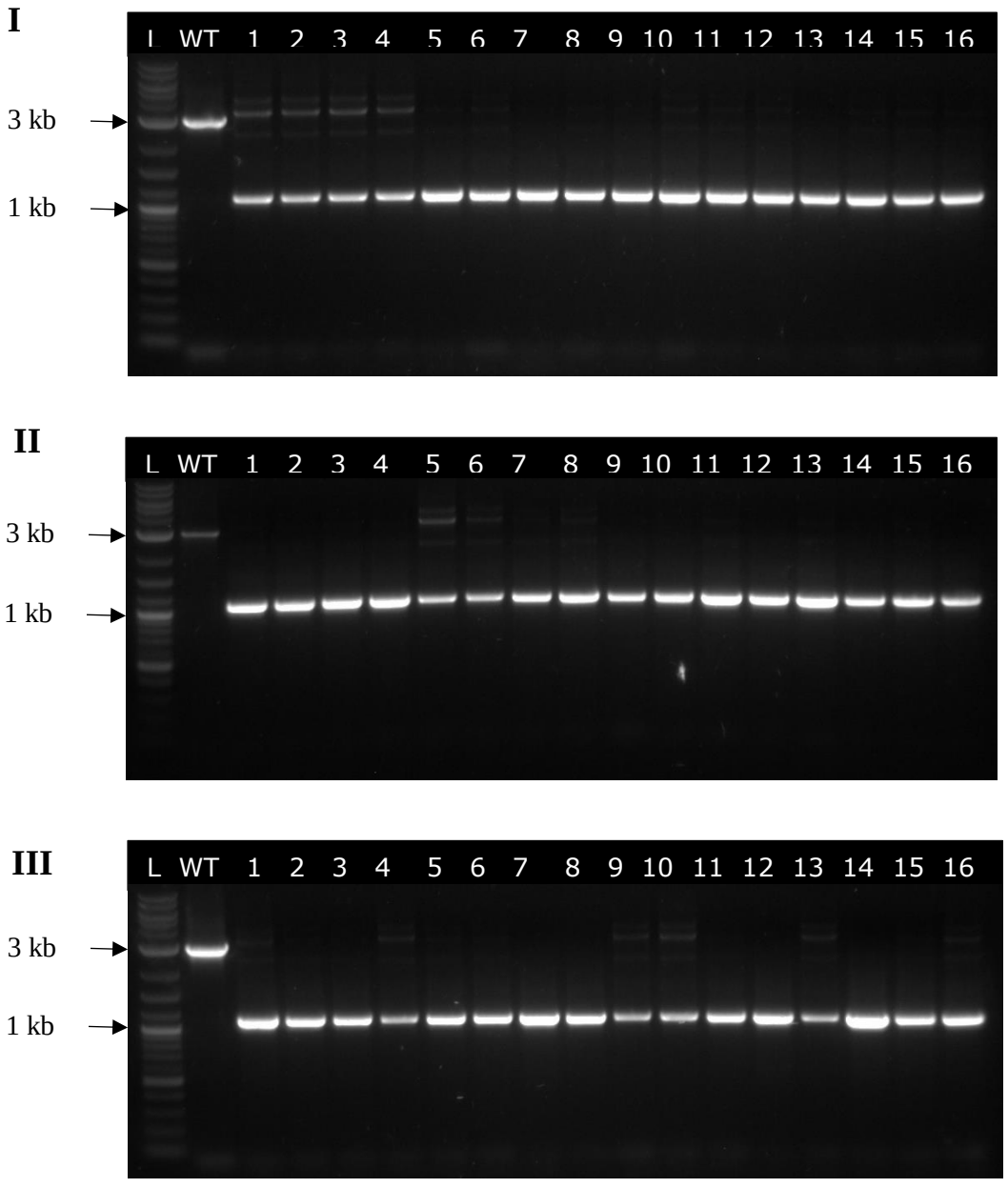
For legend see figure 4.6, p. 120-121

C: 3rd batch



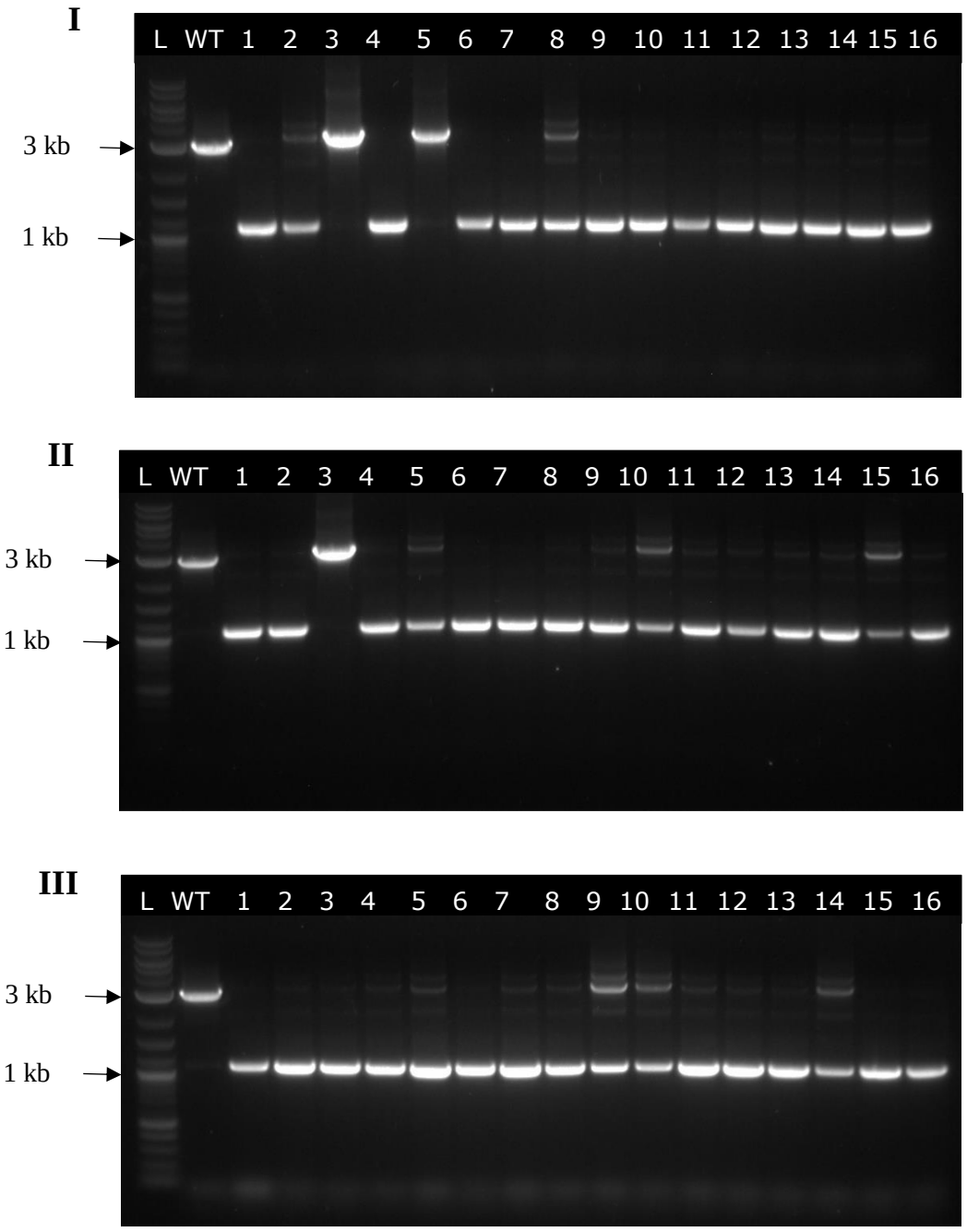
For legend see figure 4.6, p. 120-121

D: 4th batch



For legend see figure 4.6, p. 120-121

E: 5th batch



For legend see figure 4.6, p. 120-121

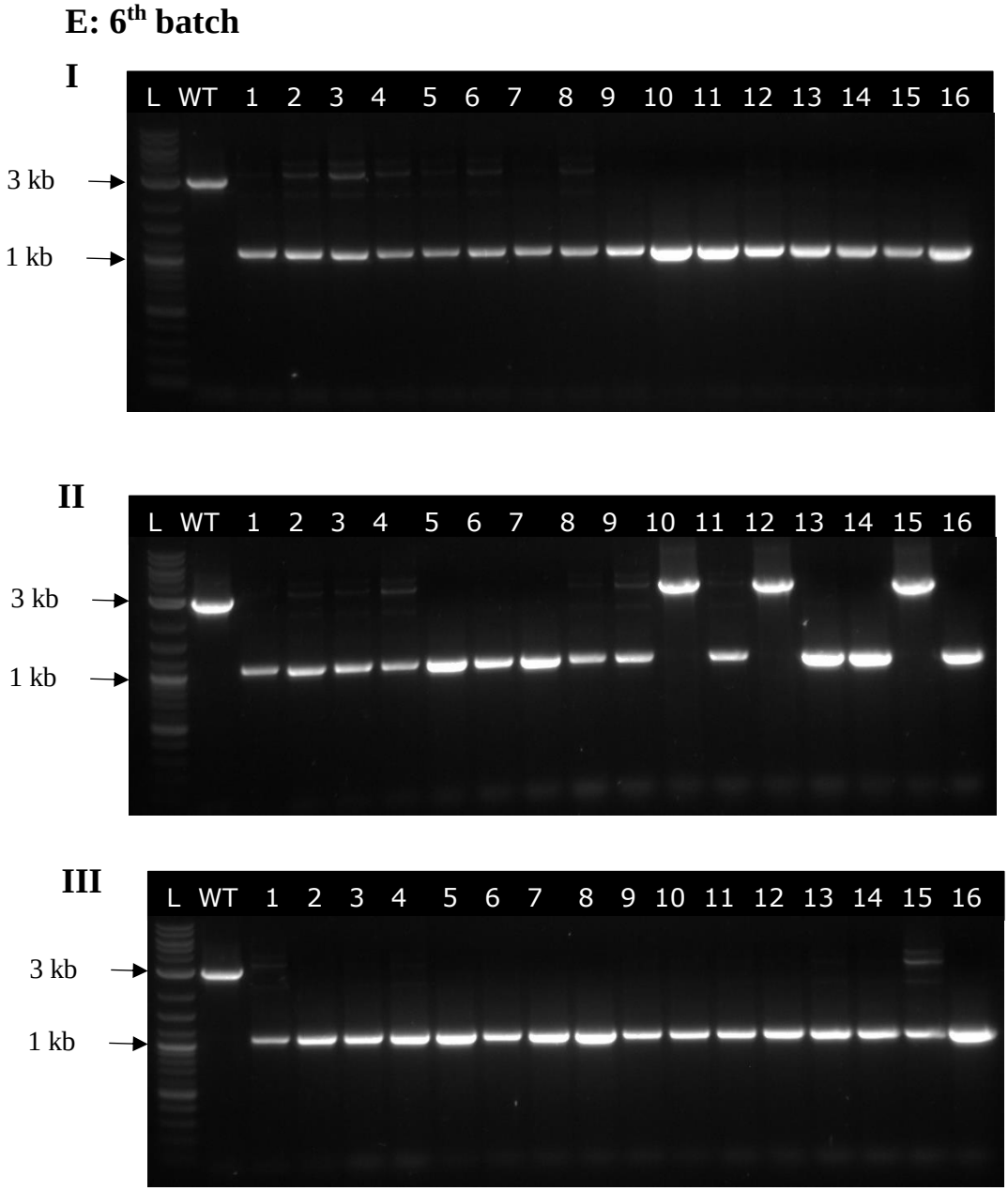


Figure 4.6: Screening of mutants for *pflB* deletion using Cas3_Pfl_F and Cas3_Pfl_R primers. The WT generated a ~ 3.3 kb DNA fragment, whereas the DNA template from Δ*ldh, pdh*^{up}*,* Δ*pfl* mutants generated a ~ 1 kb fragment. WT: PCR performed with *P. thermoglucosidasius* 11955 wild-type. L = NEB 2-log DNA Ladder (0.1 -10 kb); **A:** 1st batch; I: Lanes 1-8 = AM242_1, lanes 9-16 = AM242_2; II: Lanes

1-8 = AM242_3, lanes 9-16 = AM242_4; and III: Lanes 1-8 = AM242_5, lanes 9-16 = AM242_6; **B:** 2nd batch; I: Lanes 1-8 = AM242_1, lanes 9-16 = AM242_2; II: Lanes 1-8 = AM242_3, lanes 9-16 = AM242_4; and III: Lanes 1-8 = AM242_5, lanes 9-16 = AM242_6; **C:** 3rd batch; I: Lanes 1-8 = AM242_1, lanes 9-16 = AM242_2; II: Lanes 1-8 = AM242_3, lanes 9-16 = AM242_4; and III: Lanes 1-8 = AM242_5, lanes 9-16 = AM242_6; **D:** 4th batch; I: Lanes 1-8 = AM242_1, lanes 9-16 = AM242_2; II: Lanes 1-8 = AM242_3, lanes 9-16 = AM242_4; and III: Lanes 1-8 = AM242_5, lanes 9-16 = AM242_6; **E:** 5th batch; I: Lanes 1-8 = AM242_1, lanes 9-16 = AM242_2; II: Lanes 1-8 = AM242_3, lanes 9-16 = AM242_4; and III: Lanes 1-8 = AM242_5, lanes 9-16 = AM242_6; and **F:** 6th batch; I: Lanes 1-8 = AM242_1, lanes 9-16 = AM242_2; II: Lanes 1-8 = AM242_3, lanes 9-16 = AM242_4; and III: Lanes 1-8 = AM242_5, lanes 9-16 = AM242_6.

4.2.3 Characterisation of strains

The fermentation profiles in terms of ethanol and organic acid production of the engineered mutants, AM180 (Δ*ldh*, *pdh*^{up}) and AM242 (Δ*ldh*, *pdh*^{up}, Δ*pfl*), alongside wild-type, TM89 (Δ*ldh*), TM242 (Cripps *et al.*, 2009) and LS242 (Sheng *et al.*, 2017), was assessed during growth on ASYE medium with 20 g·l^{-1} glucose, M-ASYE with 10 g·l^{-1} xylose and M-ASYE with 10 g·l^{-1} cellobiose in sealed 50 ml falcon tubes with 40 ml of the medium containing the appropriate carbon source. Cultures were inoculated (10%) with aerobically grown pre-cultures, followed by incubation for 24 or 48 h at 60 °C with 250 rpm shaking. 40 ml media in 50 ml falcon tube represented a low aeration system. Three biological replicates (three cultures derived from the same parent culture) of each independently generated mutant strain were analysed in each case.

The comparative fermentation profiles of the thirty-six independent but genetically equivalent AM242 strains across the six different batches showed significant differences ($P^{***} \leq 0.001$) using glucose as a carbon source (Figure 4.7). The ethanol titres of the AM242 strains ranged from 51.2 – 180.1 mM achieved in 24 h compared to the 184.6 mM and 78.6 mM attained by TM242 and LS242, respectively. However, the ethanol titre of the strains AM242_2_4, AM242_2_5, AM242_4_1, AM242_4_2, AM242_4_3, AM242_4_4, and AM242_4_5 showed no statistically significant difference ($P > 0.05$) from that of the TM242 strain and were in agreement with those previously reported by Cripps *et al.* (2009) and Sheng *et al.* (2017). Whereas all of the available glucose was consumed by TM242 after 24 h, some residual glucose (1.8 – 44.2 mM) was observed in most of the AM242 strains after an equivalent period, except for the AM242_2_5, AM242_4_2, and AM242_4_3 strains that utilised all the glucose. The ethanol yields of the AM242 strains range from 0.7 – 1.6 mol/mol of glucose, with strains AM242_1_5, AM242_2_5, AM242_4_1, AM242_4_4 and AM242_4_5 having comparable yields to that of TM242 (1.7 mol/mol glucose representing over 80% theoretical yield) as shown in Table 4.1.

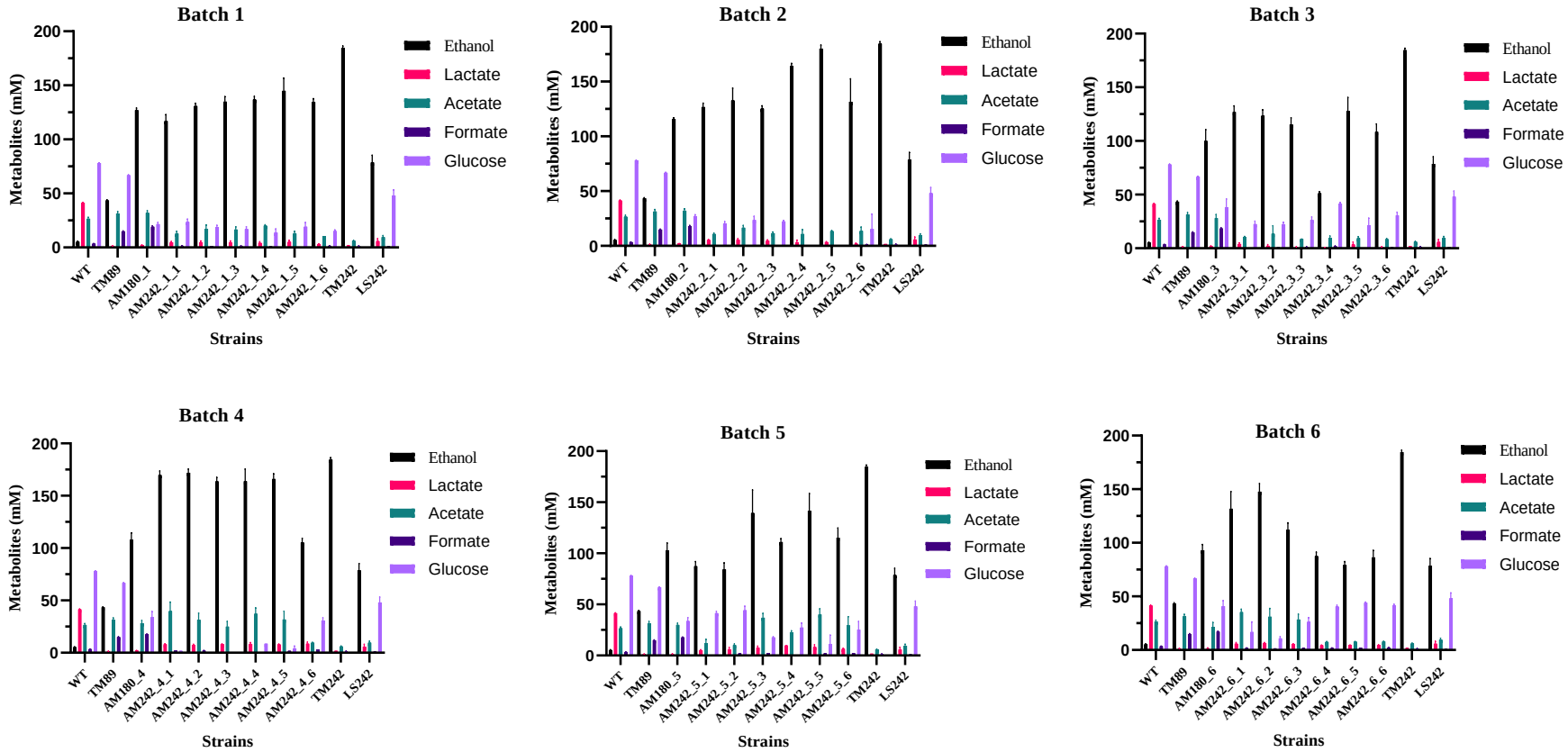


Figure 4.7: Production of metabolites by engineered strains of *P. thermoglucosidasius* NCIMB 11955 using glucose. Fermentation products (ethanol, lactate, acetate, formate and glucose) of the engineered AM242 (Δ*ldh*, *pdh*^{up}, Δ*pfl*) strains in three biological replicates are characterised by HPLC after 24 h of fermentation using 40 ml of ASYE with 1% yeast extract and 111 mM glucose in a sealed 50 ml falcon tubes incubated at 60 ˚C, 250 rpm. *P***** ≤ 0.0001 was calculated by One-way ANOVA followed by Dunnett's multiple comparisons test, and the results are shown as mean ± SD of three biological replicates.

Table 4.1: Ethanol yield of engineered strains of *P. thermoglucosidasius* using glucose

Strains	Metabolite concentration (mM) after 24 h of fermentation		
	*Consumed glucose	Ethanol	**Ethanol yield (mol/mol)
WT (NCIMB 11955)	32.9	5.3	0.2
TM89 (*ldh*$^{-}$)	44.2	43.4	1.0
AM180_1 (*ldh*$^{-}$, *pdh*^{up})	89.4	126.9	1.4
AM180_2 (*ldh*$^{-}$, *pdh*^{up})	84.1	115.9	1.4
AM180_3 (*ldh*$^{-}$, *pdh*^{up})	72.7	99.9	1.4
AM180_4 (*ldh*$^{-}$, *pdh*^{up})	77.2	108.0	1.4
AM180_5 (*ldh*$^{-}$, *pdh*^{up})	77.4	102.9	1.3
AM180_6 (*ldh*$^{-}$, *pdh*^{up})	70.1	93.1	1.3
AM242_1_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	87.4	116.9	1.3
AM242_1_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	92.0	130.9	1.4
AM242_1_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	94.1	134.9	1.4
AM242_1_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	97.2	136.7	1.4
AM242_1_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	91.8	144.9	1.6
AM242_1_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	95.9	134.5	1.4
AM242_2_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	90.7	126.7	1.4
AM242_2_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	87.3	132.7	1.5
AM242_2_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	88.7	125.2	1.4
AM242_2_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	110.7	164.4	1.5
AM242_2_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	111.0	180.1	1.6
AM242_2_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	95.5	131.5	1.4
AM242_3_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	88.5	127.0	1.4
AM242_3_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	88.6	123.5	1.4
AM242_3_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	84.6	115.3	1.4
AM242_3_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	69.4	51.2	0.7
AM242_3_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	89.7	127.9	1.4
AM242_3_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	80.3	108.5	1.4
AM242_4_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	109.2	169.8	1.6
AM242_4_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	111.0	171.9	1.5
AM242_4_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	111.0	163.6	1.5

AM242_4_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	102.3	163.8	1.6
AM242_4_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	106.9	166.0	1.6
AM242_4_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	80.1	105.2	1.3
AM242_5_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	69.7	87.4	1.3
AM242_5_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.8	84.2	1.3
AM242_5_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	93.4	139.3	1.5
AM242_5_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	83.7	111.0	1.3
AM242_5_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	99.9	141.8	1.4
AM242_5_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	85.7	115.2	1.3
AM242_6_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	94.2	131.8	1.4
AM242_6_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	100.1	147.6	1.5
AM242_6_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	84.8	112.2	1.3
AM242_6_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	70.5	87.5	1.2
AM242_6_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.8	79.4	1.2
AM242_6_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	69.7	86.4	1.2
TM242 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	111.0	184.6	1.7
LS242 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	63.0	78.6	1.2

* Initial glucose: 111 mM; ** Theoritical maximum: 2 moles of ethanol per mole of glucose

One of the advantages of *Parageobacillus* sp. as a host for biofuel production such as ethanol is their ability to grow on a wide range of sugars, both hexoses and pentoses typically found in lignocellulosic biomass. For this reason, fermentation with cellobiose and xylose was also trialled. Cellobiose as a carbon source was also converted to ethanol by the AM242 strains with titres ranging from 69.9 – 93.9 mM, as presented in Figure 4.8. There was a statistically significant difference between the ethanol titres of the strains ($P^{***} \leq 0.001$). The ethanol yields produced by all the engineered strains are presented in Table 4.2, with the highest yield of 3.2 mol/mol of cellobiose representing 80% of the theoretical yield compared to 85% by TM242 (3.4 mol/mol of cellobiose). Furthermore, all of the available cellobiose was utilised by the AM242 strains and the TM242 after 24 h of fermentation.

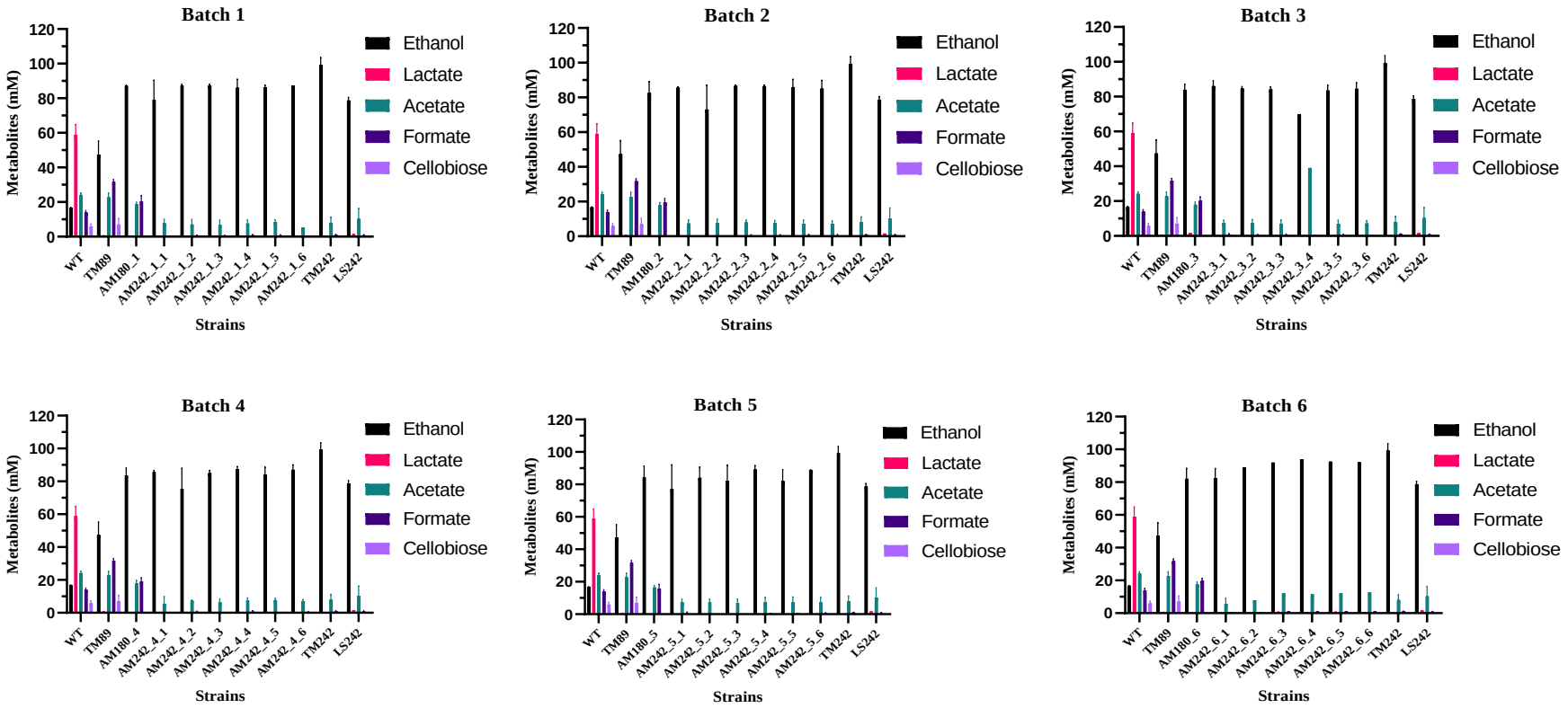


Figure 4.8: Production of metabolites by engineered strains of *P. thermoglucosidasius* NCIMB 11955 using cellobiose. Fermentation products (ethanol, lactate, acetate, formate and cellobiose) of the engineered AM242 (Δ*ldh*, *pdh*^{up}, Δ*pfl*) strains in three biological replicates are characterised by HPLC after 24 h of fermentation using 40 ml of M-ASYE with 1% yeast extract and 29.2 mM cellobiose in a sealed 50 mL falcon tubes incubated at 60 ˚C, 250 rpm. $P^{***} \leq 0.001$ was calculated by One-way ANOVA followed by Dunnett's multiple comparisons tests, and the results are shown as mean ± SD of three biological replicates.

Table 4.2: Ethanol yield of engineered strains of *P. thermoglucosidasius* using cellobiose

Strains	Metabolite concentration (mM) after 24 h of fermentation		
	*Consumed cellobiose	Ethanol	**Ethanol yield (mol/mol)
WT (NCIMB 11955)	23.3	16.7	0.7
TM89 (*ldh*$^{-}$)	22.1	47.4	2.1
AM180_1 (*ldh*$^{-}$, *pdh*^{up})	29.2	87.2	3.0
AM180_2 (*ldh*$^{-}$, *pdh*^{up})	29.2	82.6	2.8
AM180_3 (*ldh*$^{-}$, *pdh*^{up})	29.2	83.9	2.9
AM180_4 (*ldh*$^{-}$, *pdh*^{up})	29.2	83.6	2.9
AM180_5 (*ldh*$^{-}$, *pdh*^{up})	29.2	84.3	2.9
AM180_6 (*ldh*$^{-}$, *pdh*^{up})	29.2	82.0	2.8
AM242_1_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	79.2	2.7
AM242_1_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	87.4	3.0
AM242_1_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	87.4	3.0
AM242_1_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	86.1	2.9
AM242_1_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	86.4	3.0
AM242_1_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	87.4	3.0
AM242_2_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	85.6	2.9
AM242_2_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	73.0	2.5
AM242_2_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	86.6	3.0
AM242_2_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	86.5	3.0
AM242_2_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	85.9	2.9
AM242_2_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	85.0	2.9
AM242_3_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	86.0	2.9
AM242_3_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	84.6	2.9
AM242_3_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	84.2	2.9
AM242_3_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	69.9	2.4
AM242_3_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	83.5	2.9
AM242_3_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	84.5	2.9
AM242_4_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	85.6	2.9
AM242_4_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	75.4	2.6
AM242_4_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	85.1	2.9

AM242_4_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	87.5	3.0
AM242_4_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	84.1	2.9
AM242_4_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	87.0	3.0
AM242_5_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	77.2	2.6
AM242_5_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	84.1	2.9
AM242_5_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	82.3	2.8
AM242_5_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	89.2	3.1
AM242_5_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	82.1	2.8
AM242_5_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	88.8	3.0
AM242_6_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	82.5	2.8
AM242_6_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	89.0	3.0
AM242_6_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	91.9	3.1
AM242_6_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	93.9	3.2
AM242_6_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	92.5	3.2
AM242_6_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	92.0	3.1
TM242 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	99.3	3.4
LS242 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.2	78.7	2.7

* Initial cellobiose: 29.2 mM; ** Theoritical maximum: 4 moles of ethanol per mole of cellobiose

Fermentation with xylose as a carbon source to produce ethanol by the engineered AM242 strains was also investigated. The comparative fermentation profiles of the strains showed significant differences ($P^{***} \leq 0.001$) (Figure 4.9). In addition, the solvent profile of the AM242 triple deletion mutants differed significantly from that of TM242 except for AM242_4_5 (76.4 mM), with only 4.0 – 76.4 mM ethanol achieved in 24 h compared to the 76.3 mM produced by the strain TM242. It was further apparent that only about 50% of the available xylose was consumed after 24 h in the AM242 strains except for AM242_4_5. In comparison, only about 10% of the xylose was consumed by strain LS242. At the same time, all the available xylose was utilised by TM242 after an equivalent period. The ethanol yields produced by all the engineered strains in 24 h are presented in Table 4.3, with 80% of the strains producing more than

50% of the theoretical yield, while AM242_1_6, AM242_2_1 and AM242_4_5 strains having comparable yields to TM242 (1.1 mol/mol of xylose representing 66% of theoretical yield). However, when all the AM242 strains were grown on xylose under aerobic conditions (50 ml of M-ASYE with 10 $g \cdot l^{-1}$ xylose and 10 $g \cdot l^{-1}$ yeast extract medium in 250 ml baffled flasks), all the available xylose was consumed in 24 h (data not shown). This observation indicates that the incomplete utilisation of xylose by the AM242 strains under the micro-erobic conditions might be due to redox imbalance.

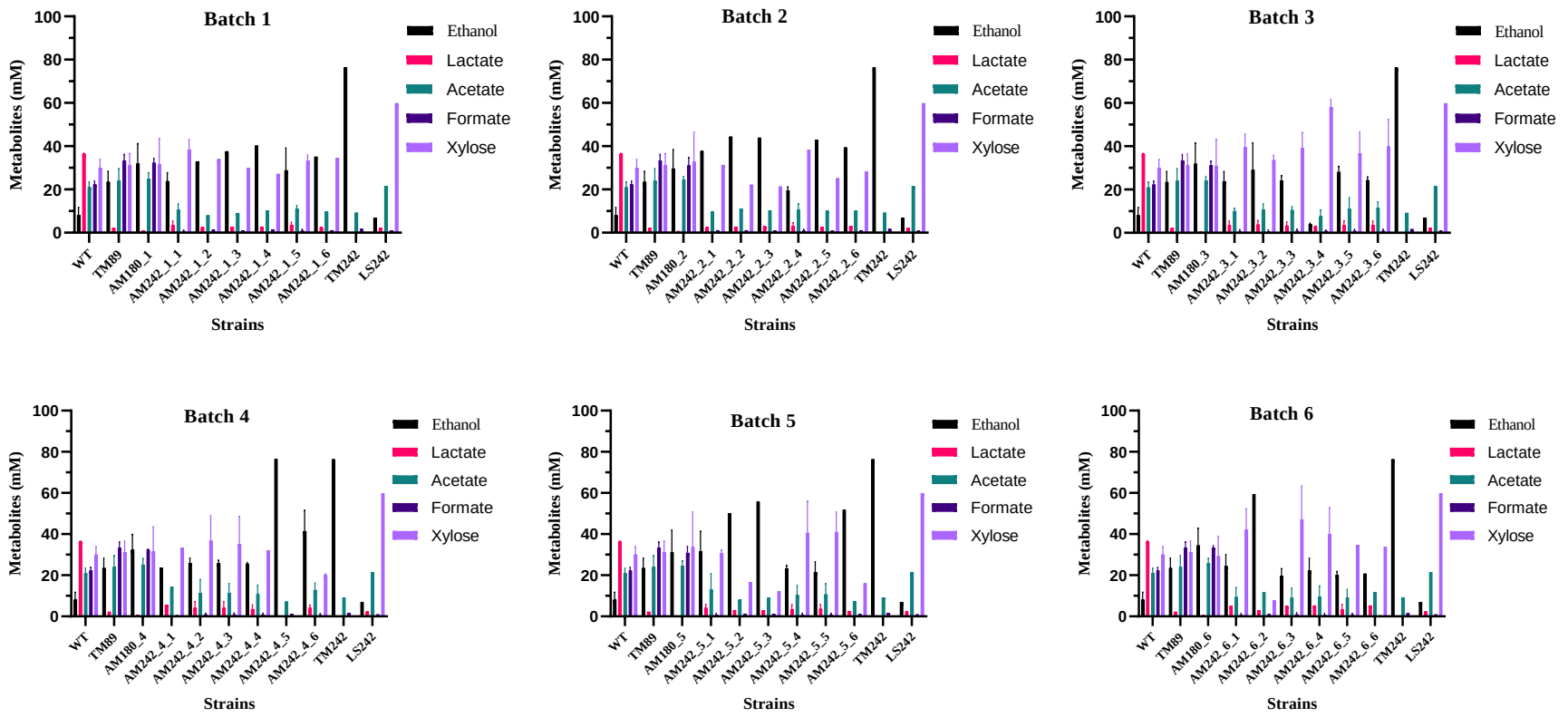


Figure 4.9: Production of metabolites by engineered strains of *P. thermoglucosidasius* NCIMB 11955 using xylose after 24 h. Fermentation products (ethanol, lactate, acetate, formate and xylose) of the engineered AM242 (Δ*ldh*, *pdh*^{up}, Δ*pfl*) strains in three biological replicates are characterised by HPLC after 24 h of fermentation using 40 ml of M-ASYE with 1% yeast extract and 66.6 mM xylose in a sealed 50 ml falcon tubes incubated at 60 ˚C, 250 rpm. *P***** ≤ 0.0001 was calculated by One-way ANOVA followed by Dunnett's multiple comparisons test, and the results are shown as mean ± SD of three biological replicates.

Table 4.3: Ethanol yield of engineered strains of *P. thermoglucosidasius* using xylose after 24 h

Strains	Metabolite concentration (mM) after 24 h of fermentation		
	*Consumed xylose	Ethanol	**Ethanol yield (mol/mol)
WT (NCIMB 11955)	36.6	8.1	0.2
TM89 (*ldh*$^{-}$)	35.4	23.5	0.7
AM180_1 (*ldh*$^{-}$, *pdh*^{up})	35.1	31.9	0.9
AM180_2 (*ldh*$^{-}$, *pdh*^{up})	33.9	29.7	0.9
AM180_3 (*ldh*$^{-}$, *pdh*^{up})	35.6	31.9	0.9
AM180_4 (*ldh*$^{-}$, *pdh*^{up})	35.1	32.4	0.9
AM180_5 (*ldh*$^{-}$, *pdh*^{up})	32.9	31.2	0.9
AM180_6 (*ldh*$^{-}$, *pdh*^{up})	37.4	34.5	0.9
AM242_1_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	28.2	23.8	0.8
AM242_1_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	32.6	33.0	1.0
AM242_1_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	36.6	37.6	1.0
AM242_1_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	39.4	40.2	1.0
AM242_1_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	33.4	28.8	0.9
AM242_1_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	32.1	35.1	1.1
AM242_2_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	35.3	37.7	1.1
AM242_2_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	44.4	44.4	1.0
AM242_2_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	45.3	43.9	1.0
AM242_2_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	28.2	19.6	0.7
AM242_2_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	41.5	42.9	1.0
AM242_2_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	38.3	39.4	1.0
AM242_3_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	27.1	23.8	0.9
AM242_3_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	33.0	29.0	0.9
AM242_3_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	27.3	24.1	0.9
AM242_3_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	8.5	4.0	0.5
AM242_3_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.8	28.1	0.9
AM242_3_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	26.7	24.3	0.9
AM242_4_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	33.3	23.6	0.7
AM242_4_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	29.9	25.9	0.9
AM242_4_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	31.4	25.9	0.8

AM242_4_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	34.6	25.6	0.7
AM242_4_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.6	76.4	1.1
AM242_4_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	46.5	41.4	0.9
AM242_5_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	35.9	31.5	0.9
AM242_5_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	49.9	50.0	1.0
AM242_5_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	54.4	55.7	1.0
AM242_5_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	26.0	23.2	0.9
AM242_5_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	25.6	21.5	0.8
AM242_5_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	50.5	51.7	1.0
AM242_6_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	24.5	24.4	1.0
AM242_6_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	58.9	59.4	1.0
AM242_6_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	19.4	19.7	1.0
AM242_6_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	26.6	22.4	0.8
AM242_6_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	31.8	20.2	0.6
AM242_6_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	32.9	20.8	0.6
TM242 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.6	76.3	1.1
LS242 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	15.7	7.0	0.4

* Initial xylose: 66.6 mM; ** Theoritical maximum: 1.67 moles of ethanol per mole of xylose

After observing that about half of the xylose was not utilised under micro-aerobic conditions after 24 h, extending the fermentation period to 48 h was tested to see if it would result in all the available xylose being consumed and a corresponding increase in ethanol being produced. Interestingly, all the available xylose was consumed in most strains after extending the fermentation period to 48 h with a corresponding increase in ethanol titres similar to that of the TM242 strain, as presented in Figure 4.10. Whereas LS242 still did not consume much of the xylose. The difference in ethanol titres between the AM242 strains was statistically significant ($P^{****} \leq 0.0001$). The ethanol yield of the strains AM242_1_2, AM242_2_4, AM242_2_5, AM242_3_1, AM242_3_2, AM242_3_3, AM242_4_2, AM242_4_3, AM242_4_5 and AM242_6_1 was over 70% of the theoretical yield and showed no significant difference from that of the TM242

strain (1.3 mol/mol of xylose) as shown in Table 4.4. This observation indicates that the AM242 strains were slower in xylose utilisation than the TM242 strain under micro-aerobic conditions for ethanol production.

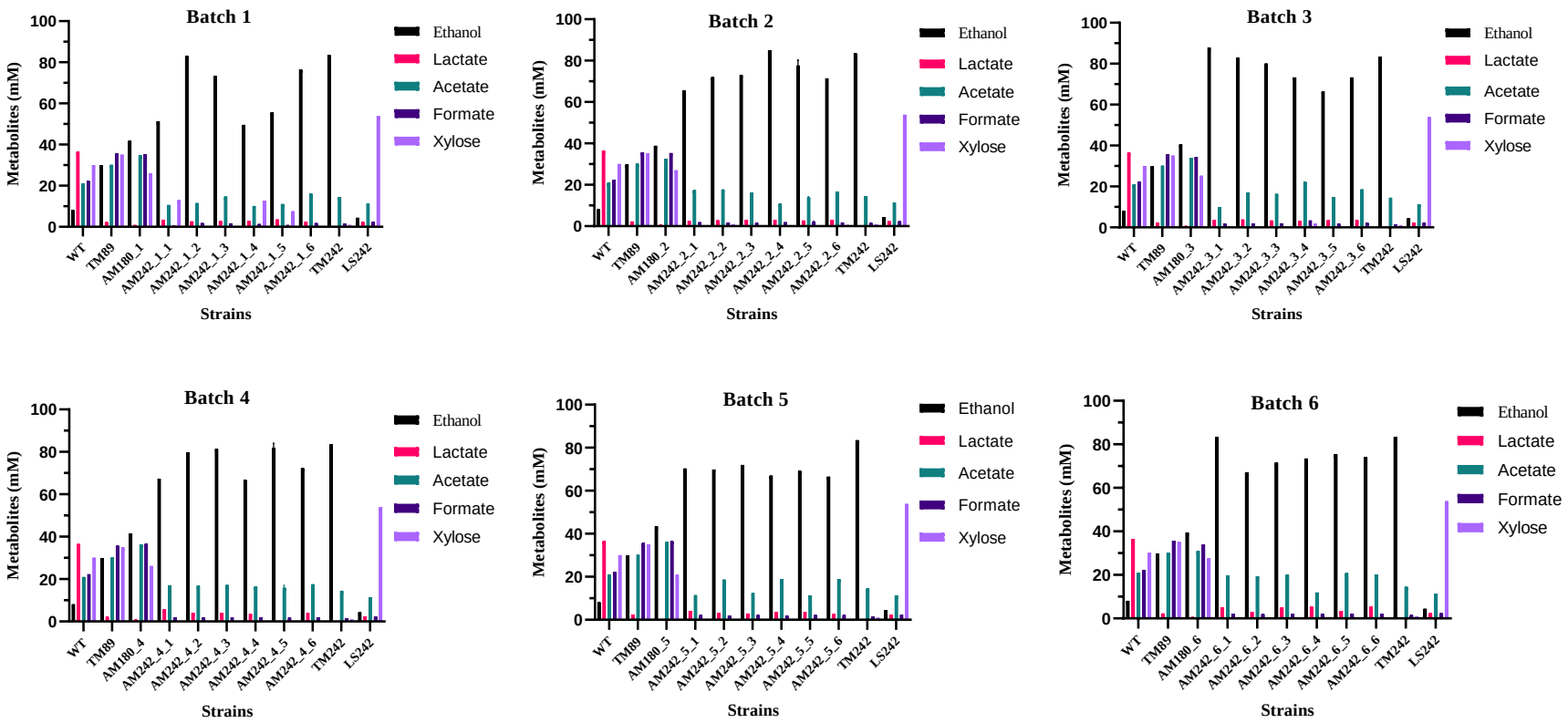


Figure 4.10: Production of metabolites by engineered strains of *P. thermoglucosidasius* NCIMB 11955 using xylose after 48 h. Fermentation products (ethanol, lactate, acetate, formate and xylose) of the engineered AM242 (Δ*ldh*, *pdh*^{up}, Δ*pfl*) strains in three biological replicates are characterised by HPLC after 48 h of fermentation using 40 ml of M-ASYE with 1% yeast extract and 66.6 mM xylose in a sealed 50 ml falcon tubes incubated at 60 ˚C, 250 rpm. *P***** ≤ 0.0001 was calculated by One-way ANOVA followed by Dunnett's multiple comparisons test, and the results are shown as mean ± SD of three biological replicates.

Table 4.4: Ethanol yield of engineered strains of *P. thermoglucosidasius* using xylose after 48 h

Strains	Metabolite concentration (mM) after 48 h of fermentation		
	*Consumed xylose	Ethanol	**Ethanol yield (mol/mol)
WT (NCIMB 11955)	36.6	8.1	0.2
TM89 (*ldh*$^{-}$)	31.6	29.9	0.9
AM180_1 (*ldh*$^{-}$, *pdh*^{up})	40.6	41.8	1.0
AM180_2 (*ldh*$^{-}$, *pdh*^{up})	39.7	38.8	1.0
AM180_3 (*ldh*$^{-}$, *pdh*^{up})	41.4	40.5	1.0
AM180_4 (*ldh*$^{-}$, *pdh*^{up})	40.4	41.4	1.0
AM180_5 (*ldh*$^{-}$, *pdh*^{up})	45.6	43.5	1.0
AM180_6 (*ldh*$^{-}$, *pdh*^{up})	39.1	39.5	1.0
AM242_1_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	53.7	51.3	1.0
AM242_1_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	83.1	1.3
AM242_1_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	73.4	1.1
AM242_1_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	53.9	49.6	0.9
AM242_1_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	59.1	55.6	0.9
AM242_1_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	76.4	1.1
AM242_2_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.4	65.5	1.0
AM242_2_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	65.7	72.0	1.1
AM242_2_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.4	73.1	1.1
AM242_2_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	85.0	1.3
AM242_2_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.6	77.6	1.2
AM242_2_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.0	71.4	1.1
AM242_3_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	87.9	1.3
AM242_3_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	82.8	1.2
AM242_3_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	80.0	1.2
AM242_3_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	64.9	73.3	1.1
AM242_3_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	66.4	1.0
AM242_3_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	73.3	1.1
AM242_4_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	67.4	1.0
AM242_4_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	79.7	1.2
AM242_4_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	81.4	1.2

AM242_4_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	66.7	1.0
AM242_4_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	81.9	1.2
AM242_4_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	72.3	1.1
AM242_5_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	70.3	1.1
AM242_5_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	69.9	1.1
AM242_5_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	72.0	1.1
AM242_5_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	67.0	1.0
AM242_5_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	69.2	1.0
AM242_5_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	66.5	1.0
AM242_6_1 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	83.4	1.3
AM242_6_2 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	67.1	1.0
AM242_6_3 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	71.5	1.1
AM242_6_4 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	73.5	1.1
AM242_6_5 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	75.5	1.1
AM242_6_6 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	66.5	74.1	1.1
TM242 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	65.7	83.5	1.3
LS242 (*ldh*$^{-}$, *pdh*^{up}, *pfl*$^{-}$)	12.8	4.4	0.3

* Initial xylose: 66.6 mM; ** Theoritical maximum: 1.67 moles of ethanol per mole of xylose

4.2.4 Genome sequencing and analysis of mutant strains

To provide insight into the possible reasons for the observed differences in the fermentation profile of the equivalent AM242 strains made here, the genomic DNA of all the thirty-six AM242 strains was subjected to Illumina paired-end sequencing. The sequence reads obtained were mapped to the *P. thermoglucosidasius* NCIMB 11955 reference genome (Sheng *et al.,* 2016) (NCBI accession number CP016622–CP016624) using CLC Genomics Workbench v20.0.4 (Qiagen, DK), as described in section 2.14.

The results of the single-nucleotide polymorphisms (SNPs)/insertions and deletions (Indels) analysis are presented in Table 4.5. From these data, it was evident that unintended mutations occurred during the genetic modifications in the form of SNPs and Indels. On one hand, during the engineering of the TM242 strain by Cripps *et al.* (2009) from *P. thermoglucosidasius* NCIMB 11955, a total of eleven unintended mutations had occurred. Nine of these mutations were present in the Δ*ldh* mutant progenitor strain TM89, and two additional mutations arose from TM89 to TM242 (Appendix IV) (Sheng *et al.*, 2017). On the other hand, the equivalent strain to TM242 made by Sheng *et al.* (2017) from the wild-type *P. thermoglucosidasius* NCIMB 11955 strain using the *pyrE*-based allelic exchange system carried four additional mutations compared to the parent strain NCIMB 11955. The equivalent strains to TM242 made here (strains AM242) from the Δ*ldh* mutant strain TM89, using the theophylline-responsive CRISPR/Cas9 genome editing system, carried additional mutations ranging from 1-7 across the thirty-six strains compared to the parental strain TM89 (Figure 4.11). Whereas strains AM242_2_1, AM242_2_4, AM242_2_5, AM242_2_6, AM242_4_1, and AM242_6_3, picked no additional mutations. Noticeably, the majority of the SNPs and Indels occurred within coding regions. Furthermore, most

SNPs were non-synonymous, causing changes in the encoded amino acid. In two strains, AM242_4_6 and AM242_6_2, premature stop codons were introduced, which truncated the encoded hypothetical protein (BCV53_05335; position 1084854) and glutamate dehydrogenase (BCV53_15805; position 3219412), respectively. Additionally, a number of SNPs occurred within hypothetical proteins. The only noticeable SNPs were in transport proteins namely; permease (BCV53_12920), ABC transporter permease (BCV53_13695), PTS sugar transporter subunit IIA (BCV53_10245), and MFS transporter (BCV53_14315). Some SNPs occurred in different batches such as 4^{th} and 6^{th} (DNA gyrase subunit B; BCV53_03070), 1^{st} and 3^{rd} (Histidine kinase; BCV53_04235), 4^{th} and 6^{th} (Hypothetical protein; BCV53_05335), 4^{th} and 5^{th} (Hypothetical protein; BCV53_06805), 1^{st} and 3^{rd} (Stage V sporulation protein AD; BCV53_07525), 4^{th} and 6^{th} (Permease; BCV53_12920), 1^{st} and 3^{rd} (Acetyladehyde dehydrogenase; BCV53_13005), 1^{st}, 3^{rd}, 4^{th} and 5^{th} (Hypothetical protein; BCV53_15235), and 4^{th} and 6^{th} (Glutamate dehydrogenase; BCV53_15805).

SNP candidates in representative strains identified by Illumina sequencing were independently verified by PCR and Sanger sequencing (Appendix V). A separate list showing SNPs per individual strain is presented in Appendix VI.

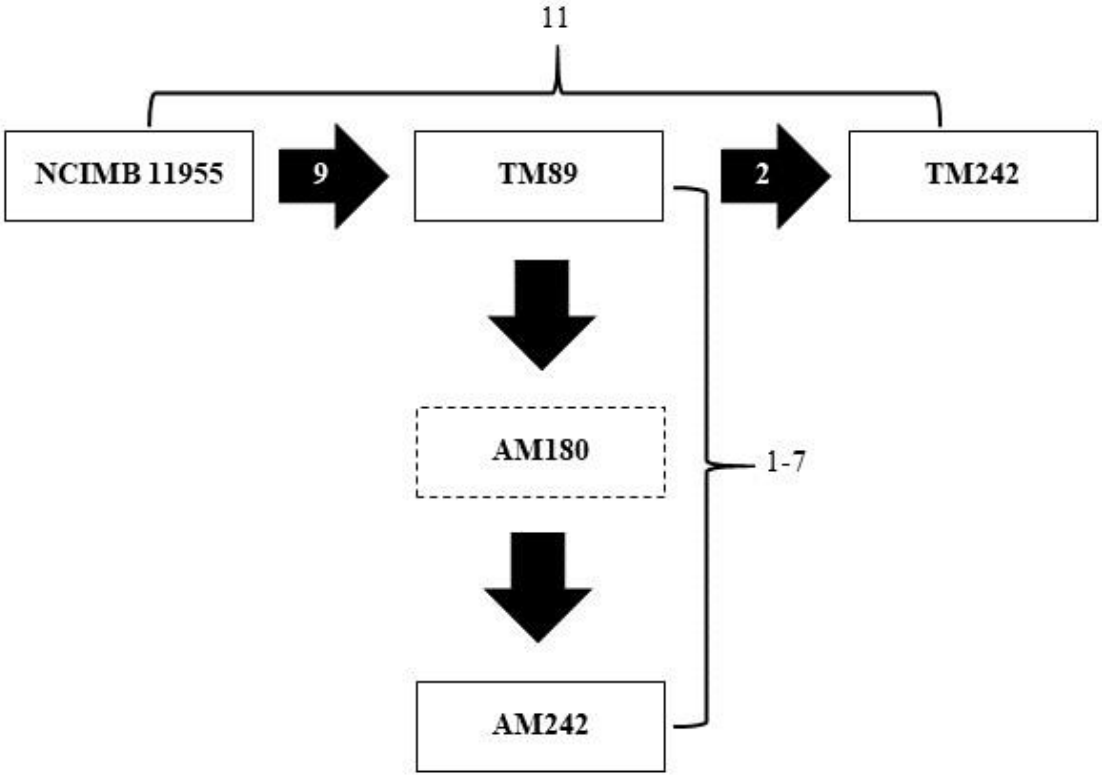


Figure 4.11: Overview of strain generation and SNPs acquired during strain generation. The engineering of TM242 equivalent strains (AM242) follows a stepwise route of *pdhA* up-regulation followed by *pflB* deletion from TM89 progenitor (TM89, AM180, AM242). Strains in solid black lines were subjected to Illumina paired-read sequencing (TM89 and TM242 by Sheng *et al.*, 2017) and analysed by mapping the reads against the deposited genome of *P. thermoglucosidasius* NCIMB 11955 (NCBI Accession Number: CP016622-CP016624) using CLC Genomics Workbench v20.0.4. Numbers indicate the mutations acquired. AM242 strains acquired mutations ranging from 1-7 across the thirty-six independent strains compared to the parental strain TM89. A total of nine mutations were noted between TM89 and NCIMB 11955 (8 SNPs, 1 Indel) with two further mutations found in TM242 (1 SNV, 1 Indel), resulting in 11-nucleotide changes compared to NCIMB 11955.

Table 4.5: Unique SNVs and Indels of AM242 strains used in this study compared to NCIMB 11955

Position	Strains	SNP	Gene	Locus	Effect
26985*	AM242_5_2, AM242_5_3	A>G	Hydroxyglutarate oxidase	BCV53_19285	K237R
296868	AM242_6_6	A>G	Transposase	BCV53_01475	C72R
405673	AM242_4_3	T>-	MerR family transcriptional regulator	BCV53_01955	V65fs
609276	AM242_4_2, AM242_6_4	G>A	DNA gyrase subunit B	BCV53_03070	T353M
856889	AM242_1_3, AM242_1_6, AM242_3_1, AM242_3_2, AM242_3_3, AM242_3_5	C>T	Histidine kinase	BCV53_04235	-
1084854	AM242_4_6, AM242_6_2	A>T	Hypothetical protein	BCV53_05335	L616*
1086291	AM242_3_4	T>C	Hypothetical protein	BCV53_05335	E137G
1367189	AM242_4_3, AM242_5_1 AM242_5_2, AM242_5_3 AM242_5_4, AM242_5_5 AM242_5_6	A>G	Hypothetical protein	BCV53_06805	-
1415815	AM242_1_1, AM242_1_2, AM242_1_4, AM242_1_5	A>T	AraC family transcriptional regulator	BCV53_07015	Q158L
1471195	AM242_1_1, AM242_1_2, AM242_1_4, AM242_1_5	G>A	Carbamoyl phosphate synthase large subunit	BCV53_07290	V152I
1514453	AM242_1_6, AM242_3_1	G>A	Stage V sporulation protein AD	BCV53_07525	G166E
1541623	AM242_4_3	C>T	Hypothetical protein	BCV53_07715	L39F
1955303	AM242_2_2	C>G	Hypothetical protein	BCV53_09715	A157P
1986844	AM242_5_1, AM242_5_2, AM242_5_4, AM242_5_5	G>A	-	-	-
2063731	AM242_5_2	C>T	PTS sugar transporter subunit IIA	BCV53_10245	-
2169641	AM242_1_1, AM242_1_2, AM242_1_4, AM242_1_5	C>T	Acyl-CoA dehydrogenase	BCV53_10705	A196T

2242729	AM242_4_2, AM242_6_4	T>C	-	-	-
2323810	AM242_1_2	G>A	Histidine kinase	BCV53_11450	G566E
2337594	AM242_1_6	G>A	Arabinose-binding protein	BCV53_11500	G314E
2360630	AM242_5_1, AM242_5_5, AM242_5_6	C>T	Xylulokinase	BCV53_11600	T180M
2570676	AM242_1_1, AM242_1_2, AM242_1_4, AM242_1_5	A>G	Acyl-CoA dehydrogenase	BCV53_12610	M208V
2638542	AM242_4_4, AM242_4_5, AM242_4_6, AM242_6_1, AM242_6_2, AM242_6_5	C>T	Permease	BCV53_12920	D231N
2657967	AM242_1_3, AM242_1_6, AM242_3_1, AM242_3_2, AM242_3_3	G>A	Acetaldehyde dehydrogenase (acetylating)	BCV53_13005	-
2791650	AM242_5_1, AM242_5_2, AM242_5_3, AM242_5_4, AM242_5_5, AM242_5_6	C>T	ABC transporter permease	BCV53_13695	-
2891366	AM242_2_3	G>A	Hypothetical protein	BCV53_14215	R56W
2916351	AM242_4_3	G>A	MFS transporter	BCV53_14315	A155T
3015609	AM242_3_4, AM242_3_6	C>T	Nucleotidyltransferase	BCV53_14775	-
3113091	AM242_1_3, AM242_1_4, AM242_3_2, AM242_3_5, AM242_4_5, AM242_5_2	G>A	Hypothetical protein	BCV53_15235	G159R
3219412	AM242_4_6, AM242_6_2	G>A	Glutamate dehydrogenase	BCV53_15805	Q148*
3228940	AM242_1_1, AM242_1_2 AM242_1_4, AM242_1_5	T>C	-	-	-
3411590	AM242_1_1, AM242_1_2, AM242_1_5	G>A	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	BCV53_16860	E114K
3523149	AM242_5_2, AM242_5_3	C>T	tRNA guanosine (34) trans glycosylase Tgt	BCV53_17445	G186D

4.2.5 Creation of E266K SNP in BCV53_09980 of AM242

According to Sheng *et al.* (2017), the TM242 strain had picked up two unique additional SNPs during its construction by Cripps *et al.* (2009) from TM89 (Δ*ldh*) strain, which also has 9 SNPs compared to the wild-type NCIMB 11955 strain. These two additional mutations are one Indel on the small megaplasmid (pNCI002) at position 34296 and one SNP (E266K) in BCV53_09980 encoding a type III restriction-modification system methylation subunit at position 2006731. To further shed light and understand why the TM242 strain produces a high ethanol yield, simultaneously utilising any carbon source used for the fermentation in just 24 h, investigation into the possible role of the unique G>A SNP (E266K) was undertaken. To make one of the AM242 strains resemble the TM242 strain as closely as possible, the strain AM242_2_5 was selected, as it had acquired no additional mutations during its construction, and a G>A SNP created (E266K). At the same time, the G>A SNP in the TM242 strain was corrected back to the original sequence, i.e., A>G. However, Lau *et al.* (2021) reported that the pNCI002 is nonessential. Instead of undergoing homologous recombination to replace the protospacer targeted by the Cas9, it is cured by the double-stranded break caused by the Cas9. Thus, indicating no selection pressure maintaining the pNCI002 like the toxin-antitoxin module of pNCI001. For this reason, it was not attempted to producethe Indel on the pNCI002 in AM242_2_5 or correct the Indel in TM242 strain. Thus, the reason for creating and restoring only the G>A SNP in AM242 and TM242 strains, respectively.

4.2.5.1 Construction of pMTL-AM731, pMTL-AM732, and pMTL-AM733 plasmids

Plasmid pMTL-AM180 was used as the vector backbone. The application-specific modules for both KI of the eCGP123 reporter gene to bookmark the SNP region and the cassettes for replacing the eCGP123 reporter gene with the original sequence containing the appropriate SNP or corrected SNP were synthesised by Integrated DNA Technologies (IDT) Ltd. For the bookmark of the SNP region (2005857-2006296 for TM242 and 2005851-2006290 for AM242), the cassette contained sgRNA targeting the SNP region (identified using Benchling CRISPR Guide Design software (www.benchling.com), T1T2 terminator, and an editing template comprising the eCGP123 reporter gene flanked with homology arms (450 bps) up and downstream of the SNP region. For TM242 SNP correction, the cassette contained sgRNA targeting the bookmarked eCGP123 reporter gene (identified using Benchling CRISPR Guide Design software (www.benchling.com), T1T2 terminator, and an editing template comprising the corrected SNP region sequence flanked with homology arms (450 bps) up and downstream of the eCGP123 reporter gene. On the other hand, the AM242 SNP creation cassette contained sgRNA targeting the bookmarked eCGP123 reporter gene (identified using Benchling CRISPR Guide Design software (www.benchling.com), T1T2 terminator, and an editing template comprising the SNP region sequence flanked with homology arms (450 bps) up and downstream of the eCGP123 reporter gene.

Plasmid pMTL-AM180 and the three application cassettes were digested with BamHI and AscI and gel purified. The vector backbone and the cassettes were ligated between the BamHI/AscI restriction recognition sites yielding plasmids pMTL-AM731, pMTL-AM732, and pMTL-AM733, respectively (Figure 4.12). pMTL-AM731 was used to bookmark the SNP region (2005857-2006296 for TM242 and 2005851-2006290 for

AM242) with the eCGP123 reporter gene. While pMTL-AM732 was used for swapping the eCGP123 gene with the region containing the corrected SNP sequence in TM242, the pMTL-AM733 was used for changing the eCGP123 gene with the region containing the created SNP sequence in AM242.

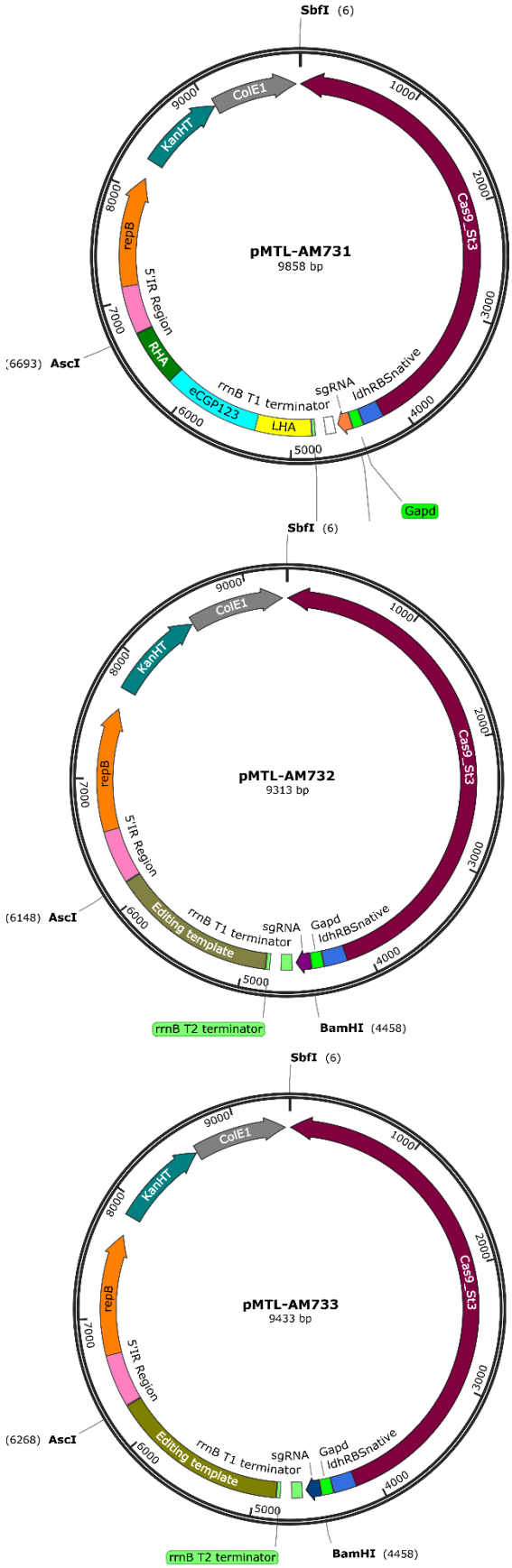


Figure 4.12: Vector map of plasmids pMTL-AM731, pMTL-AM732, and pMTL-AM733. The vectors consist of application cassettes for the bookmark of the eCGP123

reporter gene at the SNP region (pMTL-AM731), correction of G>A SNP (pMTL-AM732), and creation of G>A SNP (pMTL-AM733), respectively, ligated between BamHI and AscI sites in the pMTL-AM180 vector backbone.

Purified pMTL-AM731, pMTL-AM732, and pMTL-AM733 plasmids were screened by restriction digestion using BamHI and AscI (Figure 4.13). Sanger sequencing was performed to confirm the presence of the correct nucleotide sequence.

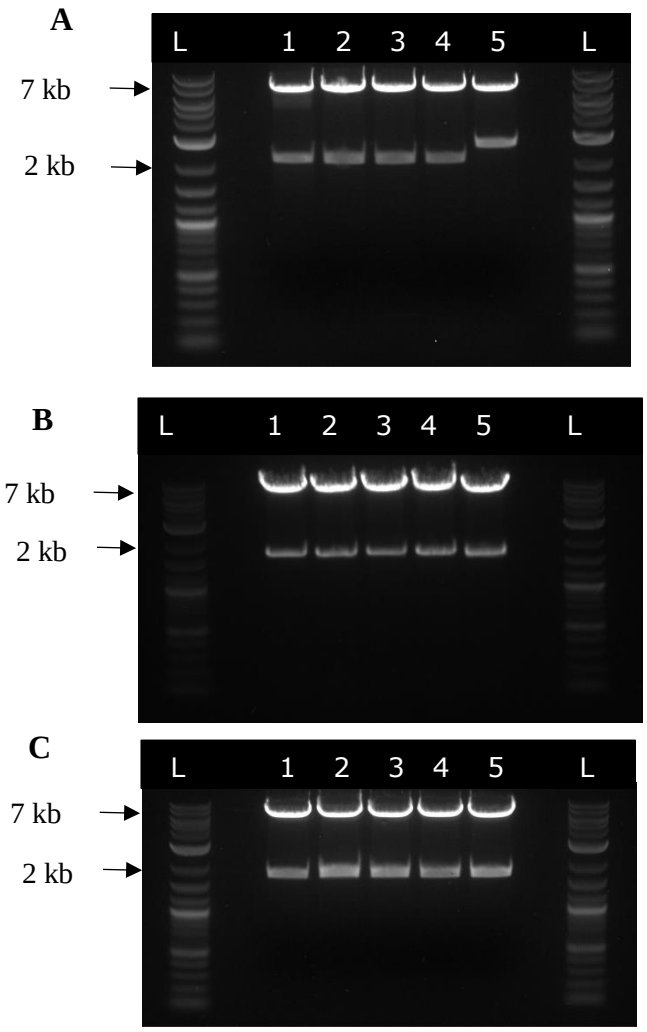


Figure 4.13: Analytical digestion of pMTL-AM731, pMTL-AM732, and pMTL-AM733. L= DNA marker (0.1-10 kb); **A**: lanes 1-5 represent digested pMTL-AM731 with BamHI and AscI; band sizes of ~ 2.2 kb and ~ 7.6 kb correspond to application cassette and vector bone, respectively. **B**: lanes 1-5 represent digested pMTL-AM732 with BamHI and AscI; band sizes of ~ 1.6 kb and ~ 7.6 kb correspond to application cassette and vector bone, respectively. **C**: lanes 1-5 represent digested pMTL-AM733 with BamHI and AscI; band sizes of ~ 1.8 kb and ~ 7.6 kb correspond to application cassette and vector bone, respectively.

4.2.5.2 Creation of AM242*::*eCGP123 and TM242*::*eCGP123 strains

Plasmid pMTL-AM731 was transformed into electro-competent cells of *P. thermoglucosidasius* AM242 and TM242 strains by electroporation as described in Methods section 2.10.4. KI of the eCGP123 reporter gene into the regions 2005857-2006296 for TM242 and 2005851-2006290 for AM242 was achieved following the method described in Methods section 2.12. Colonies from each strain were then screened by colony PCR using SNP_G>A_F and SNP_G>A_R primers (detailed primer sequences in Table 2.1) and ran on an agarose gel to identify clean mutants that had the eCGP123 reporter gene knocked in (Figure 4.14). Sanger sequencing of the PCR products confirmed the desired KI of eCGP123. Clean mutants of AM242*::*eCGP123 and TM242*::*eCGP123 were cultured in 10 ml 2SPYNG and incubated at 60 $^{\circ}$C for 12 h x 6 passages for plasmid loss. Colonies with successful plasmid loss were screened again to confirm for mutants by colony PCR.

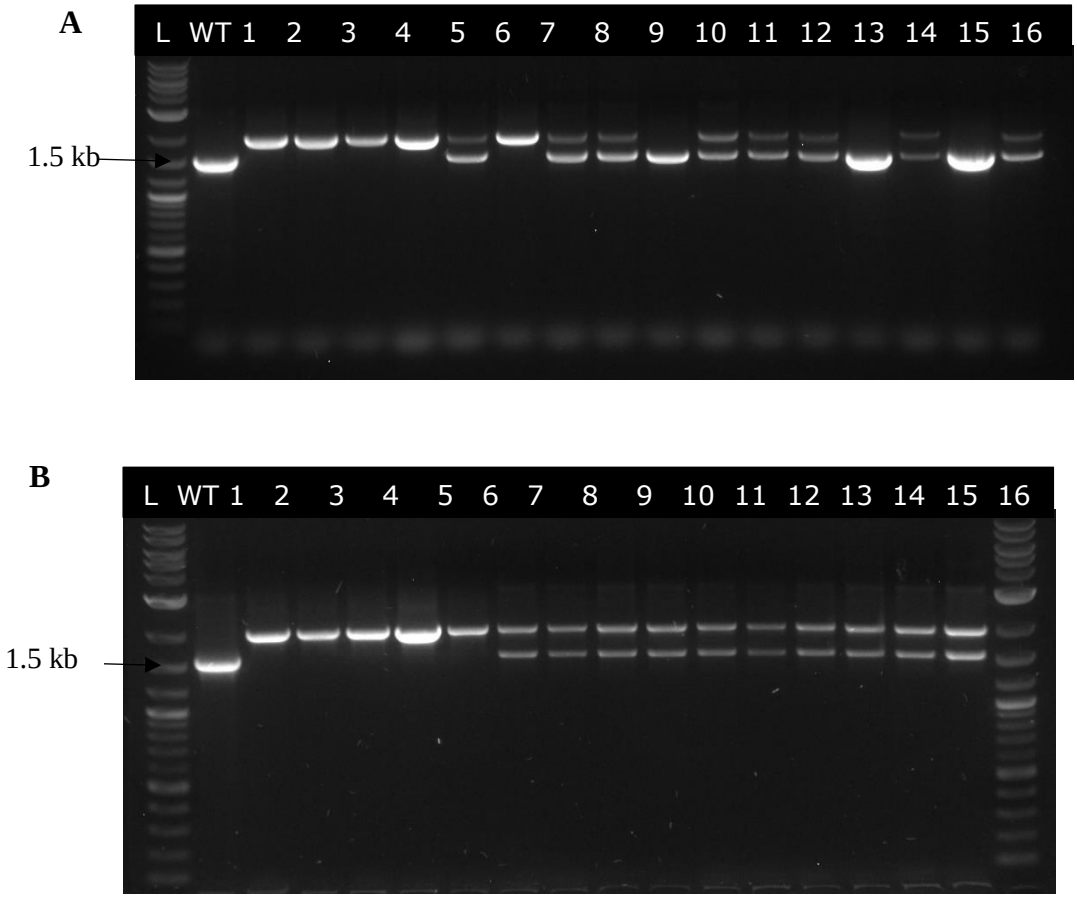


Figure 4.14: Gel electrophoresis image of a colony PCR to screen for AM242*::*eCGP123 and TM242*::*eCGP123 strains. Primers SNP_G>A_F and SNP_G>A_R were used to assess whether the eCGP123 reporter gene had been knocked into the genome of *P. thermoglucosidasius* AM242 **(A)** and TM242 **(B)** at 2005857-2006296 and 2005851-2006290 regions, respectively. A PCR of WT *P. thermoglucosidasius* was performed as a control (~ 1.5 kb). Colonies that have the eCGP123 reporter gene would yield a band at ~ 1.9 kb. L= A 2-log DNA ladder was used to determine the size of each band.

4.2.5.3 Creation of AM242_G>A and TM242_A>G strains

Plasmids pMTL-AM732 and pMTL-AM733 were transformed into electro-competent cells of *P. thermoglucosidasius* AM242::eCGP123 and TM242::eCGP123 strains, respectively, by electroporation as described in the Methods section 2.10.4. Replacements of the eCGP123 reporter gene with the 2005857-2006296 region sequence containing the corrected SNP for TM242 and 2005851-2006290 region sequence with the created SNP for AM242 were achieved following the method described in Methods section 2.11. Colonies from each strain were then screened by colony PCR using SNP_G>A_F and SNP_G>A_R primers (detailed primer sequences in Table 2.1) and electrophoresed on an agarose gel to identify clean mutants that had the eCGP123 reporter gene knocked out and replaced with the original sequence (Figure 4.15). Sanger sequencing of the PCR products confirmed the desired SNP creation in AM242 and correction in TM242. Clean AM242_G>A and TM242_A>G mutants were cultured in 10 ml 2SPYNG and incubated at 60 °C for 12 h x 6 passages for plasmid loss. Colonies with successful plasmid loss were screened again to confirm for mutants by colony PCR.

Whole-genome sequencing of three AM242_G>A and TM242_A>G strains revealed the correct SNP creation and correction, respectively, with three to four off-target mutations (Appendix VII).

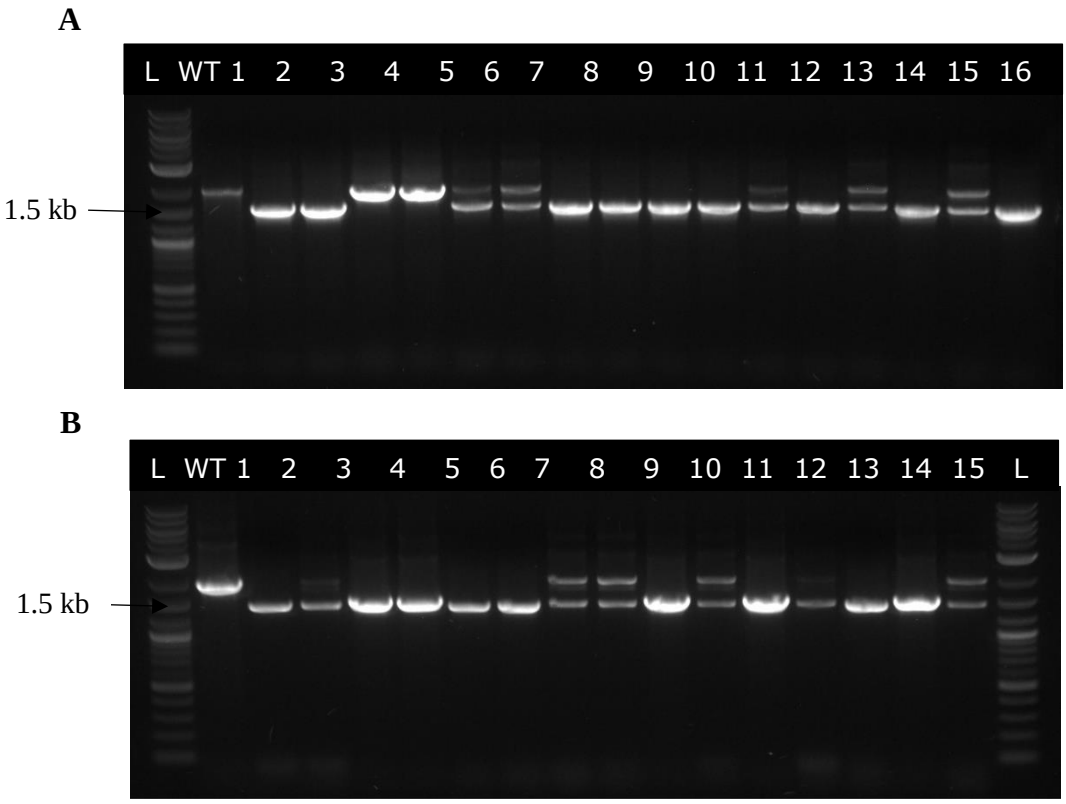


Figure 4.15: Gel electrophoresis image of a colony PCR to screen for AM242_G>A and TM242_A>G strains. Primers SNP_G>A_F and SNP_G>A_R were used to assess whether the eCGP123 reporter gene had been replaced with the 2005857-2006296 region sequence containing the corrected SNP for TM242 **(A)** and 2005851-2006290 region sequence with the created SNP for AM242 **(B)**, respectively. A PCR of AM242::eCGP123 and TM242::eCGP123 strains was performed as a WT control (~ 1.9 kb). Colonies with the eCGP123 reporter gene replaced would yield a band at ~ 1.5 kb. L= A 2-log DNA ladder was used to determine the size of each band.

4.2.5.4 Characterisation of AM242_G>A and TM242_A>G strains

The fermentation profiles of the AM242_G>A, TM242_A>G**,** AM242, and TM242 strains were assessed during growth on ASYE medium with 20 $g \cdot l^{-1}$ glucose, M-ASYE with 10 $g \cdot l^{-1}$ xylose and M-ASYE with 10 $g \cdot l^{-1}$ cellobiose in sealed 50 ml falcon tubes containing 40 ml media. In each case, three independent mutants of each generated strain were analysed. The comparative fermentation profiles of the strains AM242_G>A and AM242 and those of the TM242_A>G and TM242 showed no significant differences (Figure 4.16) using either glucose, cellobiose or xylose as carbon sources. Also, no significant differences were observed between TM242 and AM242 and TM242_A>G and AM242_G>A strains except with cellobiose and xylose at 24 h. These data suggest that the E266K SNP plays no role in the TM242 strain producing high ethanol yield, simultaneously utilising any carbon source used for the fermentation in just 24 h.

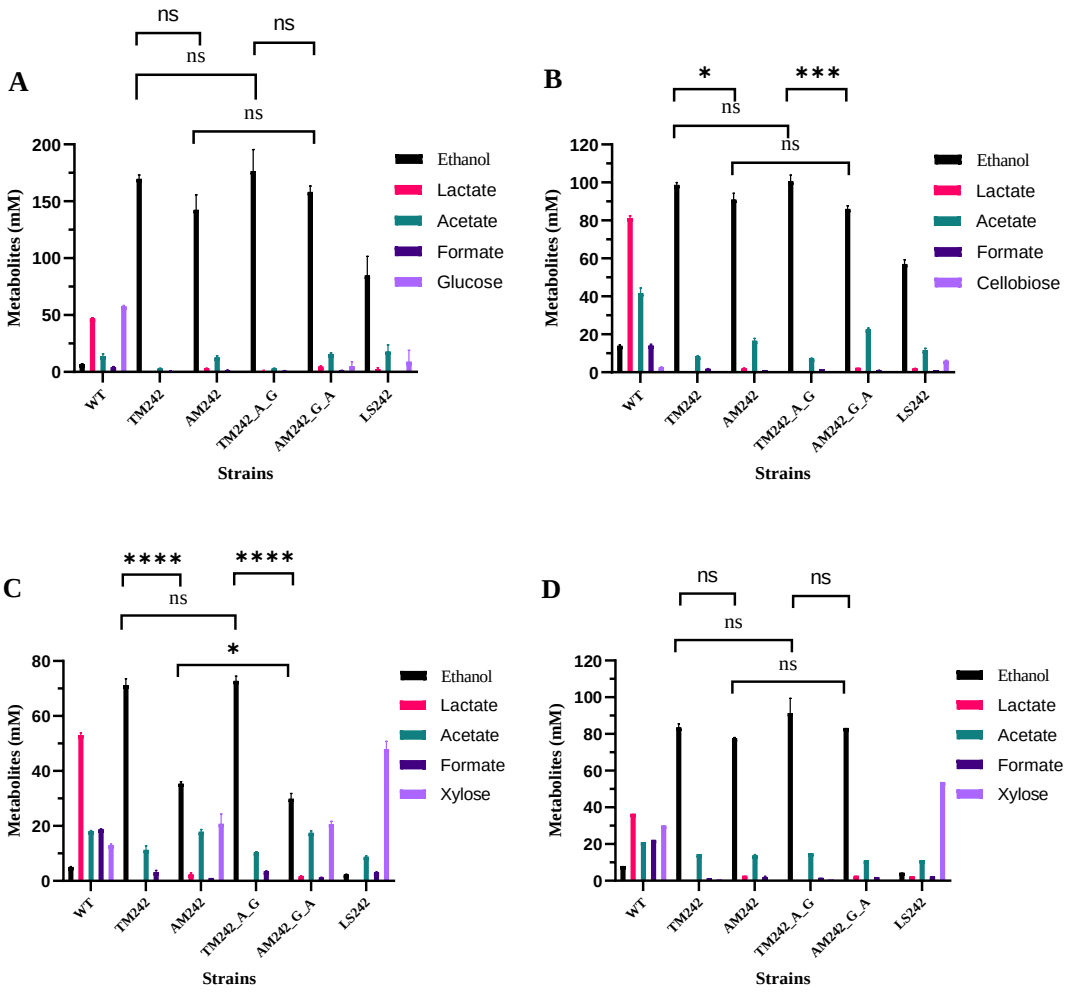


Figure 4.16: Fermentation profile of AM242_G>A and TM242_A>G strains. Fermentation profiles (ethanol, lactate, acetate, formate and glucose/cellobiose/xylose) of the AM242_G>A and TM242_A>G strains, in three biological replicates, as characterised by HPLC after 24 h of fermentation using 40 ml of ASYE with 1% yeast extract and 111 mM glucose **(A)**, M-ASYE with 29.2 mM cellobiose **(B)**, M-ASYE with 66.6 mM xylose after 24 h **(C)**, and M-ASYE with xylose after 48 h **(D)**, in sealed 50 ml falcon tubes incubated at 60 °C, 250 rpm. $P^* \leq 0.05$, $P^{****} \leq 0.0001$, $ns \geq 0.05$, were calculated by One-way ANOVA followed by Tukey's multiple comparisons test, and the results are shown as mean ± SD of three biological replicates.

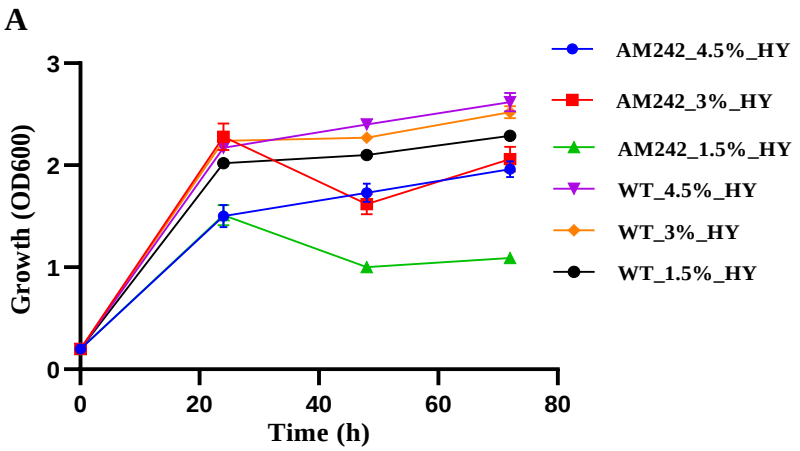
4.2.6 Fermentation of barley straw hydrolysate by *P. thermoglucosidasius* AM242

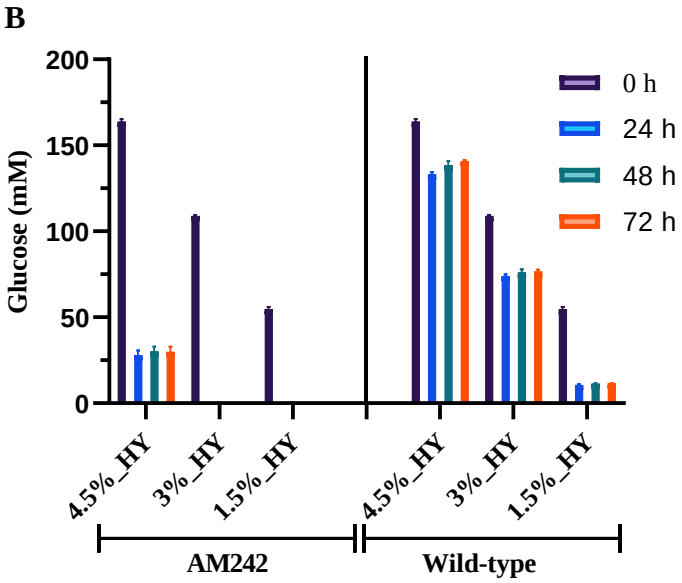
P. thermoglucosidasius AM242 strain's ability to produce ethanol from monosaccharides (glucose and xylose) and oligosaccharides (cellobiose), typical of those found in lignocellulosic biomass, was successfully demonstrated in the previous section of results. Therefore, to investigate ethanol production from lignocellulosic feedstock, fermentation of barley straw hydrolysate was carried out. The McQueen-Mason group, Department of Biology, University of York, supplied the barley straw hydrolysate as part of the MAXBIO (Maximising Conversion Yields in Biorefining) project funded by the BBSRC (Biotechnology and Biological Sciences Research Council) [grant number BB/N022718/1]. The project is a collaborative research between the University of Dundee, the University of York and the University of Nottingham Synthetic Biology Research Centre, which aims to improve conversion yields from plant biomass into chemical products and biofuels, focusing on sugar production, sugar release and sugar conversion. The barley straw hydrolysate contained approximately 60 $g \cdot l^{-1}$ glucose (6%) and 30 $g \cdot l^{-1}$ xylose (3%), equivalent to 9% total sugar.

P. thermoglucosidasius AM242 alongside *P. thermoglucosidasius* wild-type as a control was grown in a 40 ml modified-ASYE medium containing 4.5% (30 $g \cdot l^{-1}$ glucose and 15 $g \cdot l^{-1}$ xylose), 3% (20 $g \cdot l^{-1}$ glucose and 10 $g \cdot l^{-1}$ xylose) and 1.5% (10 $g \cdot l^{-1}$ glucose and 5 $g \cdot l^{-1}$ xylose) barley straw hydrolysates, respectively, as sugar sources at 60 $^{\circ}$C. The fermentation was performed as described in section 4.2.3. Growth was monitored by measuring OD_{600}, and culture supernatants from these strains were used for the measurement of ethanol production and sugar consumption by HPLC at growth time points of 0, 24, 48, and 72 h as described in the Methods section 2.15.1. It was observed that both AM242 and wild-type strains could effectively grow in all the

concentrations of the barley straw hydrolysate (Figure 4.17 A). *P. thermoglucosidasius* AM242 strain reached an OD_{600} of 1.50, 2.28, and 1.51 in 4.5%, 3%, and 1.5% hydrolysates in 24 h and then decreased in 1.5% and 3% to an OD_{600} value of 1.09 and 2.06 after 72 h while reaching the maximum OD_{600} value of 1.96 in 4.5% hydrolysate in the same period. For the wild-type strain, OD_{600} values of just above 2.0 were observed in all concentrations of hydrolysates after 24 h before reaching the maximum OD_{600} values of 2.62, 2.52 and 2.29 in 4.5%, 3% and 1.5% hydrolysates at 72 h (Figure 4.17 A).

The HPLC analysis of culture supernatant showed that around 29.9 mM residual glucose remained after 72 h of fermentation in 4.5% hydrolysate for AM242. In contrast, all of the available glucose was consumed entirely after 24 h fermentation using 3% and 1.5% hydrolysates for the AM242 strain (Figure 4.17 B). No xylose was observed at the end of the 72 h fermentation using 1.5% hydrolysate in AM242 strain, whereas 11.3 mM and 68.5 mM residual xylose remained in 3% and 4.5% hydrolysates after an equivalent period (Figure 4.17 C). By contrast, high residual glucose and xylose were observed for the wild-type strain after the fermentation (Figure 4.17 B & C). There was no consumption of xylose for the wild-type. It was also observed that glucose and xylose consumption occurred concurrently in the AM242 strain, with the latter occurring more slowly. This might be explained by a loosely controlled carbon catabolite repression (CCR) under fermentative conditions in *P. thermoglucosidasius* NCIMB 11955 (which is the same as the *P. thermoglucosidasius* DSM 2542 strain) (Liang *et al.*, 2022).


A
Growth (OD600)
Time (h)
AM242_4.5%_HY
AM242_3%_HY
AM242_1.5%_HY
WT_4.5%_HY
WT_3%_HY
WT_1.5%_HY


B
Glucose (mM)
0 h
24 h
48 h
72 h
4.5%_HY
3%_HY
1.5%_HY
AM242
Wild-type

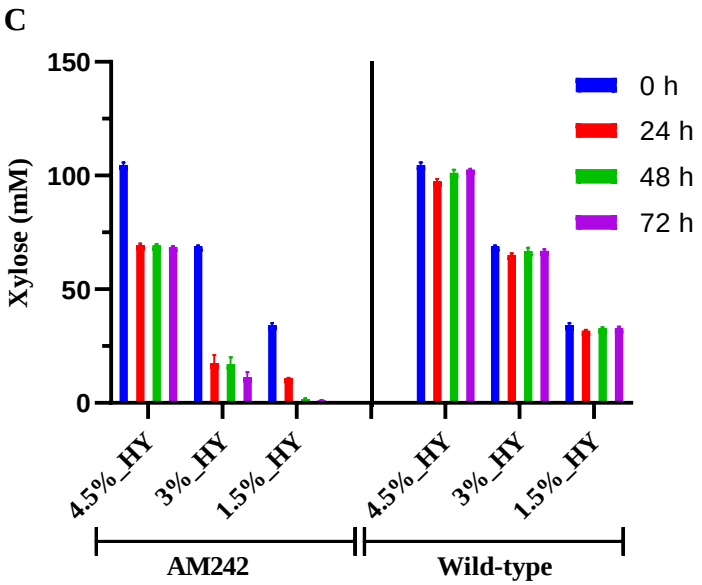


Figure 4.17: Growth and sugar consumption of *P. thermoglucosidasius* AM242 and wild-type strains on barley straw hydrolysates. (A) Growth of *P. thermoglucosidasius* AM242 and wild-type strains on 4.5% (166.5 mM glucose and ~ 100 mM xylose), 3% (111 mM glucose and 66.6 mM xylose), and 1.5% (55.5 mM glucose and 33.3 mM xylose) barley straw hydrolysate. Growth at 0, 24, 48, and 72 h time points was followed by measuring OD_{600} using a spectrophotometer. The results are shown as mean ± SD of three biological replicates. **(B)** Glucose consumption by *P. thermoglucosidasius* AM242 and wild-type strains: residual glucose concentrations (mM) were quantified by HPLC. The results are shown as mean ± SD of three biological replicates. **(C)** Xylose consumption by *P. thermoglucosidasius* AM242 and wild-type strains: residual xylose concentrations (mM) were quantified by HPLC. The results are shown as mean ± SD of three biological replicates.

P. thermoglucosidasius AM242 demonstrated a higher efficiency in the fermentation of the barley straw hydrolysate to produce ethanol than the wild-type. Highest ethanol concentrations of 279.2 mM, 281.9 mM and 130.3 mM were achieved using AM242 strain with 4.5%, 3%, and 1.5% hydrolysates, respectively, after incubation at 60 ˚C for 24 h (Figure 4.18 A). The wild-type strain produces 4.5 mM, 7.2 mM, and 10.1 mM after an equivalent period using 4.5%, 3%, and 1.5% hydrolysates, respectively (Figure 4.18 A). The concentration of ethanol in both AM242 and wild-type strains remained relatively constant up to 72 h. The ethanol yield achieved with the *P. thermoglucosidasius* AM242 strain in this study was equivalent to 0.43, 0.48, and 0.45 g/g of total sugar for 4.5%, 3%, and 1.5% barley straw hydrolysate, respectively (Figure 4.18 B). The yields correspond to 84.1%, 94%, and 88.1% of the theoretical yield based on the available sugars in the barley straw hydrolysate. Lower ethanol yields of 0.04, 0.05, and 0.06 g/g of total sugar with 4.5%, 3%, and 1.5% barley straw hydrolysate, respectively, were observed for the wild-type strain (Figure 4.18 B).

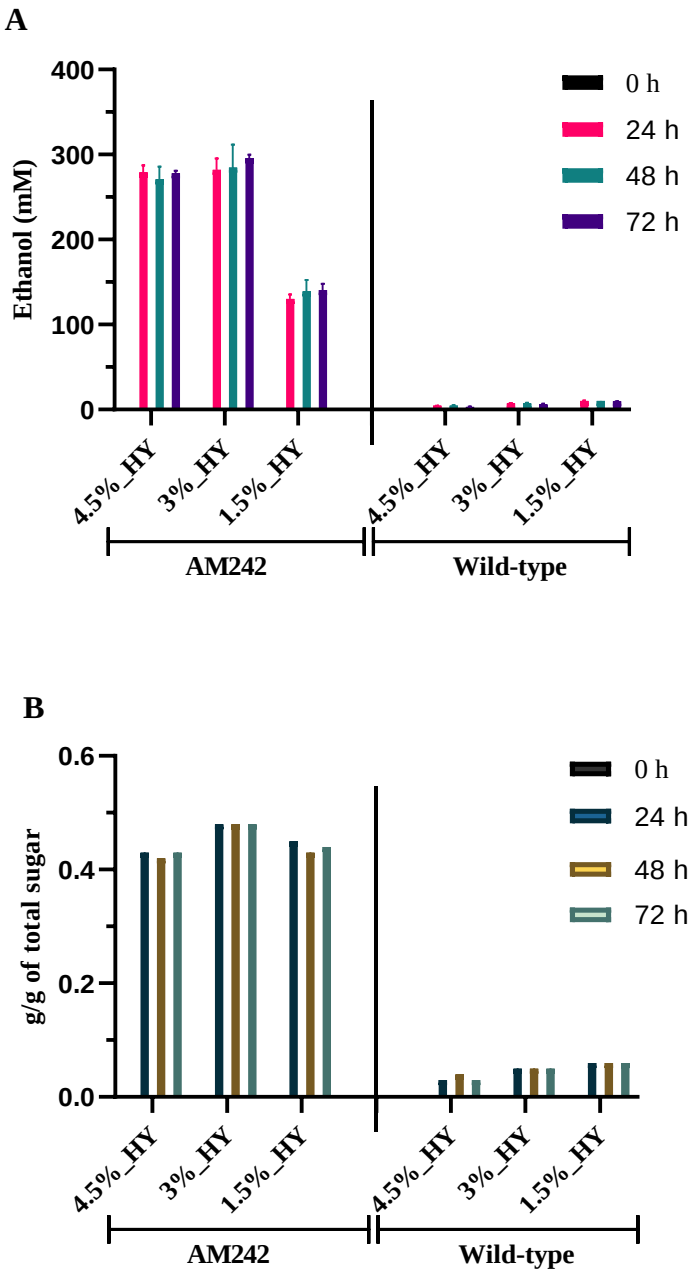


Figure 4.18: Production of ethanol by *P. thermoglucosidasius* AM242 and wild-type strains using barley straw hydrolysate. (A) Ethanol titres by *P. thermoglucosidasius* AM242 and wild-type. Ethanol concentrations of AM242 and wild-type strains in three replicates were characterised by HPLC at 0, 24, 48, and 72 h of fermentation using 40 ml of M-ASYE with 4.5% (166.5 mM glucose and ~ 100 mM xylose), 3% (111 mM glucose and 66.6 mM xylose), and 1.5% (55.5 mM glucose and 33.3 mM xylose) barley straw hydrolysate in a sealed 50 ml falcon tubes incubated at

60 °C, 250 rpm. The results are shown as mean ± SD of three biological replicates. **(B)** Ethanol yields by the AM242 and the wild-type strains. Calculated based on ethanol produced and total sugar consumed.

4.3 Discussion

The genetic manipulation of *Parageobacillus* spp. compared to other bacteria such as *B. subtilis* and *E. coli* is constrained by limited tools. However, significant progress is being made in enhancing the genetic tools for this group of microbes (Kananaviciute and Citavicius, 2015, Reeve *et al.*, 2016, Sheng *et al.*, 2017, Lau *et al.*, 2021). Currently, genetic engineering in *P. thermoglucosidasius* is achieved using allelic exchange (AE), allele-coupled exchange (ACE) (Sheng *et al.*, 2017), and CRISPR/Cas9 (Lau *et al.*, 2021), respectively. However, both tools are restricted as AE is inefficient, and ACE is a location-specific genome editing method for gene KI. Whereas the CRISPR/Cas9 system is specific and efficient in genetic engineering, the system developed for *P. thermoglucosidasius* used a constitutive promoter to control Cas9 production (Lau *et al.*, 2021) and was characterised as suffering from low transformation and/or editing efficiencies. Therefore, implementing an inducible CRISPR/Cas9 genome editing system would enhance genetic engineering within *P. thermoglucosidasius*.

Here, a novel inducible CRISPR/Cas9 system based on theophylline-responsive riboswitch E was implemented to facilitate efficient KO and KI in *P. thermoglucosidasius*. This was demonstrated by the generation of engineered strains equivalent to a previously constructed ethanol production strain – TM242 (Cripps *et al.*, 2009) and LS242 (Sheng *et al.*, 2017). The engineering involved the upregulation of the pyruvate dehydrogenase (*pdhA*) gene by replacing its native promoter with lactate

dehydrogenase promoter in an *ldhA* deficient strain and deleting the pyruvate formate lyase (*pflB*) gene, with the highest mutant generation efficiencies of approximately 75% and 100%, respectively. This study is in line with a previous study reporting the construction and characterisation of an inducible gene expression element composed of theophylline-responsive riboswitch E and a constitutive promoter in *B. subtilis*, together with efficient inducible heterologous gene expression (Cui *et al.,* 2016). Our results are also supported by the previously published study by Cañadas *et al.* (2019), reporting the versatility of the riboswitch based CRISPR/Cas9 genome editing system (RiboCas), which was demonstrated successfully in several species of *Clostridia*, including *C. sporogenes, C. pasteurianum, C. difficile, C. botulinum* and *C. autoethanogenum* (Seys *et al*., 2020). To the best of our knowledge, this study is the first implementation of a theophylline-responsive CRISPR/Cas9 genome editing system in *P. thermoglucosidasius* for genome engineering. Thus, the method represents a considerable improvement and adds to the existing tools available for the metabolic engineering of *P. thermoglucosidasius* and thermophilic bacilli generally to produce fuels and value-added chemicals.

Previous studies by Sheng *et al.* (2017) using the developed ACE technique for TM242 strain (Cripps *et al.,* 2009) re-engineering resulted in equivalent strain (LS242) that does not match TM242 in terms of ethanol production during micro-aerobic fermentation unless the medium was supplemented with acetic acid. Comparing the whole genome sequence of the engineered strains and TM242 revealed SNPs and Indels. Therefore, this observation makes it imperative to enhance the screening and selection methods to eliminate unintended mutations (Sheng *et al.,* 2017). Here, the engineering of multiple, independent strains during each modification step was explored, coupled with whole-genome sequencing and analysis. The ethanol production by the AM242 strains across

the six independent batches showed significant differences (51.2–180.1 mM). However, strains AM242_2_5, AM242_4_1, AM242_4_2, AM242_4_4, and AM242_4_5 produced ethanol yields comparable to that of TM242 with AM242_2_5 and AM242_4_4 producing ethanol from glucose at yields of 1.6 mol/mol corresponding to 80% of the theoretical value of 2 mol/mol of glucose. Also, the equivalent strains produced higher amounts of ethanol than LS242. In addition to glucose, the strains showed a rapid cellobiose metabolism to produce ethanol with the highest ethanol yield of 3.2 mol/mol of cellobiose equivalent to 80% of the theoretical maximum. Because of the preferred temperature range of available commercial cellulases and their limited β-glucosidase activity (Cripps *et al.,* 2009), a more effective simultaneous saccharification and fermentation could be achieved with these strains. Although the metabolism of xylose by these strains was slower than hexoses, utilising all the xylose in 48 h with ethanol yield of 78% of the theoretical yield compared to 24 h by TM242, it was interesting to note that both pentoses and hexoses typical of a biomass hydrolysate were metabolised to produce ethanol. Similar slow xylose metabolism was observed for TM242 (Cripps *et al.,* 2009), albeit faster than observed in this study. The comparative fermentation profile of the AM242 strains observed in this study emphasises the need for the generation of multiple, independent strains for each step in strain engineering to improve the screening and selection procedures.

Although this study generated a strain equivalent to TM242, the observed differences between the thirty-six independent AM242 strains in ethanol yields prompted a comparison of the genome sequences of the engineered strains. This analysis revealed that mutations in the form of SNPs and Indels were relatively common in the strains sequenced. Due to the multiple SNPs involved, it is impossible to determine why the AM242 strains produce different ethanol titres. However, a noticeable mutation in

TM89, the progenitor strain of both TM242 and AM242, was the presence of two non-synonymous SNPs (T65P, D119N) in the *aprt* (encoding adenine phosphoribosyltransferase) gene. While the P65 is not conserved, the D119 residue is part of a conserved 9-amino acid region (DDLLATGGT) (Sheng *et al.*, 2017). In a study by Zhou *et al.* (2016), the loss of *aprt* function in a mutant of *P. thermoglucosidasius* was observed to be due to the presence of a T123I SNP within this conserved domain. The deliberate deletion of the *aprt* gene improved both the glucose and cellobiose consumption and ethanol production in a Δ*ldh*Δ*pfl* strain of *P. thermoglucosidasius* 95A1 (Zhou *et al.*, 2016). The two mutations in the *aprt* gene of TM242 and AM242, one of which is located in the conserved domain, in which the presence of a different SNP by *P. thermoglucosidasius* 95A1 (Zhou *et al.*, 2016) led to the loss of the *aprt* function, could be the reason why the two strains produce high ethanol titres compared to the strain LS242. Furthermore, the deliberate deletion of the *aprt* gene in strain LS242 still failed to produce high ethanol yields (Sheng *et al.*, 2017). This could be due to other mutations, such as in the genes encoding heat-inducible transcription repressor (HcrA) and 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexadiene-1-carboxylate (SEPHCH C) synthase (also known as MenD), concealing the effect of *aprt* gene inactivation (Sheng *et al.*, 2017). Therefore, based on these observations, it was concluded that the TM242 produces high ethanol close to the theoretical maximum due to the SNPs/Indels it acquired during its generation.

Even though a strain equivalent to TM242 was generated, further investigations were made as to why the strain is an effective ethanol producer, simultaneously utilising any carbon source within 24 h of the fermentation. This was accomplished by seeking to make one of the AM242 strains resemble the TM242 strain as closely as possible. According to Sheng *et al.* (2017), the TM242 picked up one SNP (E266K) on a gene

encoding DNA methylase N-4 (otherwise known as type III restriction-modification system methylation subunit) and one Indel on a non-coding region on the small megaplasmid (pNCI002). Therefore, it was sought to make the E266K SNP in AM242 and corrected the SNP in TM242. However, neither the SNP creation in AM242_G>A nor the SNP correction in TM242_A>G strains showed significant differences in sugar consumption and ethanol production compared to the AM242 and TM242 strains. This suggests that the E266K SNP plays no role in the TM242 for sugar assimilation and ethanol production.

P. thermoglucosidasius is known to grow and ferment a wide range of monosaccharides and oligosaccharides sugars, typical of those derived from lignocellulosic substrates (Cripps *et al.*, 2009, Raita *et al.*, 2016). Therefore, an assessment of the ability and efficiency of *P. thermoglucosidasius* AM242 to grow and ferment barley straw hydrolysate for bioethanol production was undertaken. *P. thermoglucosidasius* AM242 efficiently fermented the barley straw hydrolysate and produced ethanol with yields equivalent to 0.48 g/g total sugar or 94% of the theoretical yield based on sugars consumed with 3% hydrolysate, indicating their capability to convert sugars in the hydrolysate to ethanol. The ethanol yield obtained in this study was relatively similar compared to the yield of 0.47 g/g sugar or 92% of the theoretical yield reported for palm kernel cake (PKC) using *P. thermoglucosidasius* TM242 (Raita *et al.*, 2016). The current study shows that *P. thermoglucosidasius* AM242 is capable of ethanol production from lignocellulosic biomass via separate hydrolysis and fermentation (SHF). To reduce the cost of enzyme addition and eliminate the time needed for separate enzyme hydrolysis, the results obtained here suggest that a CBP approach with AM242 strain that produce glycoside hydrolases may be a future prospect for ethanol production from lignocellulosic feedstocks given that it can convert both monosaccharides and

oligosaccharides present. Therefore, the AM242 strain represents an excellent candidate for developing an effective process for fermentative bioethanol production from lignocellulosic biomass.

4.4 Conclusions

In the present study, a theophylline-responsive CRISPR/Cas9 was successfully implemented as a genome-editing tool in *P. thermoglucosidasius*. As a proof of concept of the method, thirty-six independent strains equivalent to the industrial production strain, TM242 (Δ*ldh*, pdh^{up}, Δ*pfl*), were engineered. Production of ethanol matched that of TM242 in some of the AM242 strains. TM242 strain produces ethanol with efficiencies that approach the theoretical maximum due to the SNPs/Indels it acquired during its generation. Genome sequencing suggested that additional off-target mutations in the form of SNPs and Indels occurred and might have played a role in the fermentation profile of the AM242 strains generated. These findings, therefore, further emphasises the need for the generation of multiple, independent strains for each step in strain engineering, followed by whole-genome sequencing of the engineered strains. Ethanol yields of greater than 90% of the theoretical maximum was achieved by a *P. thermoglucosidasius* AM242 strain from barley straw hydrolysate. This provides the avenue for developing an effective and economic process for bioethanol production from lignocellulosic feedstocks.

CHAPTER 5

5.0 ENGINEERING *P. THERMOGLUCOSIDASIUS* FOR THE PRODUCTION OF 3-HYDROXYPROPIONIC ACID

5.1 Introduction

3-hydroxypropionic acid (3-HP, $C_3H_6O_3$, MW 90.08) is a β-hydroxy acid and a structural isomer of 2-hydroxypropionic acid (lactic acid). According to the United States Department of Energy, 3-HP is among the top 12 high-value-added chemicals derived from biomass (Werpy and Petersen, 2004). 3-HP can be used to produce chemicals such as acrylamide, acrylic acid, acrylonitrile, 1,3-propanediol, propiolactone, malonic acid, and 3-HP-containing polymers. These chemicals with high value are used in various applications, such as making adhesives, polymers, plastic packaging, cosmetics, textiles, cleaning agents, fibres, and resins (Matsakas *et al.*, 2018).

At present, the production of 3-HP is based on petrochemical routes, but the costs of these are high and unsustainable because they are derived from fossil resources (Jiang *et al.,* 2009; Kumar *et al.,* 2013). Several prokaryotes and eukaryotes produce 3-HP either as an intermediate or end product via various metabolic pathways (Kumar *et al.,* 2013). However, a high titre, yield, and productivity are required for commercial production. *Lactobacillus* sp. has been reported to co-produce 3-HP and 1,3-propanediol (PDO) using glycerol as substrate (Sobolov and Smiley, 1960; Talarico *et al.,* 1988; Garai-Ibabe *et al.,* 2008). Additionally, 3-HP was reported to be an intermediate in the 3-hydroxypropionate cycle of a thermophilic photosynthetic bacterium *C. aurantiacus* (Holo, 1989). The 3-hydroxypropionate cycle was also reported in other microorganisms such as *A. ambivalens*, *Sulfolobus* sp. strain VE6, *A. brierleyi*, *M. sedula*, and *S. metallicus* (Hügler *et al.,* 2003).

To commercialise bio-based 3-HP production and achieve higher titre, productivity and yield, several metabolic routes proposed to produce 3-HP from substrates such as

glucose or glycerol have been implemented in enteric bacterial species (*E. coli, K. pneumoniae*), yeast (*S. cerevisiae, S. pombe*), and other microorganisms (*C. glutamicum, B. subtilis, L. reuteri*) (Matsakas *et al.,* 2018). Substantial progress has been made in the bio-based production of 3-HP using organic substrates such as glycerol and glucose from biosynthetic pathways (Rathnasingh *et al.,* 2012; Ashok *et al.,* 2013; Zhou *et al.,* 2013a; Li *et al.,* 2016; Lim *et al.,* 2016; Kalantari *et al.,* 2017; Jiang *et al.,* 2018; Jers *et al.,* 2019). Several biosynthetic pathways for the production of 3-HP from glucose through intermediates such as lactate, malonyl-CoA, β-alanine, propionate/propionyl-CoA, have been proposed and patented by a US-based company, Cargill (Jiang *et al.,* 2009; Kumar *et al*., 2013).

Previous studies in our group by a prior PhD student, Bashir (2018), implemented the malonyl-CoA pathway in *P. thermoglucosidasius* NCIMB 11955 to produce 3-HP from glucose. In the study, the multi-subunit encoding biotin-dependent acetyl-CoA carboxylase gene (*accCBpccBbpl*) from *M. sedula* was integrated as an operon at the *pyrE* locus of *P. thermoglucosidasius*. In parallel, either a gene coding for a bifunctional or monofunctional malonyl-CoA reductase (MCR) from either *C. aurantiacus, M. sedula* or *Sulfolobus tokodaii*, in combination with genes encoding a malonate semialdehyde reductase (MSR) from either *M. sedula* or *Sulfolobus solfataricus*, were expressed as MCR–MSR operons from an autonomous plasmid to complete the 3-HP production pathway. The recombinant strains produced 3-HP ranging from 3.0–3.8 mM when grown aerobically in shake flasks with glucose as a carbon source. Therefore, the present study adopted the strategy of using the native acetyl-CoA carboxylase in *Parageobacillus* (involved in fatty acid biosynthesis) and introduced the genes encoding the various combinations of MCR and MSR genes on autonomous plasmids, which were previously produced by Bashir (2018).

The production of 3-HP via the malonyl-CoA pathway (Figure 5.1) involves the carboxylation of acetyl-CoA to malonyl-CoA catalysed by a biotin-dependent acetyl-CoA carboxylase (ACC), and the conversion of malonyl-CoA to malonate semi-aldehyde catalysed by MCR, and finally, the malonate semi-aldehyde is converted to 3-HP by MSR.

The malonyl-CoA route for 3-HP production has some advantages compared to the other six pathways, such as (1) several sugars (C_5 and C_6) derived from lignocellulosic biomass can be used as the substrate because acetyl-CoA is a common intermediate of sugar metabolism, (2) high 3-HP yield from glucose is anticipated since the production of 3-HP from glucose is both redox balanced and thermodynamically favourable, (3) non-requirement of coenzyme B_{12} as in the case of the glycerol dehydratase-dependent pathway (Rathnasingh *et al.*, 2012).

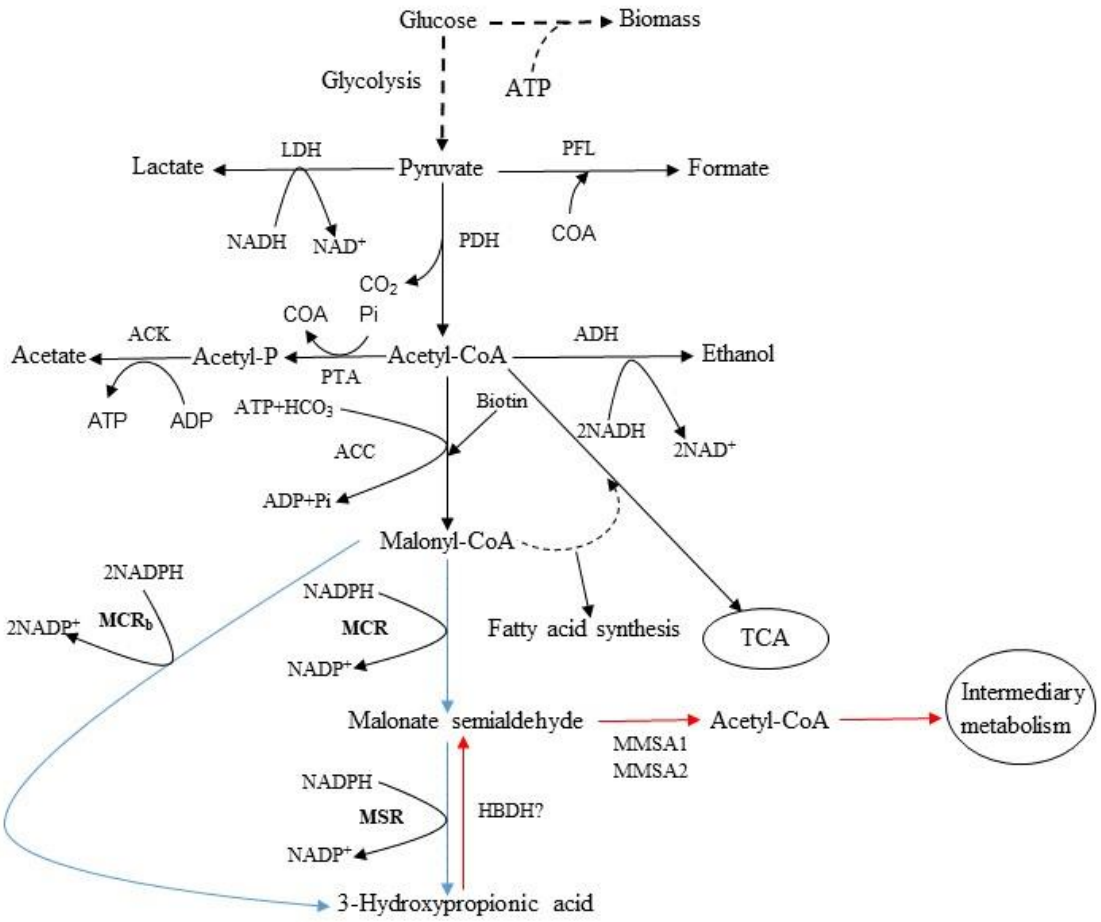


Figure 5.1: Metabolic pathway for 3-HP production from glucose via malonyl-CoA. The genetic modification steps are illustrated in blue and the heterologous enzymes in bold. 3-HP degradation pathways are shown in red. Multiple steps involved in native metabolic pathways are shown in broken arrows. LDH: lactate dehydrogenase, PFL: pyruvate formate-lyase, PTA: phosphotransacetylase, ACK: acetate kinase, ADH: aldehyde/alcohol dehydrogenase, PDH: pyruvate dehydrogenase, ACC: acetyl-CoA carboxylase, MCR_b: bifunctional malonyl-CoA reductase, MCR: malonyl-CoA reductase, MSR: malonate semialdehyde reductase. HBDH: 3-hydroxyisobutyrate dehydrogenase. Question mark indicate the role of the enzyme has not been experimentally demonstrated. MMSA1 and MMSA2: (methyl) malonate semialdehyde dehydrogenases 1 and 2. Modified from Rathnasingh *et al.* (2012).

5.1.1 Pathways for 3-HP assimilation

To achieve biological production of the platform chemical- 3-HP it is essential to ensure that the producing organism does not degrade the produced 3-HP at any stage of the production process (Arenas-López *et al.,* 2019).

Two pathways have been proposed and reported for 3-HP consumption in several organisms: an oxidative pathway and a reductive pathway (Schneider *et al.,* 2012; Otzen *et al.,* 2014; Zhou *et al.,* 2014; Arenas-López *et al.,* 2019). The reductive pathway is CoA-dependent and involves the conversion of 3-HP to propionyl-CoA through 3-hydroxypropionyl-CoA and acrylyl-CoA intermediates (Arenas-López *et al.,* 2019). This pathway has been proposed and studied in organisms like the photoheterotrophic bacterium *Rhodobacter sphaeroides* (Schneider *et al.,* 2012). Acrylyl-coenzyme A reductase was involved in the assimilation of 3-HP via the reductive conversion of acrylyl-CoA to propionyl-CoA in this organism (Asao and Alber, 2013). It has also been studied in the thermophilic photosynthetic bacterium *C. aurantiacus* as part of the 3-HP cycle used for CO_2 fixation with propionyl-CoA synthase catalysing the reaction (Alber and Fuchs, 2002). Furthermore, *Methylobacterium extorquens* AM1 was reported to re-assimilate 3-HP via the reductive route (Yang *et al.,* 2017).

In contrast, the oxidative pathway is CoA-independent and shorter. Here, 3-HP is first oxidised to malonate semialdehyde via 3-hydroxypropionate dehydrogenase, which would then be decarboxylated and converted to acetyl-CoA before entering the central carbon metabolism (Arenas-López *et al.,* 2019). Zhou *et al.* (2013a) reported the development of recombinant *Pseudomonas denitrificans* for 3-HP production from glycerol and showed that the organism utilises 3-HP as a carbon source for growth resulting in low yields. Further analysis of metabolites by GC/MS detected the presence

of malonate and methyl malonate, suggesting that the 3-HP degradation occured via the oxidative route (Zhou *et al.,* 2013a). Two putative dehydrogenases, a 3-HP dehydrogenase and a 3-hydroxyisobutyrate dehydrogenase, have been identified as responsible for 3-HP assimilation in *P. denitrificans* and deleting the two genes could eliminate 3-HP consumption by *P. denitrificans* (Zhou *et al.,* 2014). Several studies reported various dehydrogenases involved in the oxidative conversion of 3-HP to malonate semialdehyde. Their subsequent deletion led to strains incapable of 3-HP utilisation as carbon and energy sources (Zhou *et al.,* 2013b; Lee *et al.,* 2014; Arenas-López *et al.,* 2019). The existence of these genes as operons induced by 3-HP has been shown, such as a C4 operon which contains a (methyl)malonate semialdehyde dehydrogenase (*mmsA*) and a 3-hydroxyisobutyrate dehydrogenase gene (*hbdH*), and a C3 operon consisting of a single 3-hydroxyisopropionate dehydrogenase (*hpdH*) in *P. denitrificans* (Zhou *et al.,* 2014; Zhou *et al.,* 2015). Additionally, three different (methyl)malonate semialdehyde dehydrogenase (*mmsA1*, *mmsA2*, *mmsA3*) gene clusters existing as three separate *mmsA* operons were identified and confirmed to play a role in 3-HP metabolism in *C. necator* H16 (Arenas-López *et al.,* 2019). The oxidative route could be the most widely used as it is reported in other organisms such as *B. cereus, Candida albicans,* and *C. necator* H16 (Otzen *et al.,* 2014; Yao *et al.,* 2010; Arenas-López *et al.,* 2019). Consequently, a search of the *P. thermoglucosidasius* NCIMB 11955 genome, as published by Sheng *et al.* (2016), reveals two putative (methyl)malonate semialdehyde dehydrogenase genes, *mmsA1* (BCV53_10115) and *mmsA2* (BCV53_11155), 3-hydroxyisobutyrate dehydrogenase gene (*hbdH*) (BCV53_14255), and several other aldehyde dehydrogenases. This opened the possibility that 3-HP assimilation in *P. thermoglucosidasius* NCIMB 11955 could follow the oxidative route. Therefore, this study attempted to investigate the native 3-

HP catabolism in *P. thermoglucosidasius* NCIMB 11955 to engineer a strain incapable of 3-HP utilisation.

5.1.2 Chapter aims

This chapter aimed to (1) engineer *P. thermoglucosidasius* NCIMB 11955 for 3-HP production via the malonyl-CoA pathway. The strategy adopted was to use the native acetyl-CoA carboxylase present in *Parageobacillus* (essential reaction involved in fatty acid biosynthesis) and introduce the genes encoding the various combinations of MCR and MSR genes on autonomous plasmids. (2) Investigate the native catabolism of 3-HP by *P. thermoglucosidasius* NCIMB 11955 with the aim of engineering a strain incapable of utilising 3-HP as a source of carbon and energy to be used as a chassis for its production.

5.2 Results

Various recombinant *P. thermoglucosidasius* NCIMB 11955 strains were engineered for 3-HP production via the malonyl-CoA pathway by heterologously expressing different MCR–MSR operons on autonomous plasmids and using the native ACC present in the organism. The recombinant strains were characterised for 3-HP production in shake flask conditions. The monofunctional MCR and MSR were previously constructed as operons in different combinations based on the organism gene source by Bashir (2018), as shown below in Figure 5.2.

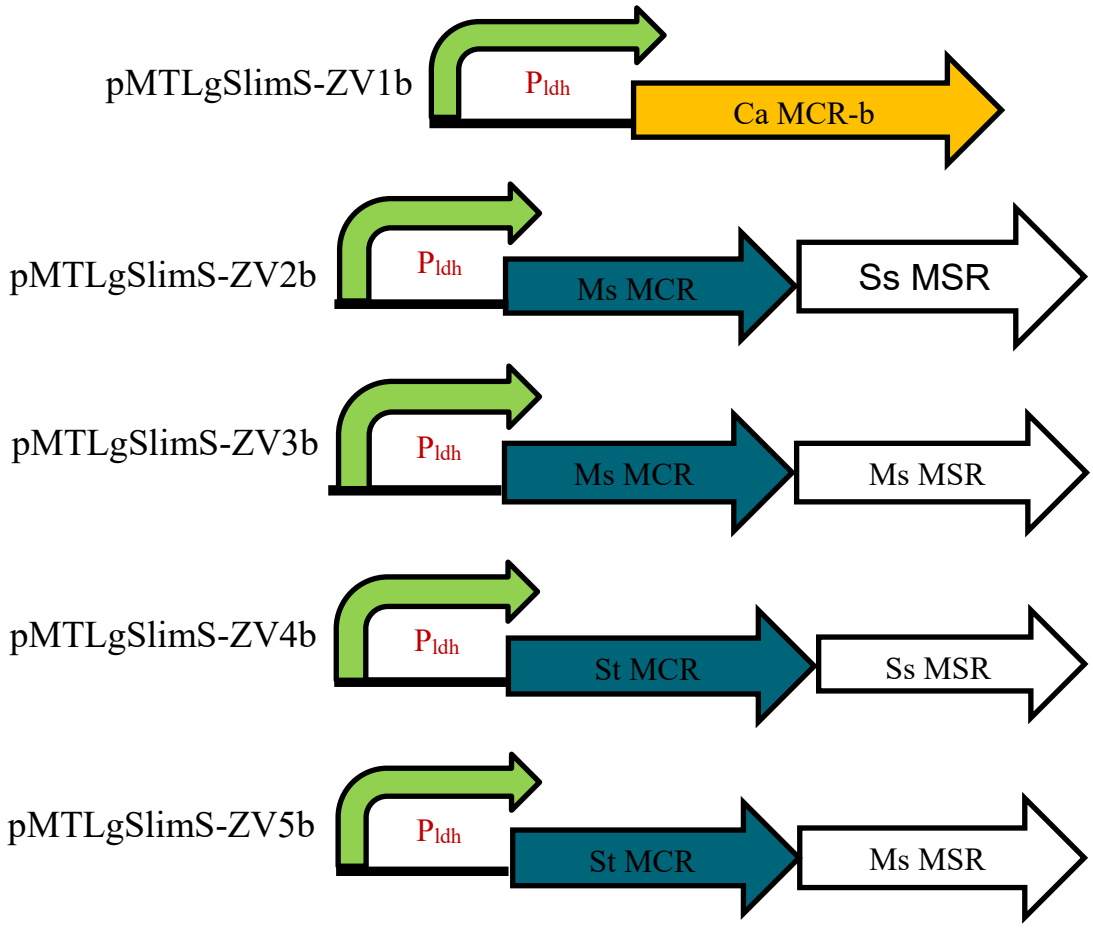


Figure 5.2: Schematic representations of operons as plasmids in different combinations of *mcr* and *msr* based on gene source. P_{ldh}: lactate dehydrogenase promoter from *Geobacillus stearothermophilus*; MCR-b: bifunctional Malonyl-CoA reductase; MCR: Malonyl-CoA reductase; MSR: Malonate semialdehyde reductase; Ca: *C. aurantiacus*; Ms: *M. sedula*; Ss: *S. solfataricus*; St: *S. tokodaii*. Plasmids previously constructed by Bashir (2018).

5.2.1 Engineering recombinant *P. thermoglucosidasius* strains for 3-HP production

To achieve 3-HP production in *P. thermoglucosidasius* via the malonyl-CoA pathway, plasmids pMTLgSlimS-ZV1b, pMTLgSlimS-ZV2b, pMTLgSlimS-ZV3b, pMTLgSlimS-ZV4b, and pMTLgSlimS-ZV5b previously made by Bashir (2018), were transformed into electro-competent cells of *P. thermoglucosidasius* NCIMB 11955, thereby resulting in recombinant *P. thermoglucosidasius* KT1, *P. thermoglucosidasius* KT2, *P. thermoglucosidasius* KT3, *P. thermoglucosidasius* KT4, and *P. thermoglucosidasius* KT5 strains, respectively as described in Methods section 2.10.4. To confirm the presence of the respective plasmids in the recombinant strains, colony PCR was carried out using plasmid-based primers ColE1_F2 and RepB_R1 (exact primer sequences in Table 2.1) as described in Methods section 2.9.4. An amplified DNA of either ~ 4 kb or ~ 3 kb was observed, corresponding to the size of MCR-b or MCR-MSR operons on the autonomous plasmids in the recombinant *P. thermoglucosidasius* KT1, KT2, KT3, KT4, and KT5 strains (Figure 5.3 A-C). Furthermore, the PCR product was purified, followed by Sanger sequencing to verify the correct nucleotide sequence of the MCR-b and MCR-MSR operons.

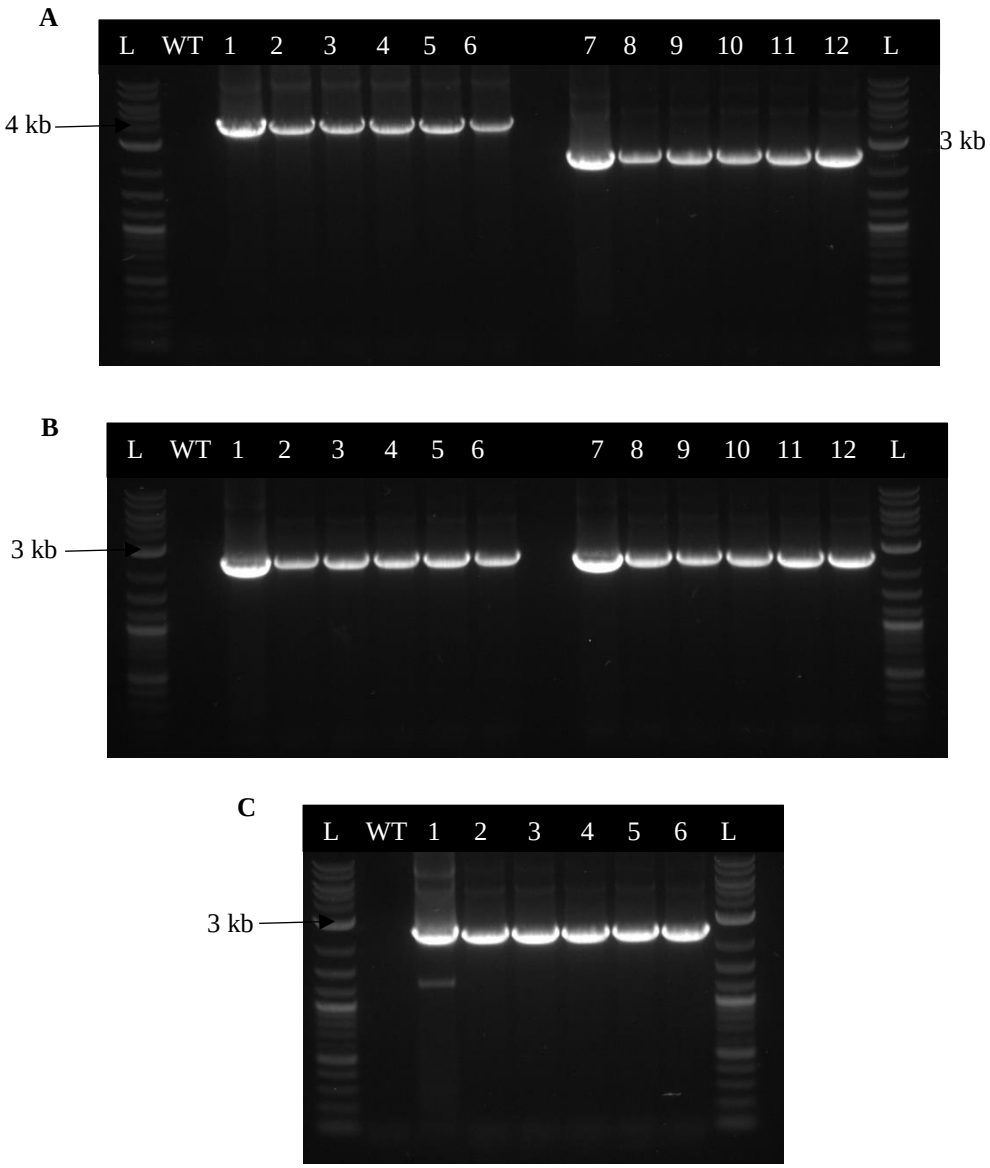


Figure 5.3: Colony-PCR screening of recombinant *P. thermoglucosidasius* KT1, KT2, KT3, KT4 and KT5 strains for MCR-b and MCR-MSR operons for 3-HP production. (A) A PCR of the WT was performed as a negative control. Lane 1= PCR of pMTLgSlimS-ZV1b as positive control (~ 4 kb); lanes 2-6 represents amplicons of ~ 4 kb corresponding to the operon from KT1 strain; lane 7= PCR of pMTLgSlimS-ZV2b as positive control (~ 3 kb); lanes 8-12 represents amplicons of ~ 3 kb corresponding to the operon from KT2 strain; L= 2-Log DNA Ladder. **(B)** A PCR of the WT was performed as a negative control. Lane 1= PCR of pMTLgSlimS-ZV3b as positive control (~ 3 kb); lanes 2-6 represents amplicons of ~ 3 kb corresponding to the

operon from KT3 strain; lane 7= PCR of pMTLgSlimS-ZV4b as positive control (~ 3 kb); lanes 8-12 represents amplicons of ~ 3 kb corresponding to the operon from KT4 strain; L= 2-Log DNA Ladder. **(C)** A PCR of the WT was performed as a negative control. Lane 1= PCR of pMTLgSlimS-ZV5b as positive control (~ 3 kb); lanes 2-6 represents amplicons of ~ 3 kb corresponding to the operon from KT5 strain; L= 2-Log DNA Ladder.

5.2.1.1 Shake-flask production of 3-HP by recombinant *P. thermoglucosidasius* KT1, KT2, KT3, KT4, and KT5 strains

To investigate the potential of the recombinant *P. thermoglucosidasius* KT1, KT2, KT3, KT4, and KT5 strains to produce 3-HP under aerobic conditions in shake-flask, the recombinant *P. thermoglucosidasius* KT1, KT2, KT3, KT4, and KT5 strains and the wild-type *P. thermoglucosidasius* strain were grown in semi-synthetic medium containing 10 $g \cdot l^{-1}$ glucose for 12 h at 52 $^{\circ}$C, 250 rpm. Samples were taken every hour and analysed by HPLC for metabolites produced by the strains as described in Methods section 2.15.1. The exponential growth phase in the wild-type and recombinant strains was observed before 6 h and 5 h, respectively (Figure 5.4 A). Glucose concentration in the supernatant analysed reveals its consumption during the exponential growth phase and reaching a minimum at 12 h in the wild-type and the recombinant strains except for the KT1 strain, getting to a minimum at 6 h (Figure 5.4 B).

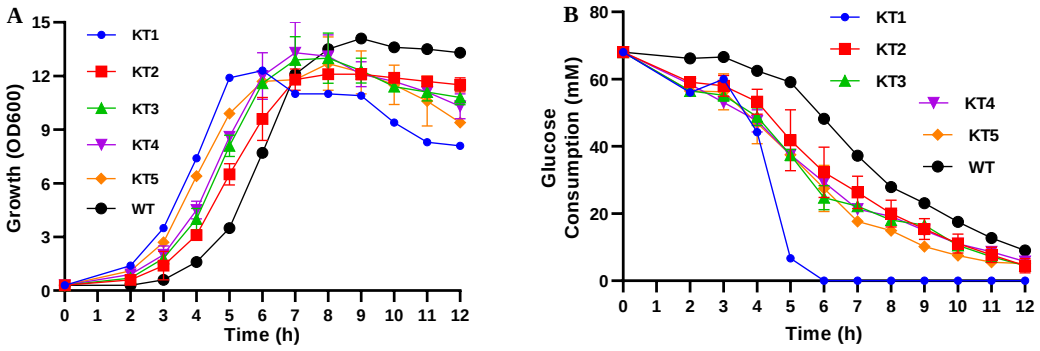


Figure 5.4: Growth curve (A) and glucose consumption (B) of recombinant *P. thermoglucosidasius* KT1, KT2, KT3, KT4, KT5 and wild-type strains. Growth (OD_{600}) was measured using a spectrophotometer at 600 nm. Residual glucose in the supernatant of the constructed strains in three replicates is characterised by HPLC for 12 h of shake-flask fermentation using 50 ml of semi-synthetic media supplemented with 10 $g \cdot l^{-1}$ glucose in a 250 ml baffled flask incubated at 52 $^{\circ}C$, 250 rpm. Results are shown as mean ± SD of three biological replicates (three cultures each derived from three different transformation events).

Additionally, 3-HP was detected by HPLC in the supernatant of the engineered strains within 4-12 h and reached a maximum between 7-11 h (Figure 5.5). The highest concentration of 3-HP produced by the recombinant *P. thermoglucosidasius* KT1, KT2, KT3, KT4, KT5 strains was found to be 0.4 ± 0.01 mM, 0.4 ± 0.1 mM, 0.4 ± 0.02 mM, 1.4 ± 0.3 mM and 1.3 ± 0.2 mM, respectively (Figure 5.5), with strains KT4 and KT5 producing the highest 3-HP.

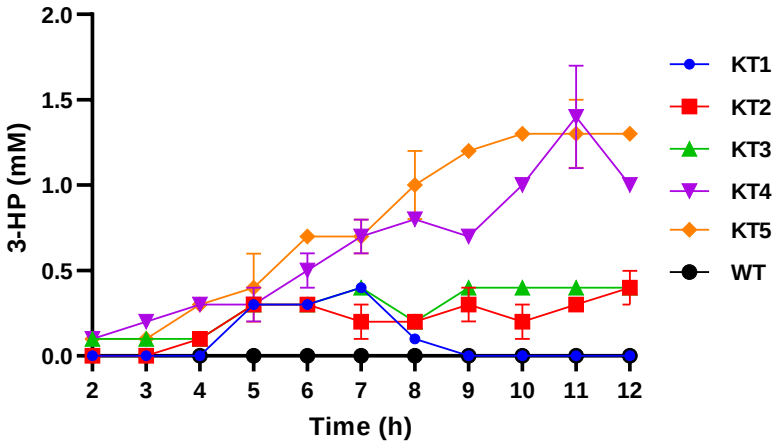


Figure 5.5: 3-HP production by recombinant *P. thermoglucosidasius* KT1, KT2, KT3, KT4, KT5 and wild-type strains. 3-HP produced by the generated recombinant strains in three replicates are characterised by HPLC for 12 h of shake-flask fermentation using 50 ml of semi-synthetic media supplemented with 10 $g \cdot l^{-1}$ glucose in a 250 ml baffled flask incubated at 52 $^{\circ}C$, 250 rpm. Results are shown as mean ± SD of three biological replicates (three cultures each derived from three different transformation events).

Furthermore, the production of acetate and lactate by the recombinant *P. thermoglucosidasius* KT1, KT2, KT3, KT4, KT5 and wild-type strains were also detected in the culture supernatant and measured by HPLC. Acetate and lactate were the significant by-products of fermentation produced by the recombinant and wild-type strains. Moreover, the acetate and lactate produced by recombinant strains KT1, KT2, and KT3 were assimilated after 7-8 h and 10 h, respectively (Figure 5.6 A-B).

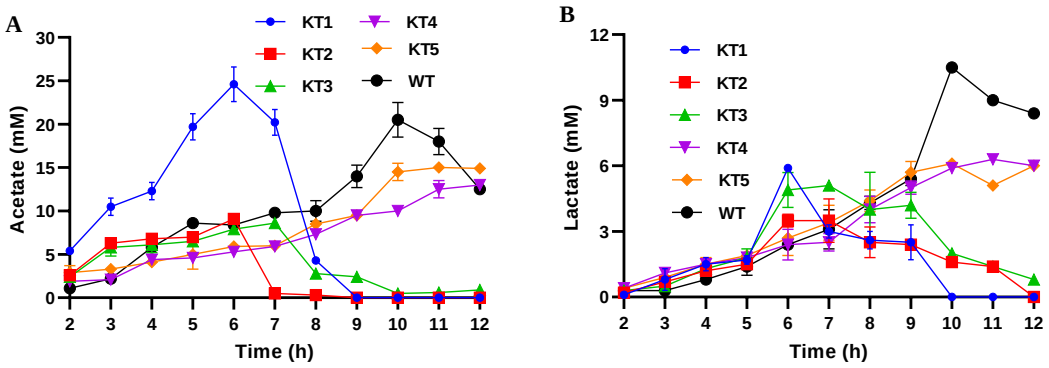


Figure 5.6: Production of acetate (A) and lactate (B) by recombinant *P. thermoglucosidasius* KT1, KT2, KT3, KT4, KT5 and wild-type strains. In three replicates, acetate and lactate profiles of the generated recombinant strains and wild-type are characterised by HPLC for 12 h of shake-flask fermentation using 50 ml of semi-synthetic media supplemented with 10 $g \cdot l^{-1}$ glucose in a 250 ml baffled flask incubated at 52 $^{\circ}C$, 250 rpm. Results are shown as mean ± SD of three biological replicates (three cultures each derived from three different transformation events).

5.2.2 Effect of lactate dehydrogenase (*ldhA*) gene deletion on 3-HP production

Having observed some lactate (1-6 mM) in the supernatant of the recombinant strains, the effect of deleting the lactate dehydrogenase gene on carbon flux towards 3-HP production was evaluated. *P. thermoglucosidasius* Δ*ldhA* (Sheng *et al.*, 2017) was used as the background strain. Furthermore, because strains KT4 and KT5 containing pMTLgSlimS-ZV4b and pMTLgSlimS-ZV5b plasmids produced the highest 3-HP, only pMTLgSlimS-ZV4b and pMTLgSlimS-ZV5b plasmids were used here.

5.2.2.1 Construction of recombinant *P. thermoglucosidasius* KT401 and KT501 strains for 3-HP production

Plasmids pMTLgSlimS-ZV4b and pMTLgSlimS-ZV5b (Bashir, 2018) were transformed into electro-competent cells of *P. thermoglucosidasius* Δ*ldhA* (Sheng *et al.,* 2017), thereby generating recombinant *P. thermoglucosidasius* KT401 and *P. thermoglucosidasius* KT501 strains, respectively, as described in Methods section 2.10.4. Additionally, to confirm the presence of the respective plasmids in the recombinant strains, colony PCR was performed using plasmid-based primers ColE1_F2 and RepB_R1 (exact primer sequences in Table 2.1) as described in Methods section 2.9.4. The observed amplified DNA product of ~ 3 kb corresponded to the expected size of mcr_{st}-msr_{ss} and mcr_{st}-msr_{ms} operons on the autonomous pMTLgSlimS-ZV4b and pMTLgSlimS-ZV5b plasmids in the recombinant *P. thermoglucosidasius* KT401 and *P. thermoglucosidasius* KT501 strains (Figure 5.7). Moreover, the PCR product was purified, followed by Sanger sequencing to verify the correct nucleotide sequence of the mcr_{st}-msr_{ss} and mcr_{st}-msr_{ms} operons.

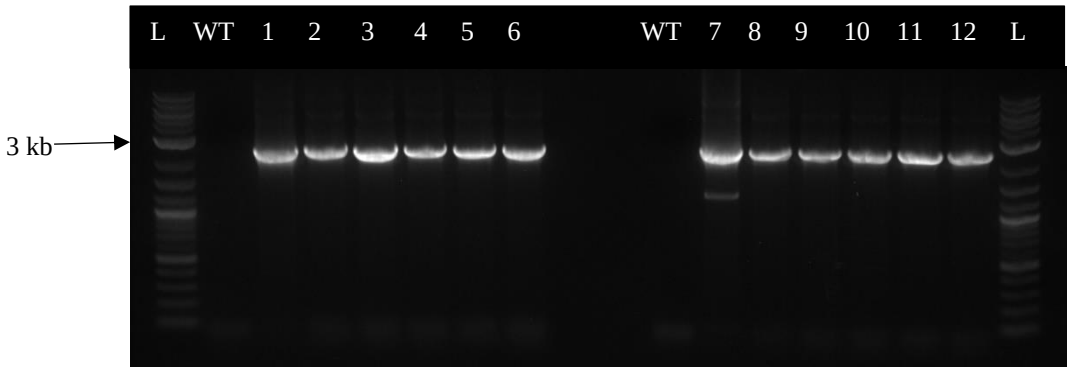


Figure 5.7: Colony-PCR screening of recombinant *P. thermoglucosidasius* KT401 and KT501 strains for mcr_{st}-msr_{ss} and mcr_{st}-msr_{ms} operons for 3-HP production. WT= A PCR of *P. thermoglucosidasius* Δ*ldhA* strain was performed as a negative control. Lane 1= PCR of pMTLgSlimS-ZV4b as positive control (~ 3 kb); lanes 2-6 represents amplicons of ~ 3 kb corresponding to the expected size of mcr_{st}-msr_{ss} in pMTLgSlimS-ZV4b plasmid from KT401 strain; lane 7= PCR of pMTLgSlimS-ZV5b as positive control (~ 3 kb); lanes 8-12 represents amplicons of ~ 3 kb corresponding to the expected size of mcr_{st}-msr_{ms} in pMTLgSlimS-ZV5b plasmid from KT501 strain; L= 2-Log DNA Ladder.

5.2.2.2 Shake-flask production of 3-HP by recombinant *P. thermoglucosidasius* KT401 and KT501 strains

To evaluate the effect of *ldhA* deletion on 3-HP production, the potential of the recombinant *P. thermoglucosidasius* KT401, KT501 and Δ*ldhA* strains to produce 3-HP under aerobic conditions in shake-flask was assessed. The recombinant strains and the Δ*ldhA* strain were grown in a semi-synthetic medium containing 10 $g \cdot l^{-1}$ glucose for 24 h at 52 °C, 250 rpm. Samples were taken every 2 h and analysed by HPLC for metabolites produced by the strains as described in Methods section 2.15.1. Recombinant KT401 and KT501 and Δ*ldhA* strains exhibited exponential growth phase

between 2-6 and 8 h, respectively (Figure 5.8 A). Furthermore, during this growth phase, the glucose concentration in the medium gradually declined, reaching a minimum at 12 h and consumed entirely by 24 h in the recombinant and Δ*ldhA* strains (Figure 5.8 B). Additionally, the production of 3-HP by the recombinant strains was detected by HPLC in the supernatant from 6 h and reached a maximum of 1.1-1.2 mM between 8-12 h (Figure 5.8 C). The highest concentration of 3-HP produced by the recombinant *P. thermoglucosidasius* KT401 and KT501 strains was found to be 1.3 ± 0.1 mM and 1.2 ± 0.2 mM, respectively. The comparative 3-HP titres observed here and in the wild-type background showed no significant differences. Moreover, the production of acetate by the recombinant *P. thermoglucosidasius* KT401, KT501 and Δ*ldhA* strains was also measured by HPLC. Acetate accumulation in the *P. thermoglucosidasius* Δ*ldhA* strain was 22.6 ± 0.5 mM, which was higher than the recombinant strains KT401 and KT501 between 4-12 h, and in all cases reabsorbed by 24 h (Figure 5.8 D). Deletion of *ldhA* abolished production of lactate under the aerobic cultivation used.

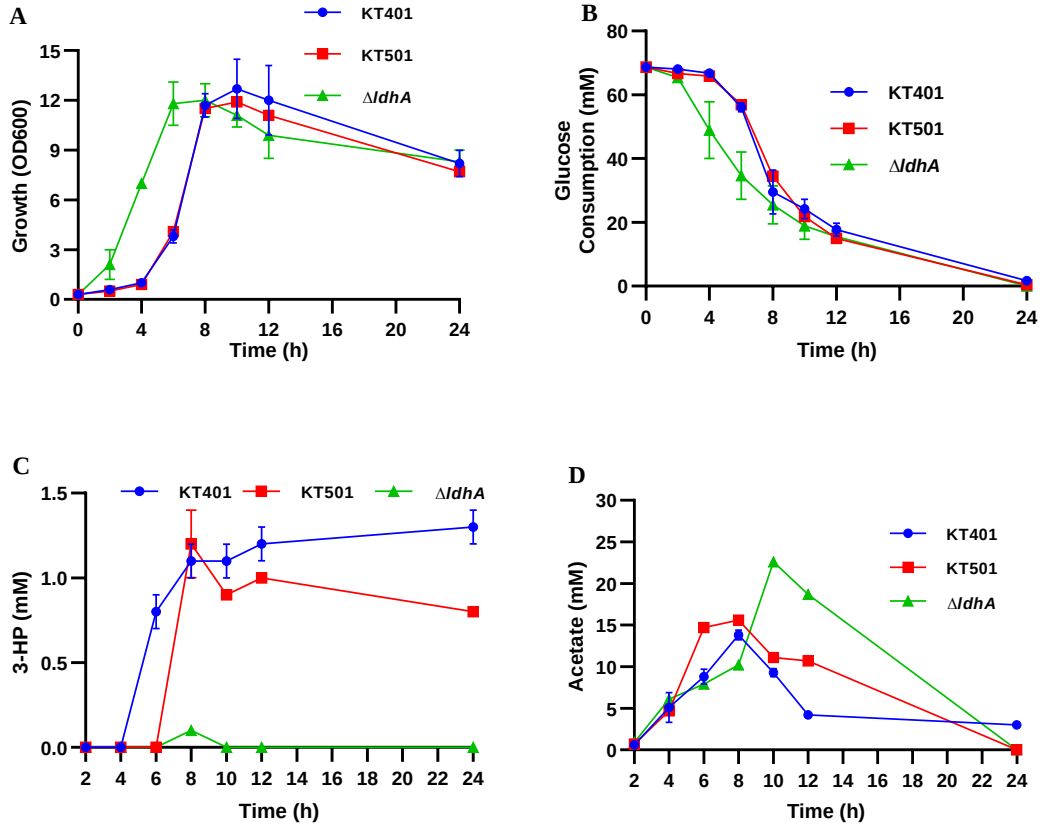


Figure 5.8: Production of metabolites by recombinant *P. thermoglucosidasius* KT401, KT501 and Δ*ldhA* strains. Fermentation products (3-HP, lactate, acetate and glucose) and growth (OD_{600}) of the constructed recombinant strains and Δ*ldhA* strain in three replicates are characterised by HPLC for 24 h of shake-flask fermentation using 50 ml of semi-synthetic media 10 $g \cdot l^{-1}$ glucose in a 250 ml baffled flask incubated at 52 $^{\circ}C$, 250 rpm. Results are shown as mean ± SD of three biological replicates (three cultures each derived from three different transformation events).

5.2.3 Investigating the native catabolism of 3-HP by *P. thermoglucosidasius* NCIMB 11955

5.2.3.1 Utilisation of 3-HP as a source of carbon and energy in *P. thermoglucosidasius*

To investigate whether *P. thermoglucosidasius* NCIMB 11955 utilises 3-HP as the sole carbon and energy source, wild-type *P. thermoglucosidasius* was grown in *Clostridium* basal medium (CBM), and CBM supplemented with 50 mM 3-HP, respectively, for 24 h at 60 ˚C, 250 rpm in a 250 ml baffled flasks. OD_{600} and samples were taken every hr until 12 h and then at 24 h. Samples were analysed by HPLC for residual 3-HP concentrations as described in Methods section 2.15.1. The maximum OD_{600} of 2.33 ± 0.03 was observed at 6 h with the exponential growth phase between 2-4 h before entering the stationary phase in CBM alone (Figure 5.9 A). On the other hand, in CBM containing 50 mM 3-HP, a maximum OD_{600} of 5.71 was reached, with the exponential growth phase observed between 2-4 h (Figure 5.9 B). The higher final optical densities observed in the presence of 3-HP compared to cultures grown in CBM alone suggested that 3-HP could be used as a carbon and energy source by *P. thermoglucosidasius*. Furthermore, under these conditions (CBM + 50 mM 3-HP), the *P. thermoglucosidasius* started to consume 3-HP after 3 h and entirely consumed the available 3-HP in 12 h (Figure 5.9 C) as measured by HPLC. This confirms that 3-HP could support the growth of the organism in CBM. It is worthy of note that the 3-HP consumption by the wild-type strain continues even after reaching the stationary phase of growth between 7-12 h. Additionally, it was observed that the wild-type *P. thermoglucosidasius* produced acetate and lactate as detected by HPLC in CBM containing 50 mM 3-HP (data not

shown). These suggest that *P. thermoglucosidasius* NCIMB 11955 consumes 3-HP with apparent production of metabolites.

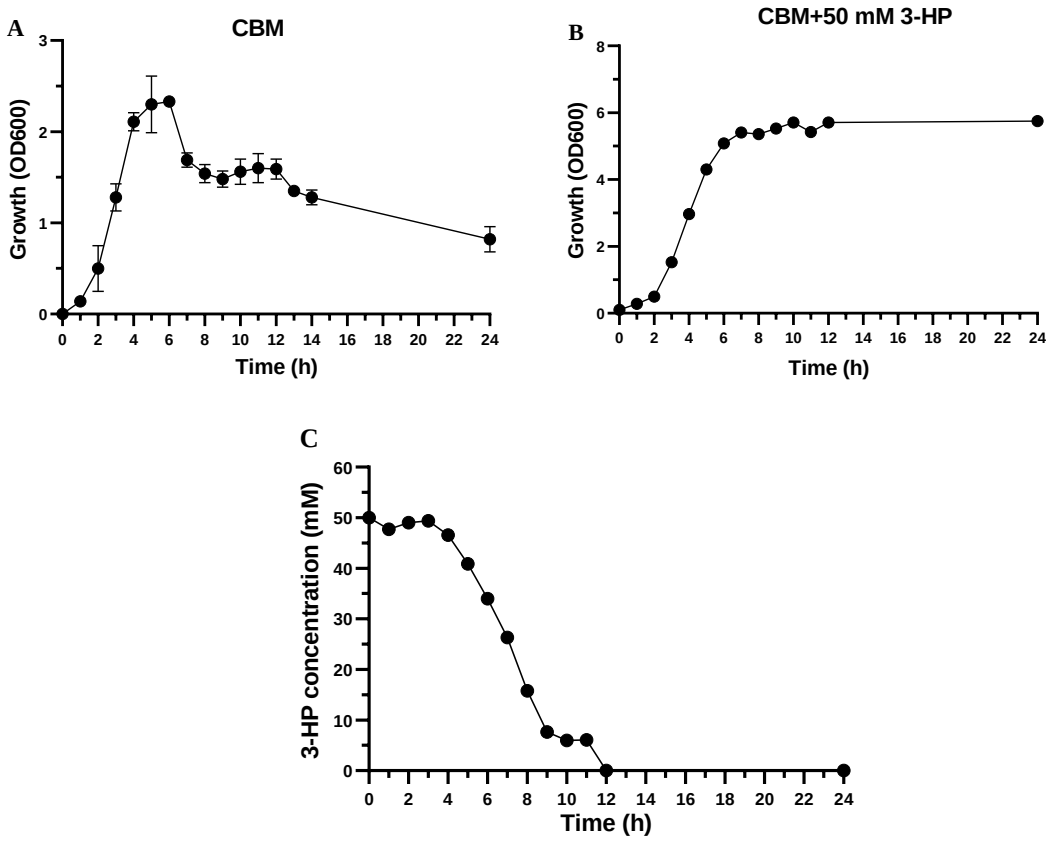


Figure 5.9: Growth and 3-HP consumption of wild-type *P. thermoglucosidasius* NCIMB 11955. (A) Growth of *P. thermoglucosidasius* on CBM at 60 °C: at hourly intervals until the 12 h time point and then, 24 h endpoint growth of *P. thermoglucosidasius* following cultivation in CBM alone. Growth was followed by measuring OD_{600}. **(B)** Growth of *P. thermoglucosidasius* on CBM with 50 mM 3-HP as the source of carbon and energy at 60 °C: at hourly intervals until the 12 h time point and then, 24 h endpoint growth of *P. thermoglucosidasius* following cultivation in CBM containing 50 mM 3-HP. Growth was followed by measuring OD_{600}. **(C)** 3-HP consumption: residual 3-HP concentrations were quantified by HPLC. Results are shown as mean ± standard deviation of three biological replicates.

5.2.3.2 Identification of *P. thermoglucosidasius* genes involved in 3-HP metabolism

Given the consumption of 3-HP by the wild-type *P. thermoglucosidasius*, its genome was interrogated for the presence of 3-HP utilisation genes. Two putative (methyl)malonate semialdehyde dehydrogenase genes, *mmsA1* (BCV53_10115) and *mmsA2* (BCV53_11155), required for the oxidative decarboxylation of malonate semialdehyde to acetyl-CoA, were found to be present in the genome (Figure 5.10). Additionally, a 3-hydroxyisobutyrate dehydrogenase gene, *hbdH* (BCV53_14255), was also found to be present (Figure 5.10). Previous studies in *P. denitrificans* demonstrate that 3-HP degradation genes are clustered in two operons, a polycistronic operon consisting of (methyl)malonate semialdehyde dehydrogenase (*mmsA*) and a 3-hydroxyisobutyrate dehydrogenase gene (*hbdH*), and a monocistronic operon comprised of a single 3-hydroxyisopropionate dehydrogenase (*hpdH*) (Zhou *et al.*, 2014: Zhou *et al.*, 2015). *C. necator* H16 has also been shown to possess three different (methyl)malonate semialdehyde dehydrogenase (*mmsA1, mmsA2, mmsA3*) genes that were shown to be responsible for the utilisation of 3-HP (Arenas-López *et al.*, 2019). Furthermore, the *hbdH* of *P. denitrificans* has been expressed heterologously in *E. coli* and shown to oxidise several 3-hydroxyacids including 3-HP (Zhou *et al.*, 2013b). This implies that the HbdH may catalyse the oxidation of 3-HP to malonate semialdehyde in *P. denitrificans*. The presence of these genes suggests that a complete oxidative 3-HP utilisation pathway does exist in *P. thermoglucosidasius* NCIMB 11955.

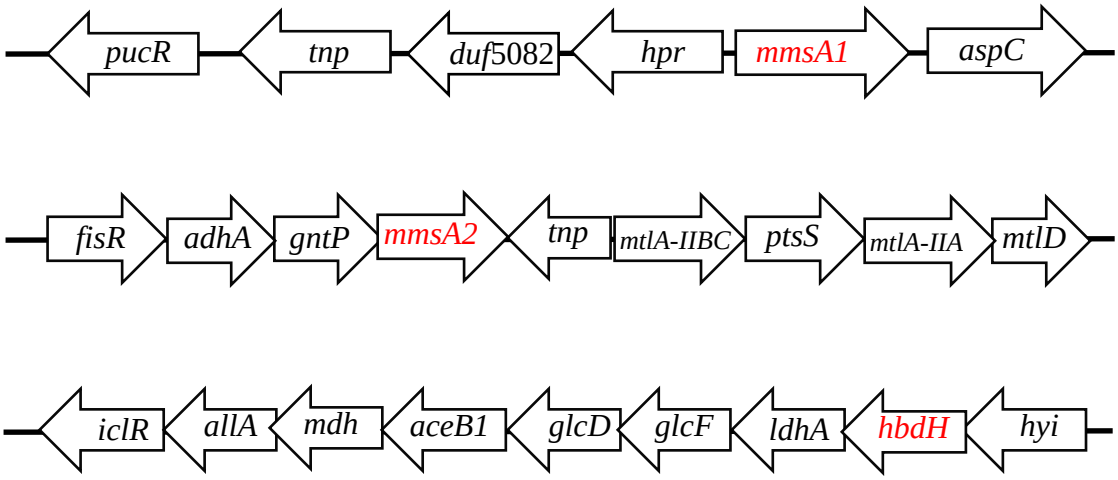


Figure 5.10: *P. thermoglucosidasius* NCIMB 11955 genes predicted to be involved in 3-HP degradation. Putative genes involved in 3-HP degradation are shown in red, together with their adjacent upstream and downstream genes in black. Genes encode the following enzymes: (methyl)malonate semialdehyde dehydrogenase (*mmsA1*, BCV53_10115; *mmsA2*, BCV53_11155), 3-hydroxyisobutyrate dehydrogenase (*hbdH*, BCV53_14255), PucR family transcriptional regulator (*pucR*, BCV53_10095), transposase (*tnp*, BCV53_10100), DUF5082 domain-containing protein (*duf5082*, BCV53_10105), hypothetical protein (*hpr*, BCV53_10110), aspartate aminotransferase (*aspC*, BCV53_10120), sigma-54-dependent Fis family transcriptional regulator (*fisR*, BCV53_11140), alcohol dehydrogenase (*adhA*, BCV53_11145), GntP family permease (*gntP*, BCV53_11150), transposase (*tnp*, BCV53_11160), PTS mannitol transporter subunit IIBC (*mtlA-IIBC*, BCV53_11165), PTS sugar transporter (*ptsS*, BCV53_11170), PTS mannitol transporter subunit IIA (*mtlA-IIA*, BCV53_11175), mannitol-1-phosphate 5-dehydrogenase (*mtlD*, BCV53_11180), IclR family transcriptional regulator (*iclR*, BCV53_14220), ureidoglycolate lyase (*allA*, BCV53_14225), malate dehydrogenase (*mdh*, BCV53_14230), malate synthase G (*aceB1*, BCV53_14235), glycolate oxidase subunit GlcD (*glcD*, BCV53_14240),

glycolate oxidase (*glcF*, BCV53_14245), lactate dehydrogenase (*ldhA*, BCV53_14250), hydroxypyruvate isomerase (*hyi*, BCV53_14260).

5.2.3.3 Generation of a *P. thermoglucosidasius* strain unable to utilise 3-HP as a source of carbon and energy

The two *mmsA* genes were first targeted for gene knockout to investigate their relative roles in 3-HP consumption based on the observation in *C. necator* H16, where inactivation of all three *mmsA* genes was required to abolish 3-HP degradation (Arenas-López *et al.*, 2019). The respective genes (*mmsA1 and mmsA2*) were inactivated by deleting their entire coding regions singly and in combination, using a newly developed CRISPR/Cas genome editing tool (Lau *et al.*, 2021) by Dr Lili Sheng, University of Nottingham (unpublished), but have not been characterised (Figure 5.11). The confirmed mutants were designated Δ*mmsA2* and Δ*mmsA2mmsA1*. Three separate attempts to obtain a single *mmsA1* mutant were unsuccessful by both Dr Sheng and in this study.

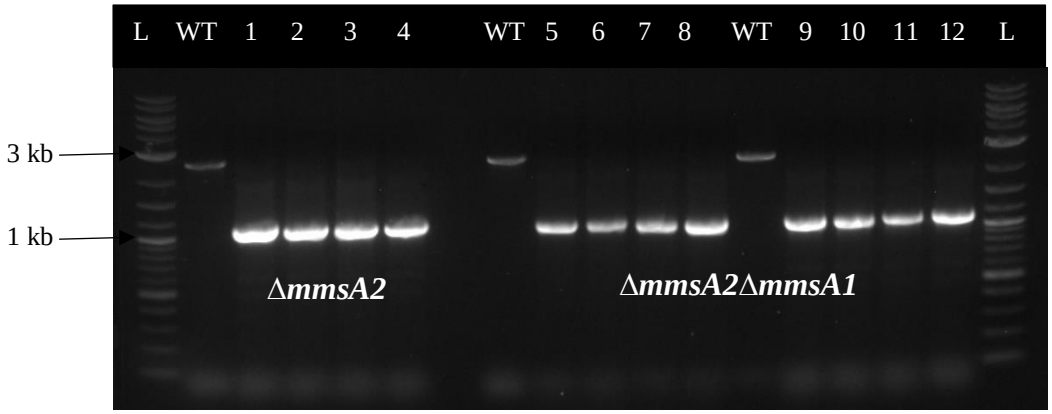


Figure 5.11: Gel electrophoresis image of colony PCR to re-confirm the deletion of *mmsA2* and *mmsA1* genes in *P. thermoglucosidasius* Δ*mmsA2* and Δ*mmsA2mmsA1* strains. The primers used were FC_mmsA2 and RC_mmsA2 (primer sequences Table 2.1) to assess *mmsA2* deletion, while FC_mmsA1 and RC_mmsA1 primers (primer sequences Table 2.1) were used for *mmsA1* deletion. WT= A PCR of *P. thermoglucosidasius* wild-type strain was performed as a control (~ 2.9 kb). Successful deletion of the *mmsA2* gene would yield a band at ~ 1.1 kb (lanes 1-4 and 5-8 in Δ*mmsA2* and ΔmmsA2mmsA1 strains, respectively). At the same time, deletion of the *mmsA1* gene would yield a band at ~ 1.1 kb (lanes 9-12 in Δ*mmsA2mmsA1* strains). L= A 2-Log DNA Ladder was used to determine the size of each band.

The Δ*mmsA2* and Δ*mmsA2mmsA1* mutants and the wild-type *P. thermoglucosidasius* strain were grown in CBM alone to determine their baseline growth rates and in CBM supplemented with 50 mM 3-HP as the source of carbon and energy in order to see whether any of the deletions could abolish or reduce growth and 3-HP utilisation. The strains were grown for 96 h at 60 °C, 250 rpm in 250 ml baffled flasks containing 50 ml of media. OD_{600} and samples were taken every hour until 14 h and then at 24 h for CBM alone, and 6 h and then 12 h intervals for CBM+50 mM 3-HP until 96 h. Samples were analysed by HPLC for residual 3-HP concentration as described in Methods section 2.15.1. Growth was followed by measuring OD_{600} using a spectrophotometer at 600 nm.

Figure 5.12 A-C shows the growth and 3-HP consumption profiles observed in CBM alone and CBM+50 mM 3-HP, respectively. In CBM alone, the Δ*mmsA2* and Δ*mmsA2mmsA1* mutants reached maximum OD_{600} of 2.79 ± 0.04 and 2.73 ± 0.15, respectively, which is comparable to the wild-type at 6 h, with the exponential growth phase between 2-4 h before entering the stationary phase (Figure 5.12 A). In CBM+50 mM 3-HP, the wild-type *P. thermoglucosidasius* reached a maximum OD_{600} of 6.27 ± 0.40 at 12 h compared to the 2.56 ± 0.20 and 2.53 ± 0.20 at 6 h for Δ*mmsA2* and Δ*mmsA2mmsA1* mutants, respectively (Figure 5.12 B). However, the OD_{600} of the Δ*mmsA2* mutant goes up to around 2.26 at 96 h after being less than 2 uptil 84 h. Furthermore, whilst the wild-type consumes the available 3-HP in 12-24 h, Δ*mmsA2* and Δ*mmsA2mmsA1* mutants could not utilise 3-HP up to 84 and 96 h, respectively (Figure 5.12 C). On the one hand, the increase in OD_{600} with the corresponding 3-HP consumption after 84 h observed for Δ*mmsA2* mutant indicates that the deletion of the *mmsA2* could only delay growth and utilisation for about 84 h but not abolish growth and 3-HP utilisation. At this point, it was observed that the experiment could not be allowed to continue because much of the medium had evaporated (due to the

combination of factors such as the high growth temperature (60 ˚C), media loss through the lids and sampling) affecting OD_{600} and residual 3-HP measurements. On the other hand, the double Δ*mmsA2mmsA1* knock-out strain could not grow or utilise 3-HP as the sole carbon and energy source within the tested timeframe of up to 96 h (Figure 5.12 C).

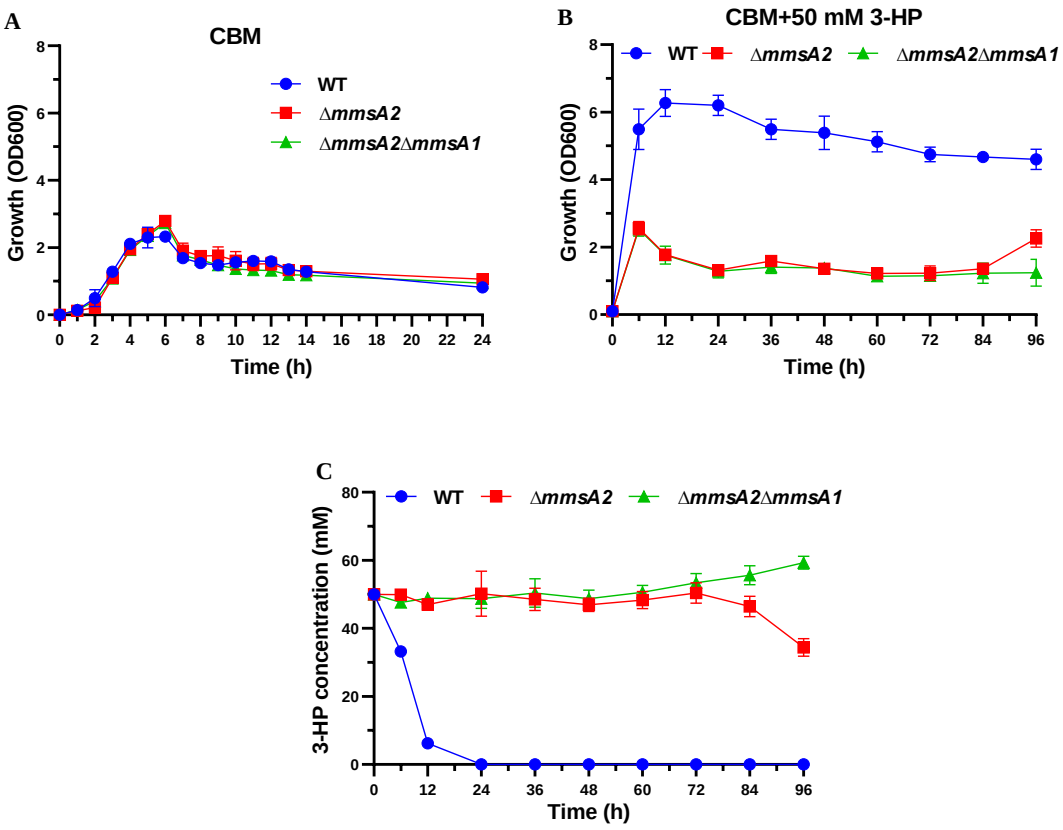


Figure 5.12: Growth and 3-HP consumption of wild-type *P. thermoglucosidasius* NCIMB 11955 and generated mutant strains. (A) Growth of wild-type *P. thermoglucosidasius,* Δ*mmsA2* and Δ*mmsA2mmsA1* mutants on CBM at 60 ˚C: measured hourly until 14 h and then 24 h endpoint growth of wild-type and mutant strains of *P. thermoglucosidasius* following cultivation in CBM alone. Growth was followed by measuring OD_{600} using a spectrophotometer. **(B)** Growth of wild-type *P. thermoglucosidasius,* Δ*mmsA2* and Δ*mmsA2mmsA1* mutants on CBM with 50 mM 3-HP as the sole source of carbon and energy at 60 ˚C: at 6 h and 12 h, and then 24 h time point intervals growth of wild-type and mutant strains of *P. thermoglucosidasius* following cultivation in CBM containing 50 mM 3-HP. Growth was followed by measuring OD_{600} using a spectrophotometer. Results are shown as mean ± SD of three biological replicates. **(C)** 3-HP consumption: residual 3-HP concentrations were quantified by HPLC. Results are shown as mean ± SD of three biological replicates.

5.3 Discussion

P. thermoglucosidasius is considered a promising platform strain for producing valuable chemicals such as isobutanol, 2,3-butanediol, and terpenes (Lin *et al.,* 2014; Xio *et al.,* 2012; Zhou *et al.,* 2020; Styles *et al.,* 2021). Furthermore, its high growth rate leads to increased cell densities in a short period and, thus, increased production of desired metabolites.

Several microorganisms have been engineered to produce 3-HP, including *E. coli*, *K. pneumonia*, *B. subtilis*, *M. extorquens* and several other organisms (Rathnasingh *et al.*, 2012; Lim *et al.*, 2016; Jiang *et al.,* 2018; Kalantari *et al.*, 2017; Yang *et al.,* 2017). Furthermore, *Pyrococcus furiosus* has been engineered to produce 3-HP from hydrogen gas and carbon dioxide (Keller *et al.,* 2013). *P. thermoglucosidasius* has previously been shown to produce 3-HP (3.0-3.8 mM) via the malonyl-CoA pathway using hereologous ACC integrated at the *pyrE* locus in the chromosome and MCR-MSR operons expressed on plasmids (Bashir, 2018). In the present study, it was demonstrated that 3-HP could be produced by *P. thermoglucosidasius* NCIMB 11955 for the first time using the native ACC present in its genome and various combinations of heterologous MCR-MSR operons expressed on autonomous plasmids, and then, secondly, investigated the native catabolism of 3-HP to engineer a strain incapable of 3-HP utilisation to prevent its consumption during production. The pMTLgSlimS-ZV1b, pMTLgSlimS-ZV2b, pMTLgSlimS-ZV3b, pMTLgSlimS-ZV4b and pMTLgSlimS-ZV5b plasmids used in this study were previously constructed by Bashir (2018).

3-HP production in the recombinant *P. thermoglucosidasius* KT1, KT2, KT3, KT4, and KT5 strains was in the range of 0.4–1.4 mM, with strains KT4 and KT5 carrying pMTLgSlimS-ZV4b and pMTLgSlimS-ZV5b plasmids producing the maximum 3-HP

of 1.4 mM and 1.3 mM, respectively. These results indicate that the native ACC can convert acetyl-CoA to malonyl-CoA for its subsequent conversion to 3-HP. Our results are supported by the previous study in *P. thermoglucosidasius* by Bashir, (2018), where all the recombinant strains BK1b, BK2b, BK3b, BK4b, and BK5b produced similar 3-HP titres with BK4b and BK5b strains carrying pMTLgSlimS-ZV4b and pMTLgSlimS-ZV5b vectors having the highest concentration of 3.68 ± 1.25 mM and 3.00 ± 0.6 mM 3-HP, respectively. The 3-HP titre in the recombinant strains in our study is low (0.4–1.4 mM) compared to 3.0-3.8 mM observed by Bashir (2018), possibly due to the limited availability of malonyl-CoA for the conversion to 3-HP. However, it does show the prospect of the native ACC being used for 3-HP production via the malonyl-CoA pathway, which can be improved upon through upregulation of the ACC gene by placing it under the control of a strong promoter.

In addition to the production of 3-HP by the recombinant *P. thermoglucosidasius* strains, by-products, namely lactate and acetate, were also detected and quantified in the culture supernatant. Moreover, the production of 3-HP can be improved by blocking competing pathways such as those in lactate and acetate production, which divert the carbon flux towards its synthesis. To address this issue, improvement in the production of 3-HP by blocking lactate formation was tested. The deletion of the *ldhA* gene abolishes lactate production under the aerobic condition employed. However, the production of 3-HP was not improved, possibly due to limiting malonyl-CoA. Similar findings were observed in engineered *E. coli,* where efforts to improve 3-HP production by deleting *ldhA* were unsuccessful (Rathnasingh *et al.,* 2012). Furthermore, the production of large amounts of acetate depletes intracellular acetyl-CoA, which is needed for the conversion of malonyl-CoA, a precursor for 3-HP production (Rathnasingh *et al.,* 2012). Thus, blocking the acetate production pathway would be necessary for further studies.

Likewise, further studies to improve the production of 3-HP via the malonyl-CoA pathway could investigate increasing the concentrations of malonyl-CoA by overexpressing the native ACC, and NADPH by expressing nicotinamide nucleotide transhydrogenase (PntAB) (Rathnasingh *et al.,* 2012).

Several microorganisms are reported to utilise and grow on 3-HP as the source of carbon and energy, including *R. sphaeroides*, *P. denitrificans*, *M. extorquens*, and *C. necator* (Schneider *et al.,* 2012; Zhou *et al.,* 2014; Yang *et al.,* 2017; Arenas-López *et al.,* 2019). Here, it was demonstrated that *P. thermoglucosidasius* NCIMB 11955 can metabolise 3-HP as the carbon and energy source. *P. thermoglucosidasius* is a promising host for the production of 3-HP from lignocellulosic biomass. However, to achieve biological production of the platform chemical, the observed consumption of the metabolite is an undesirable characteristic that requires elimination (Arenas-López *et al.,* 2019). Therefore, genes contributing 3-HP degradation were identified and investigated. Two main pathways for 3-HP degradation have been reported, an oxidative and a reductive route. The oxidative route reported in *P. denitrificans* and *C. necator* H16 is shorter and involves the oxidation of 3-HP to malonate semialdehyde, which is then decarboxylated to acetyl-CoA (Zhou *et al.,* 2014, Arenas-López *et al.,* 2019). In contrast, the reductive route as present in *R. sphaeroides* involves the conversion of 3-HP to 3-HP-CoA, then reduction to propionyl-CoA and finally the conversion of the propionyl-CoA to succinyl-CoA (Schneider *et al.,* 2012).

The presence of two putative (methyl)malonate semialdehyde dehydrogenases (*mmsA1* and *mmsA2*) and 3-hydroxyisobutyrate dehydrogenase (*hbdH*) genes in *P. thermoglucosidasius* NCIMB 11955 indicates that 3-HP might be degraded exclusively through the oxidative pathway. Whilst deletion of the *mmsA1* to get a single mutant was

unsuccessful, inactivation of the *mmsA2* only delayed growth and 3-HP consumption up to 84 h. However, the combined inactivation of the two (methyl)malonate semialdehyde dehydrogenase genes abolished 3-HP utilisation, at least for this study's timeframe (96 h). Therefore, achieving the aim of this study, investigating and engineering a *P. thermoglucosidasius* NCIMB 11955 strain to be unable to utilise 3-HP as a source of carbon and energy.

Whilst inactivation of all two *mmsA* genes was needed to abolish 3-HP degradation at least under the conditions employed in this study, the deletion of the *hbdH* gene for mutant strain construction and subsequent testing to determine its effect in 3-HP metabolism was not possible due to time constraints. However, the inactivation of the *hbdH* is still an option for further engineering and is a priority for future work. Our results are supported by previous findings involving the engineering of strains that do not degrade 3-HP, such as in *P. denitrificans*, *M. extorquens*, and *C. necator* H16 (Zhou *et al.*, 2014; Yang *et al.*, 2017; Arenas-López *et al.*, 2019). In *P. denitrificans*, it was reported that *hpdH* and *hbdH* genes are primarily responsible for 3-HP consumption, and their subsequent deletion prevents 3-HP degradation (Zhou *et al.*, 2014). It was also reported that *M. extorquens* degrade 3-HP via the reductive route where 3-HP is reduced to 3-HP-CoA and then sequentially converted to acryl-CoA and propionyl-CoA, and deleting acrylyl-CoA reductase (AcuI) resulted in the slow degradation of 3-HP in the late stationary phase (Yang *et al.*, 2017). Another study in *C. necator* H16 reported that putative genes for both reductive and oxidative routes are present in the organism, but the degradation of 3-HP occurred via the oxidative pathway. And the subsequent inactivation of the three separate (methyl)malonate semialdehyde dehydrogenases (*mmsA*) generates a strain incapable of 3-HP degradation (Arenas-López *et al.*, 2019).

5.4 Conclusions

P. thermoglucosidasius NCIMB 11955 was engineered to produce 3-HP via the malonyl-CoA pathway using the native ACC and heterologously expressing different combinations of MCR–MSR operons on autonomous plasmids. The engineered *P. thermoglucosidasius* NCIMB 11955 demonstrated the production of 3-HP on glucose with the highest titres of 1.4 mM and 1.3 mM, respectively. Efforts to improve 3-HP production by blocking lactate formation by deleting *ldhA* were unsuccessful. However, further strain optimisation is required to make the system industrially feasible. Additionally, this study engineered *P. thermoglucosidasius* NCIMB 11955 strain unable to degrade 3-HP for the timeframe tested. Inactivation of the genes encoding both (methyl)malonate semialdehyde dehydrogenases identified was required to abolish 3-HP degradation in *P. thermoglucosidasius*. The *P. thermoglucosidasius* Δ*mmsA2*Δ*mmsA1* strain represents an ideal chassis for future production of 3-HP.

CHAPTER 6

6.0 TOWARDS METABOLIC ENGINEERING OF *P. THERMOGLUCOSIDASIUS* FOR ISOBUTANOL PRODUCTION

6.1 Introduction

Despite being the world's major large-scale biofuel, ethanol is disadvantaged by its low energy density and high hygroscopicity (Lee *et al.*, 2012). These limitations underscore the need for advanced alternative biofuels such as isobutanol made through the introduction of the requisite pathways in a microbial chassis. The introduction and expression of heterologous pathways in a microbial host, however, may lead to compatibility issues such as metabolic imbalance (Atsumi *et al*., 2008). Additionally, the build-up of the non-native metabolites may cause cytotoxicity (Barbirato *et al.*, 1996; Pitera *et al.*, 2007). To overcome these issues and achieve higher productivity of the desired foreign products, native pathways compatible with the host are desirable (Atsumi *et al*., 2008). Consequently, the amino acid biosynthesis pathways are exploited for isobutanol and other higher alcohol production (Sentheshanmuganathan and Elsden, 1958).

Butanol is produced by fermentation of first-generation and second-generation feedstocks via the Acetone-Butanol-Ethanol (ABE) pathway of *Clostridium* bacteria (Gheshlaghi *et al.,* 2009). Alternatively, butanol can be produced by chemical conversion of ethanol using magnesium (Mg) or aluminium (AI) mixed oxides or hydroxyapatite catalysts (Ndaba *et al.*, 2015). However, the ABE route is limited by the low butanol yield due to hetero-fermentation (0.28-0.33 gg^{-1}), low final butanol concentration (<20 $g \cdot l^{-1}$) caused by inhibition during fermentation, and high cost of butanol recovery from low-concentration yields (Ndaba *et al.*, 2015). As an alternative, butanol can be produced through engineered decarboxylation and reduction of short-chain α-keto acids by α-keto acid decarboxylase and alcohol dehydrogenase, respectively.

6.1.1 Isobutanol production via a non-fermentative pathway

Isobutanol is an isomer of butanol and produced via the Ehrlich degradation pathway, where amino acids are converted to their corresponding 2-keto acids by branched-chain amino-acid aminotransferase (Hazelwood *et al.*, 2008; Dickinson, 2000). The 2-keto acids generated from their corresponding amino acids by branched-chain amino-acid aminotransferase can be further converted to aldehydes by 2-ketoacid decarboxylases (KDCs) through the decarboxylation step of the Ehrlich degradation pathway, and finally reduced to fusel alcohols by alcohol dehydrogenases (ADHs) (Atsumi *et al.*, 2008). For isobutanol production, 2-ketoisovalerate, the precursor for valine biosynthesis, is converted to isobutyraldehyde catalysed by 2-ketoisovalerate decarboxylase and reduced to isobutanol by alcohol dehydrogenase (Figure 6.1). Utilising this approach, only two non-native steps are required to produce isobutanol by bypassing the intermediates from amino acid biosynthesis to alcohol production (Atsumi *et al.*, 2008).

Whereas ADHs are common in many organisms, KDCs, which play an essential role in the production pathway, are less common in bacteria (Konig, 1998). According to a study by Atsumi *et al.* (2008), ketoisovalerate decarboxylase (KIVD) is the most active and versatile decarboxylase with the ability to utilise several 2-keto acids among the five KDCs tested in *E. coli* (Pdc6, Aro10, Thi3 from *S. cerevisiae*, KIVD from *Lactococcus lactis*, and PDC from *C. acetobutylicum*). Furthermore, improving the availability of the amino acid precursor is essential for high yield. Thus, the addition of various 2-keto acids to the *E. coli* culture expressing KIVD led to a 2 to 23-fold increase in the specific production of the corresponding alcohols with a marked decreased production of the other alcohols. These observations indicate the importance of increasing 2-keto acids concentration and flux towards the productivity and specificity

of production of the alcohols. Moreover, the deletion of genes that contribute to by-product formation can also improve 2-keto acids flux and subsequently increase fusel alcohol production (Atsumi *et al.,* 2008). "*The advantages of the 2-keto acid pathways are their compatibility with most organisms and many pathway enzymes available from various sources. These features underlie the relative ease in engineering isobutanol production in various organisms compared with 1-butanol*" (Chen and Liao, 2016).

Several researchers have reported isobutanol production via the ketoacid pathway in different hosts. These include *E. coli*, *C. cellulolyticum*, *P. thermoglucosidasius*, *C. thermocellum*, *S. cerevisiae*, *Synechococcus elongatus* PCC7942, *C. necator*, *B. subtilis*, and *C. glutamicum* using either glucose, cellulose, cellobiose, CO_2 or amino acids as a substrate (Atsumi *et al.,* 2008; Higashide *et al*., 2011; Lin *et al.,* 2014; Lin *et al.*, 2015; Lee *et al*., 2012; Atsumi *et al*., 2009a; Li *et al*., 2012; Choi *et al*., 2014; Smith *et al.,* 2010).

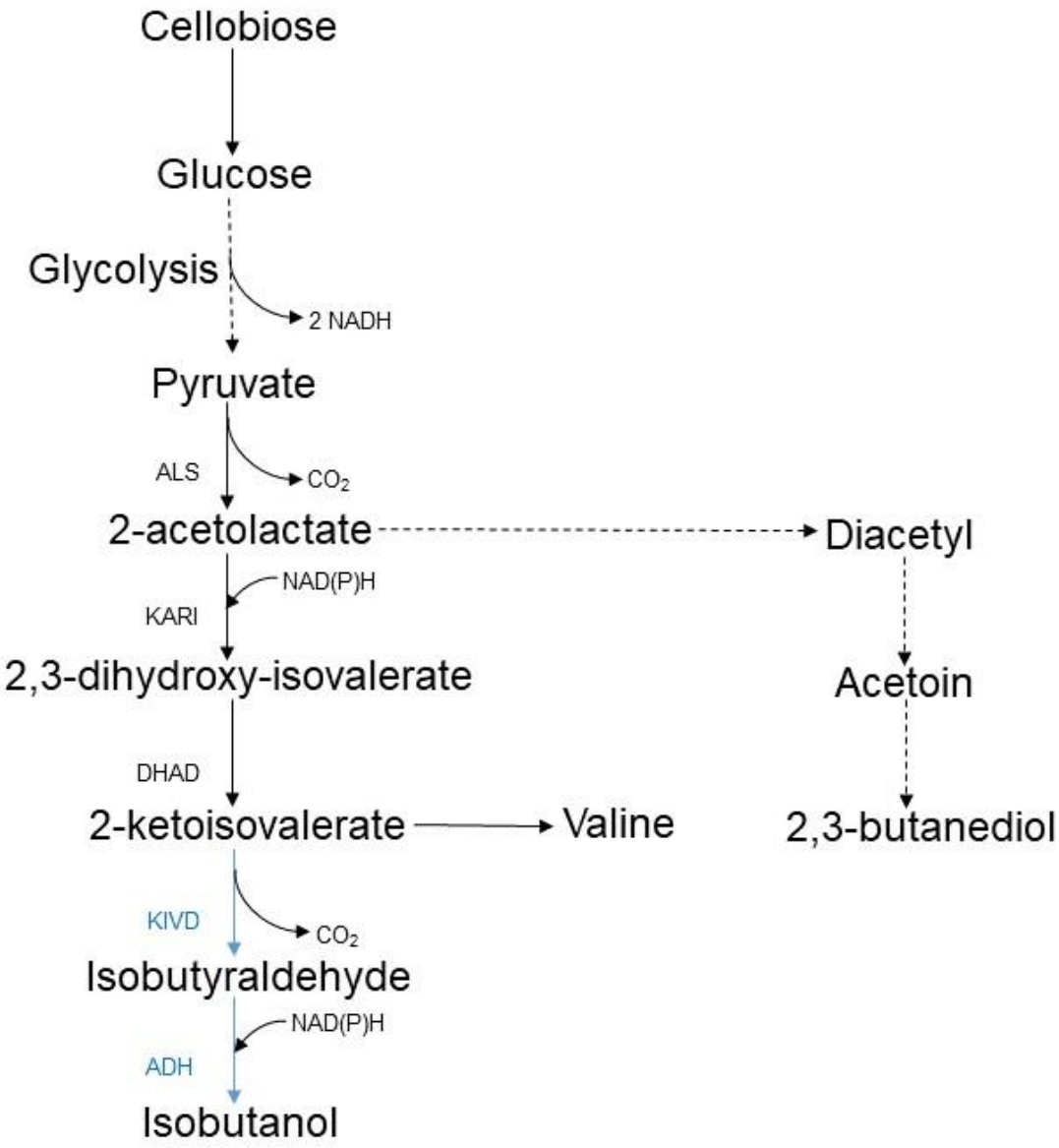


Figure 6.1: Isobutanol production pathway. Recombinant pathways are shown in blue. Enzyme abbreviations are ALS, acetolactate synthase; KARI, ketoacid reductoisomerase; DHAD, dihydroxyacid dehydratase; KIVD, 2-ketoisovalerate decarboxylase; ADH, alcohol dehydrogenase. Modified from Lin *et al*. (2014).

Additionally, the production of isobutanol via CBP has been reported by several researchers. Higashide *et al*. (2011) documented the production of isobutanol directly from cellulose by mesophilic *C. cellulolyticum* through CBP. Furthermore, Lin *et al.* (2015) also reported cellulosic isobutanol production using thermophilic *C. thermocellum* via CBP. In both cases, this was achieved by utilising the native cellulose hydrolysis ability of the organisms and the engineering of the amino acid biosynthesis pathway through the diversion of 2-ketoisovalerate to isobutyraldehyde and subsequent reduction to isobutanol. Hence, this success indicates that the strategy of CBP could be tested to produce isobutanol in *P. thermoglucosidasius* NCIMB 11955 and other thermophiles. Implementation of the isobutanol production pathway in *P. thermoglucosidasius* NCIMB 11955 could benefit from the theophylline-responsive CRISPR/Cas9 described in chapter four to inactivate competing pathways.

6.1.2 Chapter aims

This chapter aimed to engineer the metabolic pathway for cellulosic isobutanol production via the amino acid (valine) biosynthetic pathway in *P. thermoglucosidasius* NCIMB 11955. In the first instance, productions of isobutanol from glucose would be pursued to establish proof of concept before moving on to cellulosic substrates, drawing on the work of Bashir *et al.* (2019) who reported the successful expression of glycoside hydrolases (GHs) in *P. thermoglucosidasius* NCIMB 11955 for biomass utilisation to produce ethanol.

6.2 Results

To achieve isobutanol production, only two non-native steps are needed to bypass the intermediates from valine biosynthesis towards isobutanol. These steps are the decarboxylation of 2-ketoisovalerate to isobutyraldehyde by ketoisovalerate decarboxylase, and finally, the reduction of isobutyraldehyde to isobutanol by alcohol dehydrogenase.

In addition, the native acetolactate synthase (ALS), which catalyses the condensation of two pyruvate molecules to acetolactate, is anabolic, belonging to the branched-chain amino acid biosynthetic pathways, which are generally regulated through a feedback inhibition mechanism (Elisáková *et al.*, 2005). Hence, there is a need for a catabolic ALS, which is resistant to feedback inhibition. Previous studies indicate that *B. subtilis* ALS is catabolic and has a high affinity for pyruvate. Hence, it is a good candidate (Lin *et al.*, 2014; Zhou *et al.*, 2020).

6.2.1 Determination of active ketoisovalerate decarboxylase in *P. thermoglucosidasius*

It was first determined whether the KIVD was active before designing expression cassettes for the pathway. Thus, the activity of native/wild-type KIVD from *L. lactis* and its variants, namely, KIVD (LLKF_1386) (Lin *et al.*, 2014), KIVD [LLKF (mut.)] (LLKF_1386 with 20 bp [1451-1471] swapped with wild-type sequence because of homology with *B. subtilis* ALS), KIVD (LLM4) (Soh *et al.*, 2017), and KIVD (WT_LLM4) [LLKF sequence containing four amino acid mutations (Q34H, V130I, A290V, S386P) of LLM4], under the control of P_{gapd} and P_{ldh} promoters, respectively, were investigated in *P. thermoglucosidasius*. The KIVD (LLKF_1386) was named according to its accession number, LLKF_1386, and differ from the wild-type KIVD

by seven amino acids, with a T_{50} (the temperature at which the enzyme loses half of its activity upon a 20 min incubation) of 57 °C (Lin *et al.*, 2014). The LLM4 variant contained top four single-mutation variants generated random mutagenesis and has a T_{50} of 60.4 °C (Soh *et al.*, 2017).

6.2.1.1 Construction of plasmids for ketoisovalerate decarboxylase expression

The gene encoding alpha-ketoisovalerate decarboxylase (*kivD*) was amplified using the PCR from *L. lactis* genomic DNA (*L. lactis* supplied by Sarah Chapman, University of Nottingham). The ~ 1.6 kb native *kivD* gene was ligated to glyceraldehyde-3-phosphate dehydrogenase (P_{gapd}) and lactate dehydrogenase (P_{ldh}) promoters, respectively, using SOEing PCR. The P_{gapd}*_kivD* and P_{ldh}*_kivD* fragments were ligated between the NotI/XhoI, and SbfI/XhoI sites of the vector pMTL61110 (Sheng *et al.*, 2017), respectively, to yield plasmids pMTL-KIVD1 and pMTL-KIVD2, followed by transformation in *E. coli* Top10. The same procedure was carried out to construct plasmids pMTL-LLKF1, pMTL-LLKF2 under the control of the P_{gapd} promoter, and pMTL-LLKF3, pMTL-LLKF4, pMTL-LLM4, and pMTL-WT_LLM4 under the control of the P_{ldh} promoter, using synthesised variants of the wild-type *kivD* gene by Azenta Life Sciences Ltd, namely; *kivD* (LLKF_1386), *kivD* [LLKF (mut.)], *kivD* (LLM4), and *kivD* (WT_LLM4) genes, respectively. Colony PCR was carried out to screen for the presence of the respective gene fragments in the *E. coli* Top10 transformants (Figure 6.2). Sanger sequencing confirmed the presence of the expected nucleotide sequences in the various plasmids. Furthermore, all the nucleotide sequences correspond to alpha-keto acid decarboxylase from *L. lactis* when translated using the ExPASy tool (https://web.expasy.org/translate/) and blast search carried out using the Blastp algorithm of the NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

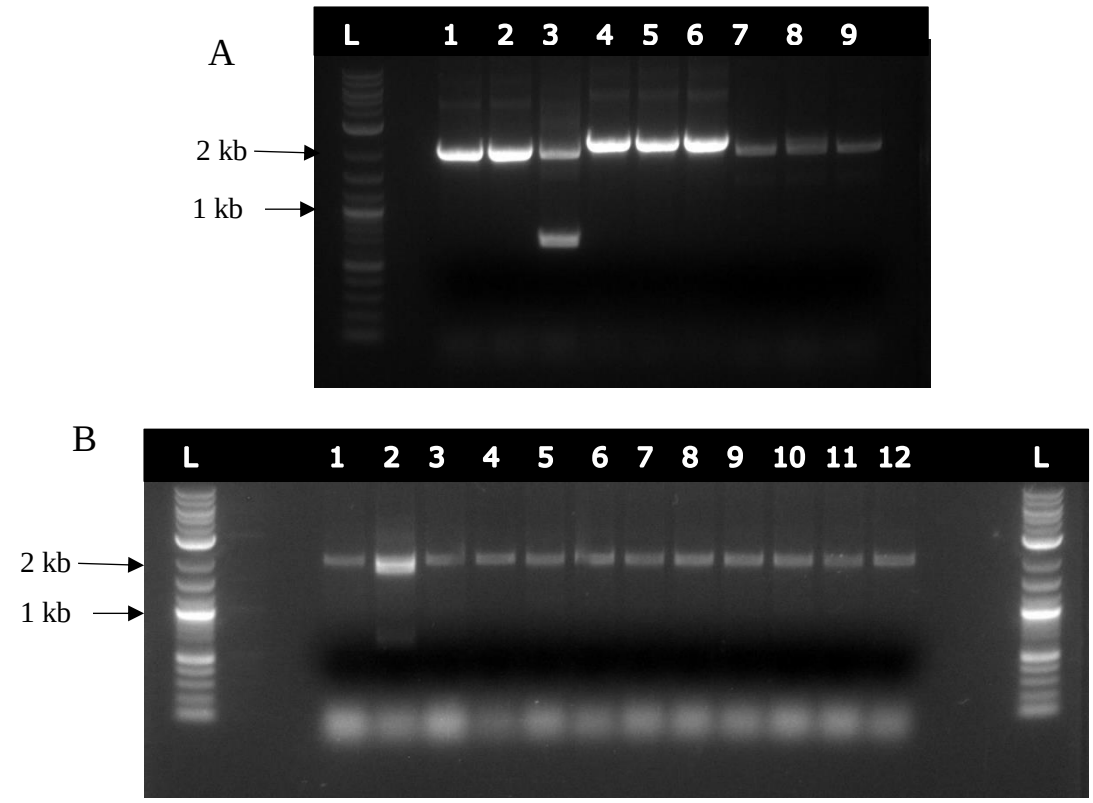


Figure 6.2: Colony-PCR for screening of *kivD* expression plasmids in transformed *E. coli* Top10. A) L= DNA Ladder (0.1-10 kb), lanes 1-3= pMTL-LLKF1, lanes 4-6 = pMTL-LLKF2, and lanes 7-9 = pMTL-KIVD1. Primers used: pGAPD_Seq and RepB_R1. Expected size of inserts ~ 2 kb. B) L= DNA Ladder (0.1-10 kb), lanes 1-3= pMTL-LLKF3, lanes 4-6 = pMTL-LLKF4, lanes 7-9 = pMTL-KIVD2, and lanes 10-12 = pMTL-LLM4. Primers used: pLDH_Seq and RepB_R1. Expected size of inserts ~ 2 kb.

6.2.1.2 Development of recombinant *P. thermoglucosidasius* strains expressing *kivD*

To produce strains expressing *kivD* genes on autonomous plasmids, pMTL-LLKF1, pMTL-LLKF2, pMTL-KIVD1, pMTL-LLKF3, pMTL-LLKF4, pMTL-KIVD2, pMTL-LLM4, and pMTL-WT_LLM4 plasmids were transformed into electro-competent cells of *P. thermoglucosidasius* NCIMB 11955 by electroporation as described in Methods section 2.10.4. Colonies of the resulting transformants were subsequently chosen and screened for the presence of P_{gapd}*_kivD* and P_{ldh}*_kivD* fragments by colony PCR screening using oligonucleotide primers pGAPD_Seq and RepB_R1 and pLDH_Seq and RepB_R1 respectively, (primer sequences in Table 2.1), which anneal to the promoters upstream of the *kivD* gene and *repB* replicon on the plasmid backbone. As observed, the ~ 2 kb band size corresponding to the P_{gapd}*_kivD* and P_{ldh}*_kivD* from the transformants on agarose gel confirms the presence of the plasmids harbouring the *kivD* gene in the recombinant *P. thermoglucosidasius* strains (Figure 6.3). Additionally, Sanger sequencing of correct PCR products confirms the nucleotide sequence of the gene fragments. The recombinant strains generated were designated as *P. thermoglucosidasius* AK1, AK2, AK3, AK4, AK5, AK6, AK7, and AK8 strains, respectively.

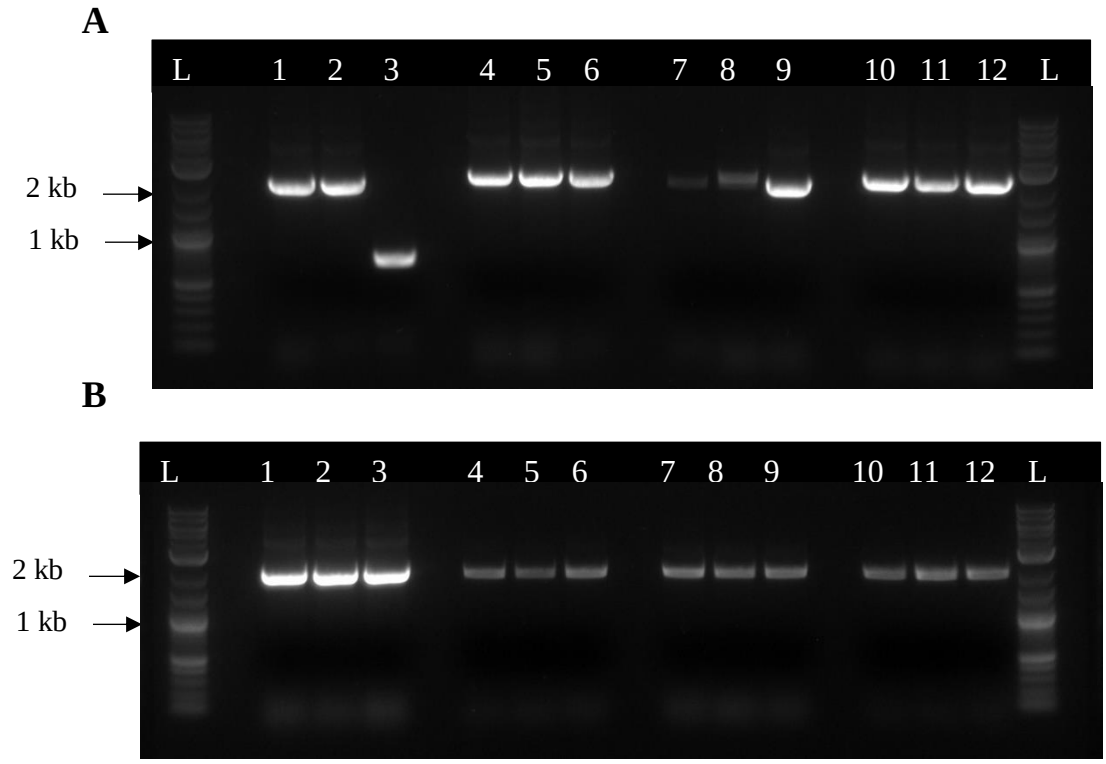


Figure 6.3: Colony-PCR for screening of *kivD* expression plasmids in recombinant strains of *P. thermoglucosidasius*. (A) L= DNA Ladder (0.1-10 kb), lanes 1-3 = AK1- pMTL-LLKF2 [P_{gapd}_*kivD* [LLKF (mut.)], lanes 4-6 = AK2- pMTL-LLKF1 [P_{gapd}_*kivD* (LLKF_1386)], lanes 7-9 = AK3- pMTL-KIVD1 (P_{gapd}_*kivD*), and lanes 10-12 = AK4- pMTL-LLKF4 [P_{ldh}_*kivD* [LLKF (mut.)]. Primers used: pGAPD_Seq and RepB_R1 for AK1, AK2, and AK3, while pLDH_Seq and RepB_R1 for AK4. Expected size of inserts ~ 2 kb. **(B)** L= DNA Ladder (0.1-10 kb), lanes 1-3 = AK5- pMTL-LLKF3 [P_{ldh}_*kivD* (LLKF_1386)], lanes 4-6 = AK6- pMTL-KIVD2 (P_{ldh}_*kivD*), lanes 7-9 = AK7- pMTL-LLM4 [P_{ldh}_*kivD* (LLM4)], and lanes 10-12 = AK8- pMTL-WT_LLM4 [P_{ldh}_*kivD* (WT_LLM4)]. Primers used: pLDH_Seq and RepB_R1. Expected size of inserts ~ 2 kb.

6.2.1.3 Characterization of recombinant *P. thermoglucosidasius* AK1, AK2, AK3, AK4, AK5, AK6, AK7, and AK8 strains

To determine the activity of the various ketoisovalerate decarboxylase in the recombinant *P. thermoglucosidasius* AK1, AK2, AK3, AK4, AK5, AK6, AK7, and AK8 strains, and wild-type *P. thermoglucosidasius* as control, fermentation experiments were performed at 50 ˚C for 48 h using M-ASYE medium + 20 $g \cdot l^{-1}$ glucose supplemented with 5 $g \cdot l^{-1}$ 2-ketoisovalerate, which is a precursor of isobutyraldehyde. Culture supernatants from these strains were used for quantification of isobutanol production by HPLC as described in the Methods section 2.11.1. Isobutanol titres of 9.1 mM, 8.1 mM, and 8.5 mM were observed for strains AK1 (P_{gapd}_*kivD* [LLKF (mut.)]), AK2 [P_{gapd}_*kivD* (LLKF_1386)], and AK5 [P_{ldh}_*kivD* (LLKF_1386)] respectively when 2-Ketoisovalerate was added to the medium. In comparison, strains AK3, AK4, AK6, AK7, and AK8 produced only 0.4 mM isobutanol comparable to the wild-type strain (Figure 6.4). These observations indicate that the KIVD (LLKF_1386) was expressed and functional; hence, the conversion of supplemental 2-ketoisovalerate to isobutyraldehyde and finally reduced to isobutanol.

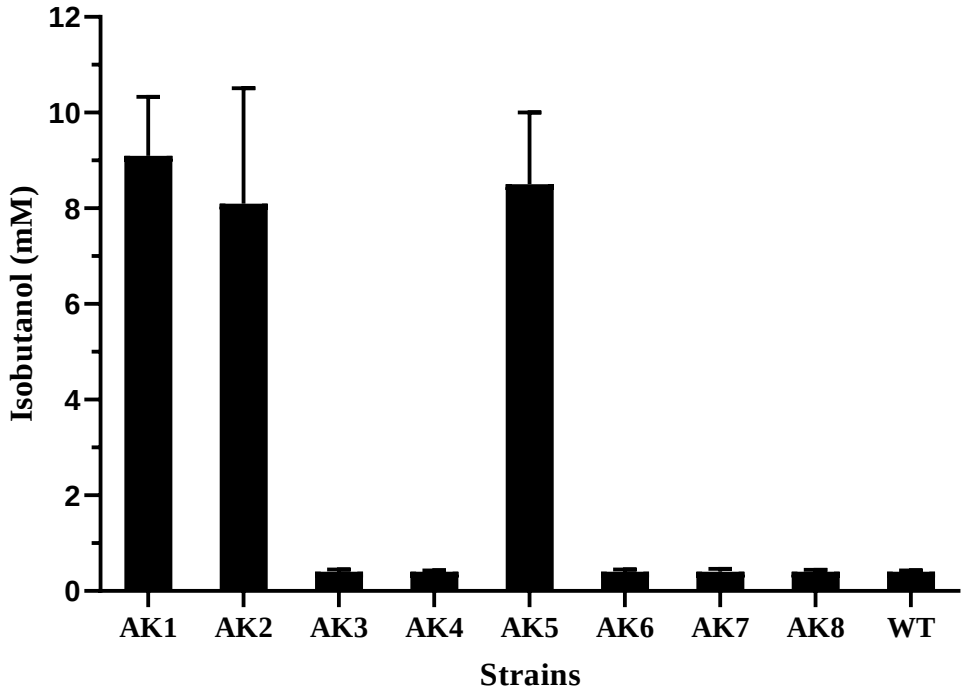


Figure 6.4: Isobutanol (mM) produced by the recombinant strains of *P. thermoglucosidasius* as measured by HPLC to determine active ketoisovalerate decarboxylase. AK1, AK2 and AK3 denote strains expressing pMTL-LLKF2 [P_{gapd}_*kivD* {LLKF (mut.)}], pMTL-LLKF1 [P_{gapd}_*kivD* (LLKF_1386)], and pMTL-KIVD1 (P_{gapd}_*kivD*), respectively. AK4, AK5, AK6, AK7 and AK8 denote strains expressing pMTL-LLKF4 [P_{ldh}_*kivD* {LLKF (mut.)}], pMTL-LLKF3 [P_{ldh}_*kivD* (LLKF_1386)], pMTL-KIVD1 (P_{ldh}_*kivD*), pMTL-LLM4 [P_{ldh}_*kivD* (LLM4)], and pMTL-WT_LLM4 [P_{ldh}_*kivD* (WT_LLM4)], respectively. WT denote Wild-type *P. thermoglucosidasius* strain. In three biological replicates, the recombinant strains and wild-type isobutanol profile were characterised by HPLC after 48 h of fermentation using 10 ml of M-ASYE with 20 $g \cdot l^{-1}$ glucose and 10 $g \cdot l^{-1}$ yeast extract+5 $g \cdot l^{-1}$ 2-ketoisovalerate+12.5 $\mu g \cdot ml^{-1}$ Kan in 50 ml falcon tubes at 50 $^{\circ}C$, 250 rpm. Results are shown as mean ± SD of three biological replicates.

6.2.2 Development of recombinant *P. thermoglucosidasius* strains for isobutanol production

After determining P_{gapd}_*kivD* [LLKF (mut.)]), P_{gapd}_*kivD* (LLKF_1386)], and P_{ldh}_*kivD* (LLKF_1386) to be expressed and functional, experiments proceeded to establish isobutanol production in *P. thermoglucosidasius*. Different combinations of P_{gapd}_*kivD_alsS* and P_{ldh}_*kivD_alsS* operons were constructed by utilising three different catabolic acetolactate synthase (*alsS*) genes from the mesophilic *B. subtilis* str. 168 (accession number: CP053102.1) and the thermophiles *Bacillus coagulans* (accession number: CP009709.1) and *Thermoactinomyces daqus* (accession number: GCA_000763315.1).

6.2.2.1 Construction of P_{gapd}_*kivD_alsS* and P_{ldh}_*kivD_alsS* plasmids

The genes encoding ALS were amplified using the PCR from synthesised *B. subtilis* (ALS1), *B. coagulans* (ALS2), and *T. daqus* (ALS3) DNA employing ALS1_F_XhoI/ALS1_R_NheI, ALS2_F_XhoI/ALS2_R_NheI, and ALS3_F_XhoI /ALS3_R_NheI, primers (detailed primer sequences in Table 2.1). The ~ 1.7 kb $alsS_{bs}$, $alsS_{bc}$, $alsS_{td}$, DNA fragments were ligated between the XhoI/NheI sites of the vector pMTL-LLKF2, which contained the P_{gapd}_*kivD* [LLKF (mut.)] fragment upstream of the XhoI site, thereby yielding plasmids pMTL-AM1, pMTL-AM2, and pMTL-AM3, respectively (Figure 6.5). The same procedure was carried out to construct pMTL-AM4, pMTL-AM5, and pMTL-AM6 (Figure 6.5), using pMTL-LLKF1, which contained P_{gapd}_*kivD* (LLKF_1386) fragment. Similarly, plasmids pMTL-AM7, pMTL-AM8, and pMTL-AM9 (Figure 6.5) were constructed using pMTL-LLKF3 containing P_{ldh}_*kivD* (LLKF_1386) fragment. Complete plasmid maps are presented in appendix VIII.

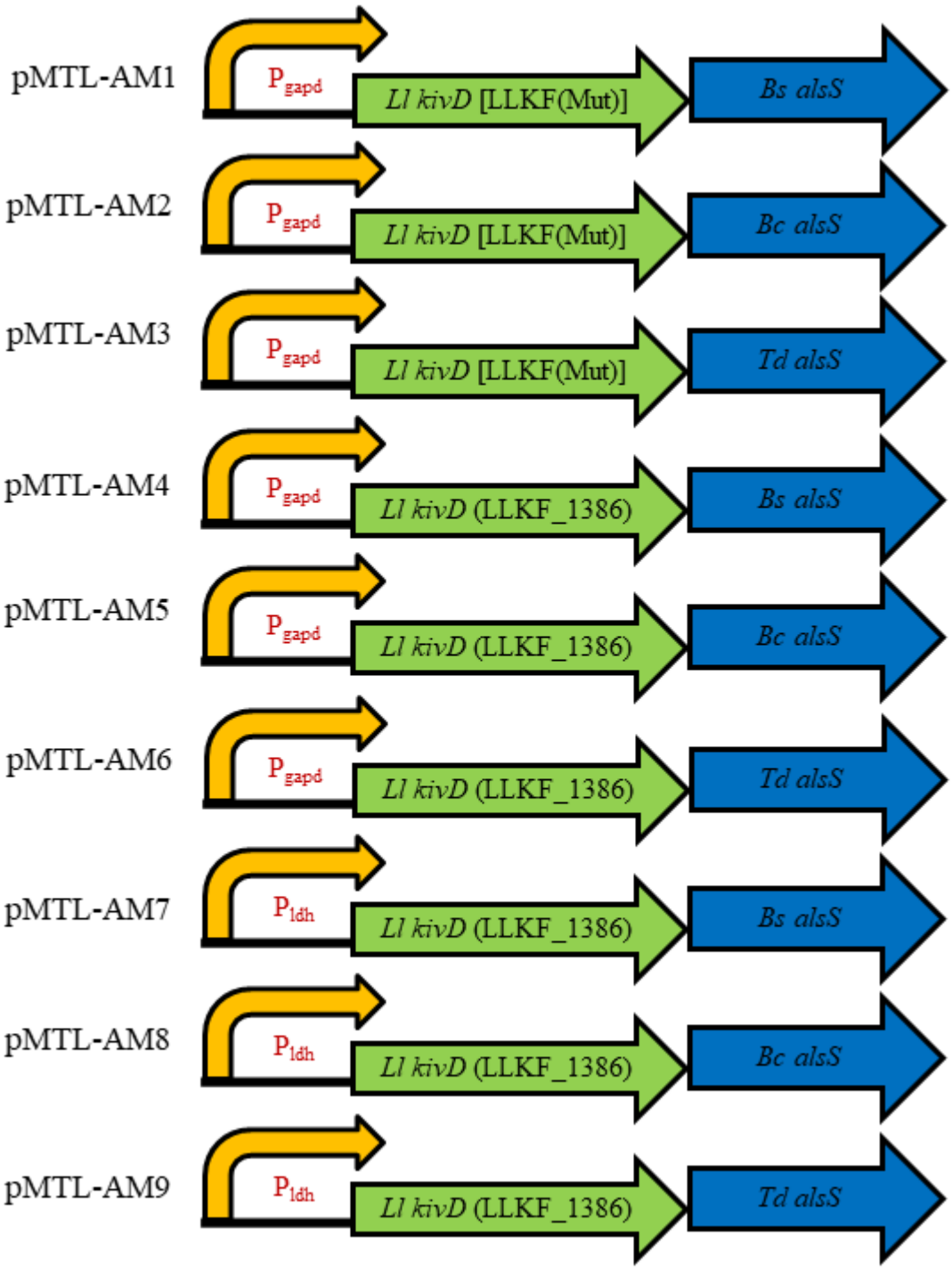


Figure 6.5: Schematic of P_{gapd}_*kivD_alsS* and P_{ldh}_*kivD_alsS* operons. P_{gapd}: glyceraldehyde-3-phosphate dehydrogenase promoter, P_{ldh}: lactate dehydro-genase promoter, *Ll*: *L. lactis*, *kivD*: ketoisovalerate decarboxylase gene, *Bs*: *B. subtilis*, *Bc*: *B. coagulans*, *Td*: *T. daqus*, *alsS*: acetolactate synthase gene.

To select the transformed colonies from *E. coli* Top10 carrying the respective plasmids, colony PCR was carried out to screen for the presence of the respective P_{gapd}*_kivD_alsS* and P_{ldh}*_kivD_alsS* operons using plasmid-based ColE1_F2 and RepB_R1 primers (detailed primer sequence Table 2.1) as shown in Figure 6.6. Sanger sequencing confirmed the presence of the expected nucleotide sequences in the plasmids.

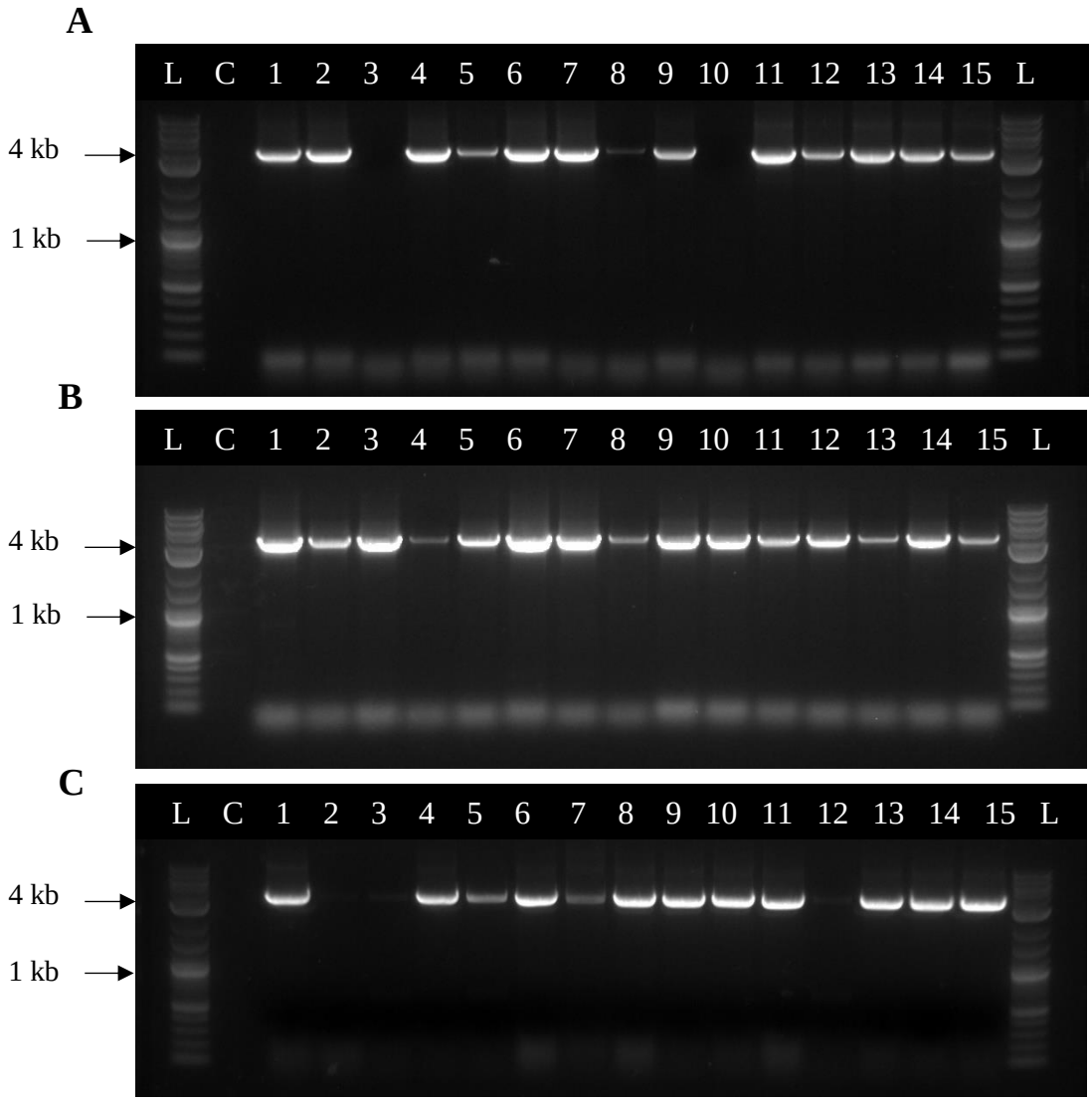


Figure 6.6: Colony-PCR screening of P$_{gapd}$_*kivD_alsS* and P$_{ldh}$_*kivD_alsS* operons in recombinant *E. coli* Top10. (A) L=DNA Ladder (0.1-10 kb), C=Wild-type *E. coli* Top10 as negative control, lanes 1-5=pMTL-AM1, lanes 6-10=pMTL-AM2, and lanes 11-15=pMTL-AM3. Expected size of P$_{gapd}$_*kivD* (LLKF (Mut.)_*alsS* fragment ~ 4 kb. **(B)** C=Wild-type *E. coli* Top10 as negative control, lanes 1-5=pMTL-AM4, lanes 6-10=pMTL-AM5, and lanes 11-15=pMTL-AM6. Expected size of P$_{gapd}$_*kivD* (LLKF_1386)_*alsS* fragment ~ 4 kb. **(C)** C=Wild-type *E. coli* Top10 as negative control, lanes 1-5=pMTL-AM7, lanes 6-10=pMTL-AM8, and lanes 11-15=pMTL-AM9. Expected size of P$_{ldh}$_*kivD* (LLKF_1386)_*alsS* fragment ~ 4 kb.

6.2.2.2 Construction of recombinant *P. thermoglucosidasius* AM001, AM002, AM003, AM004, AM005, AM006, AM007, AM008, and AM009 strains

After confirming the presence of correct *kivD_alsS*$_{bs}$, *kivD_alsS*$_{bc}$, *kivD_alsS*$_{td}$, operons by nucleotide sequencing, plasmids pMTL-AM1, pMTL-AM2, and pMTL-AM3 expressing P$_{gapd}$_*kivD* [LLKF (Mut.)]_*alsS*$_{bs}$, P$_{gapd}$_*kivD* [LLKF (Mut.)]_*alsS*$_{bc}$, and P$_{gapd}$_*kivD* [LLKF (Mut.)]_*alsS*$_{td}$, respectively were transformed into electro-competent cells of *P. thermoglucosidasius* NCIMB 11955 as described in Methods section 2.10.4. The recombinant strains generated were designated as *P. thermoglucosidasius* AM001, AM002, and AM003 strains. Similarly, plasmids pMTL-AM4, pMTL-AM5, and pMTL-AM6 expressing P$_{gapd}$_*kivD* (LLKF_1386)_*alsS*$_{bs}$, P$_{gapd}$_*kivD* (LLKF_1386) _*alsS*$_{bc}$, and P$_{gapd}$_*kivD* (LLKF_1386)_*alsS*$_{td}$, respectively, were transformed into electro-competent cells of *P. thermoglucosidasius*, and the recombinant strains generated were designated as *P. thermoglucosidasius* AM004, AM005, and AM006 strains. Plasmids pMTL-AM7, pMTL-AM8, and pMTL-AM9 expressing P$_{ldh}$_*kivD* (LLKF_1386)_*alsS*$_{bs}$, P$_{ldh}$_*kivD* (LLKF_1386)_*alsS*$_{bc}$, and P$_{ldh}$_*kivD* (LLKF_1386) _*alsS*$_{td}$, respectively, were additionally transformed into electro-competent cells of *P. thermoglucosidasius*, and the recombinant strains generated were designated as *P. thermoglucosidasius* AM007, AM008, and AM009 strains. Colony PCR was carried out to confirm the presence of the *kivD_alsS* operons using plasmid-based oligonucleotide primers ColE1_F2 and RepB_R1 (detailed primer sequence in Table 2.1) which generated a ~ 4 kb product for the colonies screened (three for each strain) (Figure 6.7). The PCR products were purified, and Sanger sequencing confirmed their sequence as correct.

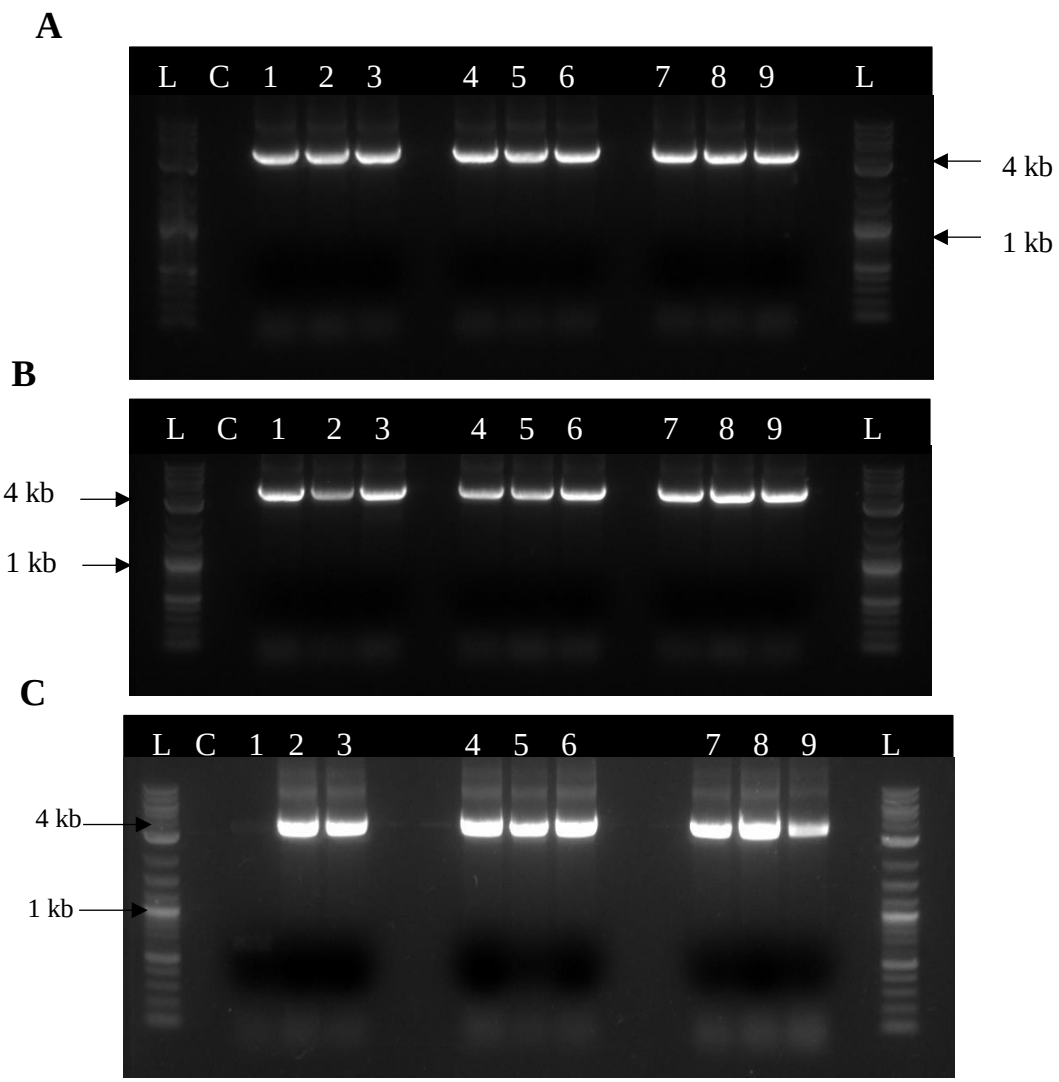


Figure 6.7: Colony-PCR screening of P$_{gapd}$_*kivD_alsS* and P$_{ldh}$_*kivD_alsS* operons in recombinant *P. thermoglucosidasius*. (A) L=DNA Ladder (0.1-10 kb), C=Wild-type *P. thermoglucosidasius* as negative control, lanes 1-3=AM001- P$_{gapd}$_*kivD* [LLKF (mut.)]_*alsS*$_{bs}$; lanes 4-6=AM002- P$_{gapd}$_*kivD* [LLKF (mut.)]_*alsS*$_{bc}$; and lanes 7-9=AM003- P$_{gapd}$_*kivD* [LLKF (mut.)]_*alsS*$_{td}$. **(B)** C=Wild-type *P. thermoglucosidasius* as negative control, lanes 1-3=AM004- P$_{gapd}$_*kivD* (LLKF_1386)_*alsS*$_{bs}$; lanes 4-6=AM005- P$_{gapd}$_*kivD* (LLKF_1386) _*alsS*$_{bc}$; and lanes 7-9=AM006- P$_{gapd}$_*kivD* (LLKF_1386)_*alsS*$_{td}$. **(C)** C=Wild-type *P. thermoglucosidasius* as negative control, lanes 1-3=AM007- P$_{ldh}$_*kivD* (LLKF_1386)_*alsS*$_{bs}$; lanes 4-6=AM008- P$_{ldh}$_*kivD* (LLKF_1386) _*alsS*$_{bc}$; and lanes 7-9=AM009- P$_{ldh}$_*kivD* (LLKF_1386)_*alsS*$_{td}$. Expected size of P$_{gapd}$_*kivD_alsS* and *P*$_{ldh}$_*kivD_alsS* operons ~ 4 kb.

6.2.2.3 Characterization of recombinant *P. thermoglucosidasius* AM001, AM002, AM003, AM004, AM005, AM006, AM007, AM008, and AM009 strains

The production of isobutanol by the recombinant *P. thermoglucosidasius* AM001, AM002, AM003, AM004, AM005, AM006, AM007, AM008, and AM009 strains and wild-type *P. thermoglucosidasius* as control was investigated. Three replicates of each strain were grown in M-ASYE with 20 $g \cdot l^{-1}$ glucose and 10 $g \cdot l^{-1}$ yeast extract at 50 ˚C for 48 h in sealed 50 ml falcon tubes, according to the methods described by Lin *et al.* (2014). Culture supernatants from the growth of these strains were used to quantify isobutanol production by HPLC as described in the Methods section 2.11.1. Isobutanol was not detected in any of the samples analysed for strains AM001, AM002, AM003, AM004, AM005, and AM006 containing P_{gapd}*_kivD* [LLKF (mut.)]*_alsS*$_{bs}$, P_{gapd}*_kivD* [LLKF (mut.)]*_alsS*$_{bc}$, P_{gapd}*_kivD* [LLKF (mut.)]*_alsS*$_{td}$, P_{gapd}*_kivD* (LLKF_1386)*_alsS*$_{bs}$, P_{gapd}*_kivD* (LLKF_1386)*_alsS*$_{bc}$, and P_{gapd}*_kivD* (LLKF_1386) *_alsS*$_{td}$ operons. To check the stability and confirm the presence of the plasmids, the recombinant strains were sub-cultured on TSA+Kan plates and the resulting colonies screened by colony PCR for the respective operons. This demonstrated that the recombinant strains retained the plasmids. Additionally, Sanger sequencing confirmed their sequences as correct. This observation suggested that the P_{gapd} promoter was probably inactive.

When *alsS*$_{bs}$, *alsS*$_{bc}$, and *alsS*$_{td}$ was overexpressed along with *kivD* under the control of P_{ldh} promoter, 0.3, 0.4, and 0.3 mM isobutanol was produced by the recombinant *P. thermoglucosidasius* AM007, AM008, and AM009 strains, respectively (Figure 6.8) from glucose in 48 h at 50 ˚C. Simultaneously, 13.8, 19.1, and 4.3 mM total 2,3-butanediol (BDO) was produced by the recombinant strains compared to the 1.4 mM produced by the wild-type. BDO is a reduced form of acetoin, which can be produced

from acetolactate via spontaneous cleavage at high temperatures (Zhou *et al.,* 2020; Xiao *et al.,* 2012). The BDO observed suggested that the metabolic flux from acetolactate to 2,3-dihydroxy-isovalerate was potentially insufficient for isobutanol production. One way to possibly increase the flux towards isobutanol was to overexpress the ketoacid reductoisomerase (*ilvC*) gene.

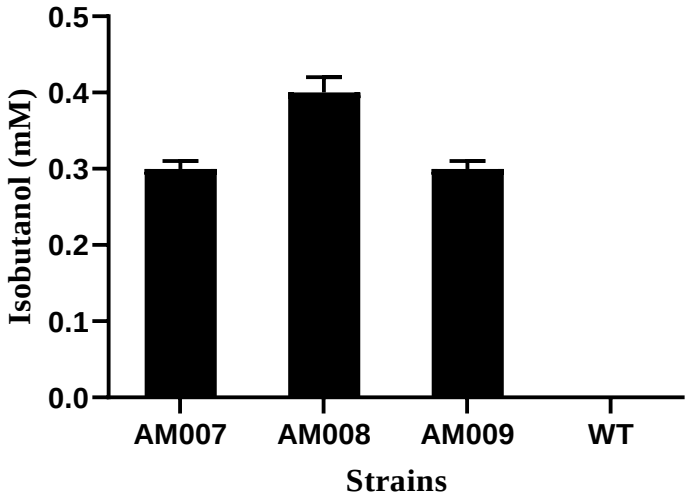


Figure 6.8: Concentration of isobutanol (mM) produced by the recombinant strains of *P. thermoglucosidasius* transformed with P_{ldh}_*kivD_alsS*. AM007, AM008, and AM009 denotes strains expressing P_{ldh}_*kivD* (LLKF_1386) _$alsS_{bs}$, P_{ldh}_*kivD* (LLKF_1386) _$alsS_{bc}$, and P_{ldh}_*kivD* (LLKF_1386) _$alsS_{td}$, respectively. WT denote wild-type *P. thermoglucosidasius*. In case of the *alsS* gene designations, the subscripts indicate the source of the gene, viz., bs, *B. subtilis*, bc, *B. coagulans*, td, *T. daqus*. In three biological replicates, the recombinant strains and wild-type's isobutanol profiles were characterised by HPLC after 48 h of fermentation using 10 ml of M-ASYE with 20 $g \cdot l^{-1}$ glucose and 10 $g \cdot l^{-1}$ yeast extract +12.5 $\mu g \cdot ml^{-1}$ Kan in 50 ml falcon tubes at 50 $^{\circ}$C, 250 rpm. Results are shown as mean ± SD of three biological replicates.

6.2.3 Effect of overexpression of the ketoacid reductoisomerase (*ilvC*) gene

The *ilvC* gene was integrated at the *pyrE* locus of *P. thermoglucosidasius* using ACE (Sheng *et al.,* 2017) for overexpression. The generated strain was designated *P. thermoglucosidasius::ilvC*; work carried out by Dr Lili Sheng, University of Nottingham (unpublished).

6.2.3.1 Construction of recombinant *P. thermoglucosidasius* AM007b, AM008b, and AM009b strains

Plasmids pMTL-AM7, pMTL-AM8, and pMTL-AM9 expressing P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bs}$, P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bc}$, and P_{ldh}*_kivD* (LLKF_1386) *_alsS*$_{td}$, respectively, were transformed into electro-competent cells of *P. thermoglucosidasius::ilvC* (kindly provided by Dr Lili Sheng) as described in Methods section 2.10.4. Transformant colonies were screened by colony PCR to confirm the presence of the P_{ldh}*_kivD* (LLKF_1386)*_alsS* operons using plasmid-based ColE1_F2 and RepB_R1 oligonucleotide primers (detailed primer sequence in Table 2.1) which generated a ~ 4 kb product for the colonies screened (three for each strain) (Figure 6.9). The PCR products were gel purified, and Sanger sequenced to confirm the presence of the correct nucleotide sequences. Furthermore, the strains generated were designated as *P. thermoglucosidasius* AM007b, AM008b, and AM009b strains expressing genome-integrated *ilvC* at *pyrE* locus along with the autonomous plasmid-based expression of P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bs}$, P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bc}$, and P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{td}$, respectively.

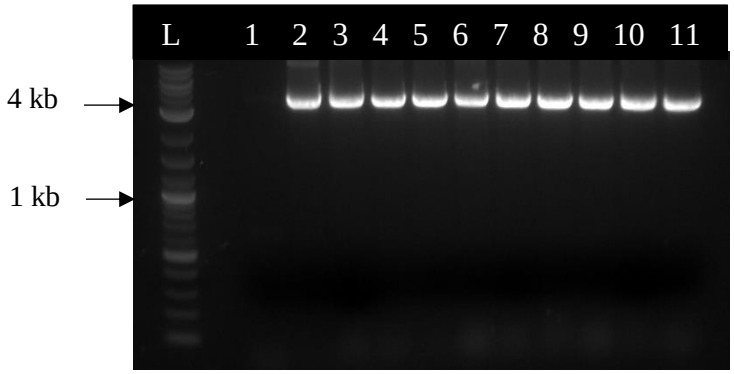


Figure 6.9: Colony-PCR screening of P*ldh*_*kivD_alsS* operons in recombinant *P. thermoglucosidasius::ilvC*. L=DNA Ladder (0.1-10 kb), lane 1=*P. thermoglucosidasius::ilvC* as negative control, lane 2=pMTL-AM7 vector as positive control, lanes 3-5=AM007b - P*ldh*_*kivD* (LLKF_1386)_*alsSbs*, lanes 6-8=AM008b - P*ldh*_*kivD* (LLKF_1386)_*alsSbc*, and lanes 9-11=AM009b -P*ldh*_*kivD* (LLKF_1386) _*alsStd*. In case of the *alsS* gene designations, the subscripts indicate the source of the gene, viz., bs, *B. subtilis*, bc, *B. coagulans*, td, *T. daqus*. Expected size of P*ldh*_*kivD_alsS* operons ~ 4 kb.

6.2.3.2 Characterization of recombinant *P. thermoglucosidasius* AM007b, AM008b, and AM009b strains

To evaluate the effect of overexpressing *ilvC* in the genome, isobutanol production by the recombinant *P. thermoglucosidasius* AM007b, AM008b, and AM009b strains, and the *P. thermoglucosidasius::ilvC* control strain, was investigated. The strains were grown in three biological replicates in M-ASYE with 20 $g \cdot l^{-1}$ glucose and 10 $g \cdot l^{-1}$ yeast extract at 50 $^{\circ}C$ for 48 h in sealed 50 ml falcon tubes according to the Methods described by Lin *et al.* (2014). Culture supernatants from these strains were used to quantify isobutanol production by HPLC as described in the Methods section 2.11.1. The AM008b strain with P*ldh*_*kivD* (LLKF_1386)_*alsSbc* and *ilvC* integrated at the *pyrE*

locus on the genome achieved the highest isobutanol production of 0.6 mM, while *ilvC* made no difference in strains AM007b and AM009b (Figure 6.10). Additionally, no 2,3-butanediol production was observed for all the strains. This, therefore, indicates that the overexpression of *ilvC* at the *pyrE* locus on the genome might still be insufficient due to low copy number or the native flux down the pathway (2,3-dihydroxy-isovalerate to 2-ketoisovalerate, and isobutyraldehyde to isobutanol) are still inadequate to produce considerable isobutanol. The latter was chosen and tested to see whether overexpression of native dihydroxy acid dehydratase (*ilvD*) and alcohol dehydrogenase (*adhA*) genes could increase isobutanol production.

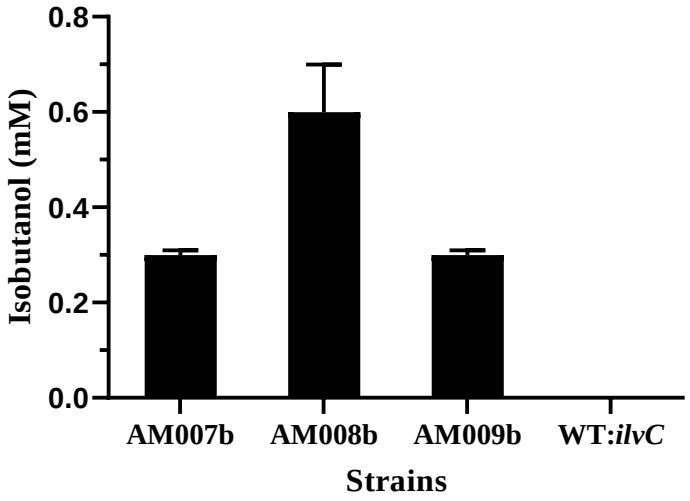


Figure 6.10: Concentration of isobutanol (mM) produced by the recombinant strains of *P. thermoglucosidasius::ilvC* transformed with P_{ldh}*_kivD_alsS* operon. AM007b, AM008b, and AM009b denotes strains expressing genome integrated *ilvC* at *pyrE* locus along with plasmid-based expression of P_{ldh}*_kivD* (LLKF_1386)*_*$alsS_{bs}$, P_{ldh}*_kivD* (LLKF_1386)*_*$alsS_{bc}$, and P_{ldh}*_kivD* (LLKF_1386)*_*$alsS_{td}$, respectively. WT:*ilvC* denote wild-type *P. thermoglucosidasius::ilvC*. In case of the *alsS* gene designations, the subscripts indicate the source of the gene, viz., bs, *B. subtilis*, bc, *B. coagulans*, td, *T. daqus*. In three biological replicates, the recombinant strains and wild-type's isobutanol profiles were characterised by HPLC after 48 h of fermentation using 10 ml of M-ASYE with 20 $g \cdot l^{-1}$ glucose and 10 $g \cdot l^{-1}$ yeast extract +12.5 $\mu g \cdot ml^{-1}$ Kan in 50 ml falcon tube at 50 $^{\circ}$C, 250 rpm. Results are shown as mean ± SD of three biological replicates.

6.2.4 Effect of overexpression of the dihydroxy acid dehydratase (*ilvD*) gene

To further increase the carbon flux towards isobutanol production, the native *P. thermoglucosidasius* dihydroxy acid dehydratase (*ilvD*) gene was also overexpressed. In essence, *ilvD* was cloned and overexpressed in the same operon downstream of *kivD* and *alsS* with genome-integrated *ilvC*.

6.2.4.1 Construction of pMTL-AM10, pMTL-AM11, and pMTL-AM12 plasmids

To construct *kivD* (LLKF_1386)_$alsS_{bs}$_*ilvD*, *kivD* (LLKF_1386)_$alsS_{bc}$_*ilvD*, and *kivD* (LLKF_1386)_$alsS_{td}$_*ilvD*, operons respectively, under the transcriptional control of P_{ldh} promoter, NEBuilder-HIFI DNA Assembly method was used as described in Methods section 2.9.11. Oligonucleotide primers were designed using NEBuilder Assembly Tool (https://nebuilder.neb.com/#!/), and the detailed sequences are presented in Table 2.1. The native *ilvD g*ene was amplified from *P. thermoglucosidasius* genomic DNA using ilvD_F_alsS1/ilvD_R_Cpa, ilvD_F_alsS2/ilvD_R_Cpa, and ilvD_F_alsS3 /ilvD_R_Cpa primers, respectively. Plasmids pMTL-AM7 (P_{ldh}_*kivD* (LLKF_1386)_$alsS_{bs}$), pMTL-AM8 (P_{ldh}_*kivD* (LLKF_1386)_$alsS_{bc}$), and pMTL-AM9 (P_{ldh}_*kivD* (LLKF_1386)_$alsS_{td}$) respectively, were linearized by restriction digestion using NheI enzyme. The linearized vectors and the amplified *ilvD* gene fragments were assembled using NEBuilder HIFI DNA Assembly Master Mix as described in Methods section 2.9.11, yielding plasmids pMTL-AM10 [P_{ldh}_*kivD* (LLKF_1386)_$alsS_{bs}$_*ilvD*], pMTL-AM11 [P_{ldh}_*kivD* (LLKF_1386)_$alsS_{bc}$_*ilvD*), and pMTL-AM12 [P_{ldh}_*kivD* (LLKF_1386)_$alsS_{td}$_*ilvD*] respectively (Figure 6.11). Complete plasmid maps are presented in appendix IX.

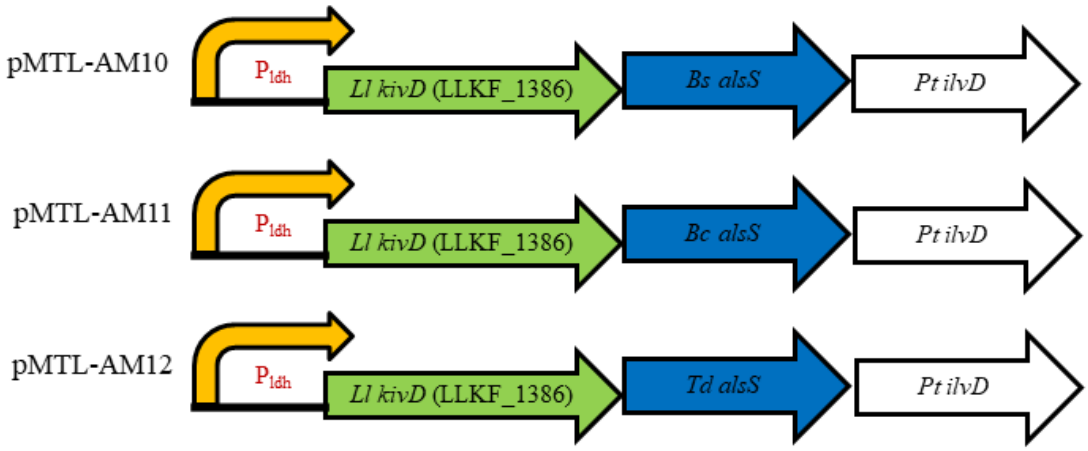


Figure 6.11: Schematic of P_{ldh}_*kivD_alsS_ilvD* operons in pMTL-AM10, pMTL-AM11 and pMTL-AM12 plasmids. P_{ldh}: lactate dehydrogenase promoter, *Ll*: *L. lactis*, *kivD*: ketoisovalerate decarboxylase gene, *Bs*: *B. subtilis*, *Bc*: *B. coagulans*, *Td*: *T. daqus*, *alsS*: acetolactate synthase gene, *Pt*: *P. thermoglucosidasius*, *ilvD*: dihydroxy acid dehydratase gene.

Four transformed *E. coli* Top10 colonies were grown overnight, and then plasmids pMTL-AM10, pMTL-AM11, and pMTL-AM12 were extracted to screen for the correct constructs. Purified pMTL-AM10, pMTL-AM11, and pMTL-AM12 plasmids were screened by PCR for the presence of *ilvD* gene fragments (Figure 6.12) using pairs of ilvD_F_alsS1/ilvD_R_Cpa, ilvD_F_alsS2/ilvD_R_Cpa, and ilvD_F_alsS3/ilvD_ R_ Cpa primers, respectively (detailed primer sequence Table 2.1). Additionally, Sanger sequencing confirmed the presence of the expected nucleotide sequences in the plasmids.

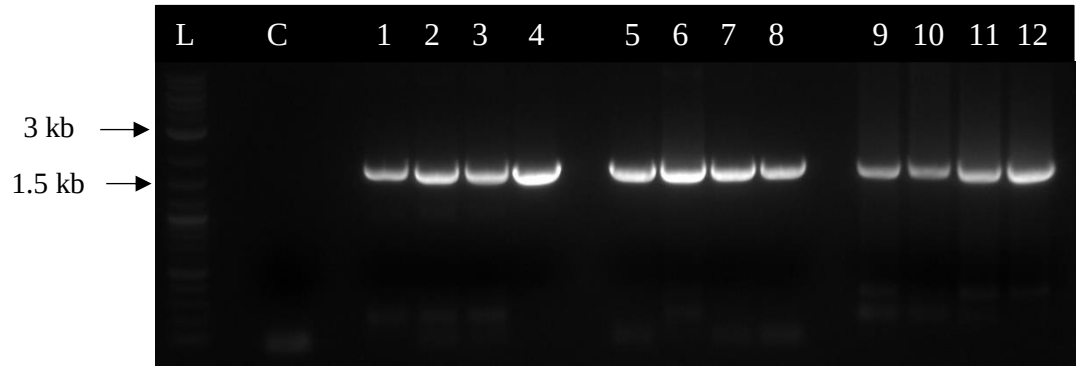


Figure 6.12: PCR screening of *ilvD* gene fragments from pMTL-AM10, pMTL-AM11, and pMTL-AM12 plasmids. L = DNA Ladder (0.1-10 kb), lane C = pMTL-AM7 plasmid as negative control, lanes 1-4 = pMTL-AM10 [P_{ldh}*_kivD* (LLKF_1386) *_alsS*$_{bs}$*_ilvD*], lanes 5-8 = pMTL-AM11 [P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bc}$*_ilvD*), and lanes 9-12 = pMTL-AM12 [P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{td}$*_ilvD*]. In case of the *alsS* gene designations, the subscripts indicate the source of the gene, viz., bs, *B. subtilis*, bc, *B. coagulans*, td, *T. daqus*. Expected size of *ilvD* gene fragments ~ 1.6 kb.

6.2.4.2 Construction of recombinant *P. thermoglucosidasius* AM010, AM011, and AM012 strains

After confirming the correct nucleotide sequence, plasmids pMTL-AM10, pMTL-AM11, and pMTL-AM12 expressing P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bs}$*_ilvD*, P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bc}$*_ilvD*, and P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{td}$*_ilvD*, respectively, were transformed into electro-competent cells of *P. thermoglucosidasius::ilvC* as described in Methods section 2.10.4. The transformants were screened by colony PCR to confirm the presence of the P_{ldh}*_kivD* (LLKF_1386)*_alsS_ilvD* operons using pLDH_Seq and RepB_R1 oligonucleotide primers (detailed primer sequence in Table 2.1) which are plasmid-based, generating a ~ 5.6 kb product for the colonies screened (three for each strain) (Figure 6.13). The PCR products were purified, and Sanger

sequenced to confirm the presence of the correct nucleotide sequences. The strains generated were designated as *P. thermoglucosidasius* AM010, AM011, and AM012 strains expressing genome-integrated *ilvC* at *pyrE* locus along with plasmid-based expression of P_{ldh}*_kivD* (LLKF_1386)*_*$alsS_{bs}$*_ilvD*, P_{ldh}*_kivD* (LLKF_1386) *_*$alsS_{bc}$*_ilvD*, and P_{ldh}*_kivD* (LLKF_1386)*_*$alsS_{td}$*_ilvD* operons, respectively.

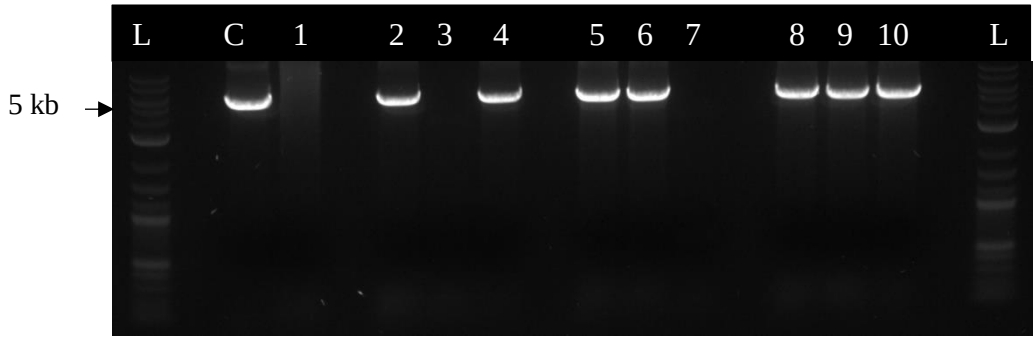


Figure 6.13: Colony-PCR screening of P_{ldh}*_kivD_alsS_ilvD* operons in recombinant *P. thermoglucosidasius::ilvC*. L = DNA Ladder (0.1-10 kb), C= pMTL-AM10 vector as positive control, lane 1 = *P. thermoglucosidasius::ilvC* as negative control, lanes 2-4 = AM010 - P_{ldh}*_kivD* (LLKF_1386)*_*$alsS_{bs}$*_ilvD*, lanes 5-7 = AM011 -P_{ldh}*_kivD* (LLKF_1386)*_*$alsS_{bc}$*_ilvD*, and lanes 8-10 = AM012 -P_{ldh}*_kivD* (LLKF_1386)*_*$alsS_{td}$*_ilvD*. In case of the *alsS* gene designations, the subscripts indicate the source of the gene, viz., bs, *B. subtilis*, bc, *B. coagulans*, td, *T. daqus*. Expected size of P_{ldh}*_kivD_alsS_ilvD* operons ~ 5.6 kb.

6.2.4.3 Characterization of recombinant *P. thermoglucosidasius* AM010, AM011, and AM012 strains

Isobutanol production by the recombinant *P. thermoglucosidasius* AM010, AM011, and AM012 strains, as well as the *P. thermoglucosidasius::ilvC* control strain, was investigated as described previously in sections 6.2.2.3 and 6.2.3.2. Interestingly, overexpression of *ilvD* alongside *kivD, alsS,* and *ilvC* only led to a slight increase in isobutanol production by the recombinant strains AM010 and AM011 producing 0.7 and 0.8 mM, respectively (Figure 6.14). This result likely indicated the need to overexpress *adhA* (encoding alcohol dehydrogenase), which catalyse the final step of isobutyraldehyde to isobutanol, to see whether isobutanol titres could be improved.

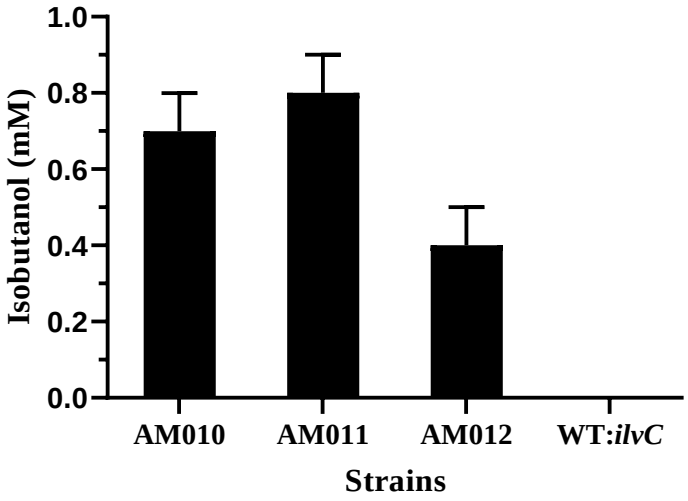


Figure 6.14: Concentration of isobutanol (mM) produced by the recombinant strains of *P. thermoglucosidasius::ilvC* transformed with P_{ldh}*_kivD_alsS_ilvD* operon. AM010, AM011, and AM012 denotes strains expressing genome-integrated *ilvC* at *pyrE* locus along with plasmid-based expression of P_{ldh}*_kivD* (LLKF_1386) *_alsS*$_{bs}$*_ilvD*, P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bc}$*_ilvD*, and P_{ldh}*_kivD* (LLKF_1386) *_alsS*$_{td}$*_ilvD* respectively. WT:*ilvC* denote wild-type *P. thermoglucosidasius::ilvC*. In case of the *alsS* gene designations, the subscripts indicate the source of the gene, viz., bs, *B. subtilis*, bc, *B. coagulans*, td, *T. daqus*. In three biological replicates, the recombinant strains and wild-type's isobutanol profiles were characterised by HPLC after 48 h of fermentation using 10 ml of M-ASYE with 20 $g \cdot l^{-1}$ glucose and 10 $g \cdot l^{-1}$ yeast extract +12.5 $\mu g \cdot ml^{-1}$ Kan in 50 ml falcon tubes at 50 ˚C, 250 rpm. Results are shown as mean ± SD of three biological replicates.

6.2.5 Effect of overexpression of the alcohol dehydrogenase (*adhA*) gene

To test whether increasing alcohol dehydrogenase concentration could improve the isobutanol titres, the encoding gene (*adhA*) was cloned and overexpressed, essentially by inserting *adhA* in the same operon downstream of *kivD, alsS* and *ilvD* with genome-integrated *ilvC*.

6.2.5.1 Construction of pMTL-AM13, pMTL-AM14, and pMTL-AM15 plasmids

The P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bs}$*_ilvD_adhA*, P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bc}$ *_ilvD_adhA*, and P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{td}$*_ ilvD_adhA*, operons were constructed using NEBuilder-HIFI DNA Assembly method as described in Methods section 2.9.11. Oligonucleotide primers were designed using NEBuilder Assembly Tool (https://nebuilder.neb.com/#!/), and the detailed sequences are presented in Table 2.1. The native *ilvD g*ene was PCR amplified from *P. thermoglucosidasius* genomic DNA using pairs of ilvD_F_alsS1/ilvD_R_adhA, ilvD_F_alsS2/ilvD_R_adhA, and ilvD_F_alsS3 /ilvD_R_adhA primers, respectively. Similarly, the *adhA* gene from *P. thermoglucosidasius* was PCR amplified using adhA_F_ilvD and adhA_R_Cpa. Plasmids pMTL-AM7 (P*ldh_kivD* (LLKF_1386)*_alsS*$_{bs}$), pMTL-AM8 (P*ldh_kivD* (LLKF_1386)*_alsS*$_{bc}$), and pMTL-AM9 (P*ldh_kivD* (LLKF_1386)*_alsS*$_{td}$) respectively, were linearized by restriction digestion using NheI enzyme. The linearized vectors and the amplified *ilvD* and *adhA* gene fragments were assembled together using NEBuilder HIFI DNA Assembly Master Mix as described in Methods section 2.9.11, yielding plasmids pMTL-AM13 [P*ldh_kivD*(LLKF_1386) *_alsS*$_{bs}$*_ilvD_adhA*], pMTL-AM14 [P*ldh_kivD* (LLKF_1386)*_alsS*$_{bc}$*_ilvD_adhA*), and pMTL-AM15 [P*ldh_kivD* (LLKF_1386) *_alsS*$_{td}$*_ilvD_adhA*], respectively (Figure 6.15). Complete plasmid maps are presented in appendix X.

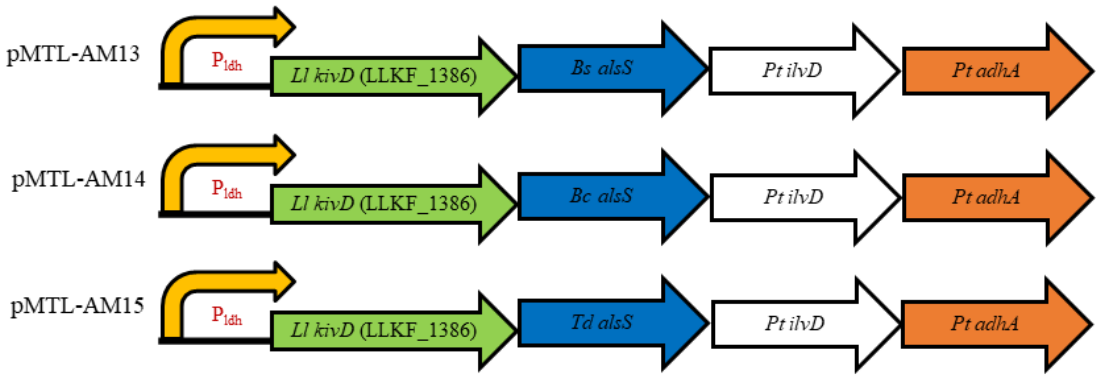


Figure 6.15: Schematic of P_{ldh}_*kivD_alsS_ilvD_adhA* operons in pMTL-AM13, pMTL-AM14 and pMTL-AM15 plasmids. P_{ldh}: lactate dehydrogenase promoter, *Ll*: *L. lactis*, *kivD*: ketoisovalerate decarboxylase gene, *Bs*: *B. subtilis*, *Bc*: *B. coagulans*, *Td*: *T. daqus*, *alsS*: acetolactate synthase gene, *Pt*: *P. thermoglucosidasius*, *ilvD*: dihydroxy acid dehydratase gene, *adhA*: alcohol dehydrogenase gene.

Four transformed *E. coli* Top10 colonies were grown overnight, and then plasmids pMTL-AM13, pMTL-AM14, and pMTL-AM15 were extracted to screen for the correct constructs. Purified pMTL-AM13, pMTL-AM14, and pMTL-AM15 plasmids were screened by PCR for the presence of *ilvD_adhA* gene fragments (Figure 6.16) using pairs of ilvD_F_alsS1/adhA_R_Cpa, ilvD_F_alsS2/adhA_R_Cpa, and ilvD_F_alsS3 /adhA_R_Cpa primers, respectively (detailed primer sequence Table 2.1). Furthermore, Sanger sequencing confirmed the presence of the expected nucleotide sequences in the plasmids.

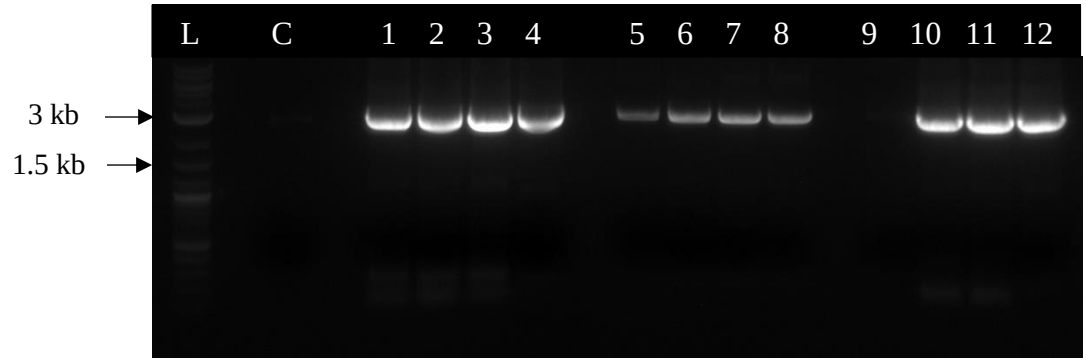


Figure 6.16: PCR screening of *ilvD_adhA* gene fragments from pMTL-AM13, pMTL-AM14, and pMTL-AM15 plasmids. L = DNA Ladder (0.1-10 kb), lane C = pMTL-AM7 plasmid as negative control, lanes 1-4=pMTL-AM13 [P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bs}$*_ilvD_adhA*], lanes 5-8 = pMTL-AM14 [P_{ldh}*_kivD* (LLKF_1386) *_alsS*$_{bc}$*_ilvD_adhA*), and lanes 9-12=pMTL-AM15 [P_{ldh}*_kivD* (LLKF_1386) *_alsS*$_{td}$*_ilvD_adhA*]. In case of the *alsS* gene designations, the subscripts indicate the source of the gene, viz., bs, *B. subtilis*, bc, *B. coagulans*, td, *T. daqus*. Expected size of *ilvD_adhA* gene fragments ~ 2.7 kb.

6.2.5.2 Construction of recombinant *P. thermoglucosidasius* AM013, AM014, and AM015 strains

Plasmids pMTL-AM13, pMTL-AM14, and pMTL-AM15 containing P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bs}$*_ilvD_adhA*, P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bc}$*_ilvD_adhA*, and P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{td}$*_ilvD_adhA* operons respectively, were transformed into electro-competent cells of *P. thermoglucosidasius::ilvC* as described in Methods section 2.10.4. Colony PCR was used to confirm the presence of the P_{ldh}*_kivD* (LLKF_1386) *_alsS_ilvD_adhA* operons in the transformed colonies using plasmid-based pLDH_Seq and RepB_R1 oligonucleotide primers (detailed primer sequence in

Table 2.1), generating a ~ 6.7 kb product for the colonies screened (three for each strain) (Figure 6.17). Furthermore, the PCR products were purified, and Sanger sequenced to confirm the presence of the correct nucleotide sequences. The strains generated were designated as *P. thermoglucosidasius* AM013, AM014, and AM015 strains expressing genome-integrated *ilvC* at *pyrE* locus along with plasmid-based expression of P_{ldh}*_kivD* (LLKF_1386)*_*$alsS_{bs}$*_ilvD_adhA*, P_{ldh}*_kivD* (LLKF_1386)*_*$alsS_{bc}$*_ilvD_adhA*, and P_{ldh}*_kivD* (LLKF_1386)*_*$alsS_{td}$*_ilvD_adhA*, respectively.

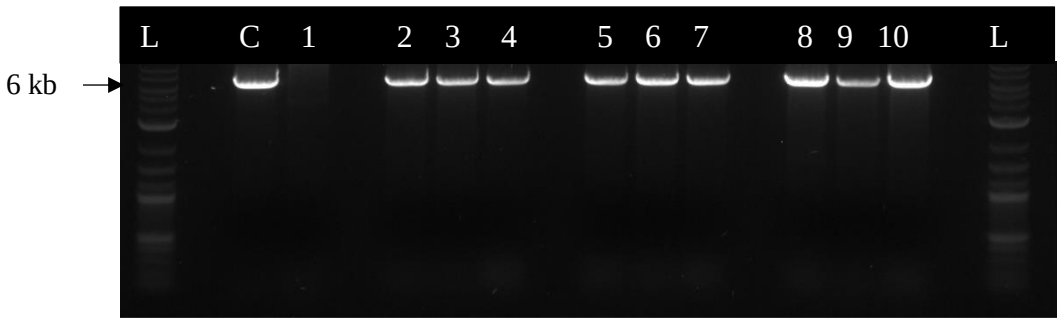


Figure 6.17: Colony-PCR screening of P_{ldh}*_kivD_alsS_ilvD_adhA* operons in recombinant *P. thermoglucosidasius*. L = DNA Ladder (0.1-10 kb), C= pMTL-AM13 vector as positive control, lane 1 = *P. thermoglucosidasius::ilvC* as negative control, lanes 2-4 = AM013 - P_{ldh}*_kivD* (LLKF_1386) *_*$alsS_{bs}$*_ilvD_adhA*, lanes 5-7 = AM014 -P_{ldh}*_kivD* (LLKF_1386) *_*$alsS_{bc}$*_ilvD_adhA*, and lanes 8-10 = AM015 -P_{ldh}*_kivD* (LLKF_1386) *_*$alsS_{td}$*_ilvD_adhA*. In case of the *alsS* gene designations, the subscripts indicate the source of the gene, viz., bs, *B. subtilis*, bc, *B. coagulans*, td, *T. daqus*. Expected size of P_{ldh}*_kivD_alsS_ilvD_adhA* operons ~ 6.7 kb.

6.2.5.3 Characterization of recombinant *P. thermoglucosidasius* AM013, AM014, and AM015 strains

Isobutanol production by the recombinant *P. thermoglucosidasius* AM013, AM014, and AM015 strains, as well as the *P. thermoglucosidasius::ilvC* control strain, was investigated as described previously in sections 6.2.2.3 and 6.2.3.2. Amazingly, overexpression of *adhA* alongside *kivD, alsS, ilvD* and *ilvC* only improved isobutanol production in recombinant *P. thermoglucosidasius* AM014 strain with 1 mM isobutanol produced and not in AM013 and AM015 strains (Figure 6.18).

The combined results observed in this study indicate that the system produced isobutanol but at low titres, possibly due to the expression of *ilvC* as a single copy on the genome, or the P_{ldh} promoter is not sufficient to drive the expression of the whole operon.

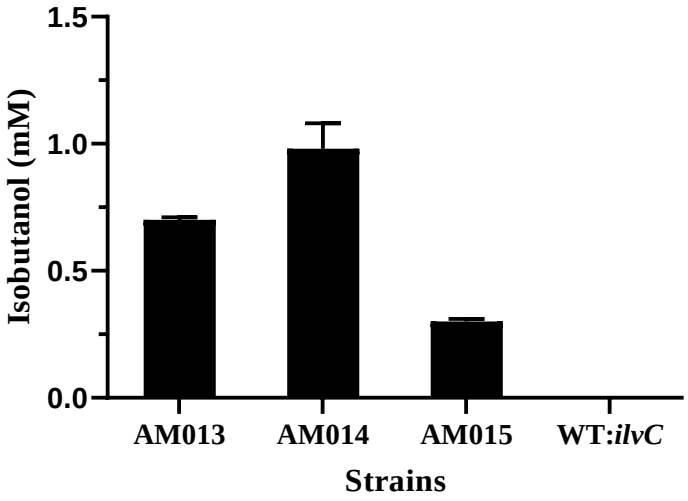


Figure 6.18: Concentration of isobutanol (mM) produced by the recombinant strains of *P. thermoglucosidasius::ilvC* transformed with P_{ldh}*_kivD_alsS_ilvD_adhA* operon. AM013, AM014, and AM015 denotes strains expressing genome-integrated *ilvC* at *pyrE* locus along with plasmid-based expression of P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bs}$*_ilvD_adhA*, P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{bc}$*_ilvD_adhA*, and P_{ldh}*_kivD* (LLKF_1386)*_alsS*$_{td}$*_ilvD_adhA* respectively. WT:*ilvC* denote wild-type *P. thermoglucosidasius::ilvC*. In case of the *alsS* gene designations, the subscripts indicate the source of the gene, viz., bs, *B. subtilis*, bc, *B. coagulans*, td, *T. daqus*. In three biological replicates, the recombinant strains and wild-type's isobutanol profiles were characterised by HPLC after 48 h of fermentation using 10 ml of M-ASYE with 20 $g \cdot l^{-1}$ glucose and 10 $g \cdot l^{-1}$ yeast extract +12.5 $\mu g \cdot ml^{-1}$ Kan in 50 ml falcon tubes at 50 ˚C, 250 rpm. Results are shown as mean ± SD of three biological replicates.

6.3 Discussion

The alcohol isobutanol is proposed as a superior fuel to ethanol because of its relatively lower hygroscopicity, higher energy density (Lee *et al.,* 2012) and its compatibility with current infrastructure (Smith *et al.,* 2010). In the present study, engineering of a recombinant strain of *P. thermoglucosidasius* NCIMB 11955 to produce isobutanol from glucose was demonstrated.

The first step was the characterisation of the native/wild-type KIVD from *L. lactis* and its several variants for isobutanol production in a medium supplemented with the precursor, 2-ketoisovalerate. Interestingly, KIVD (LLKF_1386) was found to be the most active at 50 °C. An earlier study in *P. thermoglucosidasius* supports this observation, where the mesophilic *L. lactis* KIVD (LLKF_1386) was active at 50 °C and was utilised for isobutanol biosynthesis with a titre of 3.3 $g \cdot l^{-1}$ (Lin *et al.,* 2014). Furthermore, the wild-type *P. thermoglucosidasius* and the recombinant strains expressing the other KIVD variants produce similar isobutanol titres (0.4 mM). Acetolactate synthase has been shown to convert 2-ketoisovalerate to isobutyraldehyde (Atsumi *et al.,* 2009a). Thus, the isobutanol produced by the wild-type and the other recombinant strains of *P. thermoglucosidasius* might be due to acetolactate synthase present in *P. thermoglucosidasius,* which has been reported to have KIVD activity (Lin *et al.,* 2014).

With the best combination of enzymes (*L. lactis* KIVD, *B. coagulans* ALS, *P. thermoglucosidasius* KARI, DHAD and ADH) under the control of P_{ldh} promotor of the *ldhA* gene, a titre of 1 mM isobutanol from glucose at 50 °C was achieved after two days. Although a low concentration of isobutanol was produced, the system does demonstrate the production of isobutanol from glucose. This low isobutanol titre could

possibly be due to the fact that the single genomic copy of *ilvC* results in inadequate expression of this gene or that the P_{ldh} promoter is insufficient to promote expression of the entire operon. Future work aimed at increasing the isobutanol titre would need to investigate these possibilities before looking at the potential of *P. thermoglucosidasius* as a CBP organism, where both cellulose hydrolysis and the fermentation of the sugars to isobutanol within a single process could be explored. Increased isobutanol production titres may be achieved by overexpressing the *ilvC* on a multicopy plasmid (Lee *et al.,* 2012; Lin *et al.,* 2014; Lin *et al.,* 2015) or overexpressing the operon genes under the control of different promoters of relatively higher strength (Lin *et al.,* 2015). Furthermore, inactivating competing pathways potentially consuming isobutanol precursors such as phosphoenolpyruvate, pyruvate, and 2-ketoisovalerate might improve isobutanol production (Smith *et al.,* 2010). For example, the K_m of ALS for pyruvate is 13.6 mM (Atsumi *et al.,* 2009b), whereas lactate dehydrogenase (LDH) has a lower K_m 7.3 mM (Dietrich *et al.,* 2009), indicating increased pyruvate consumption for LDH to produce lactate as against ALS for isobutanol production. Therefore, inactivation of the LDH can improve isobutanol production.

Previous studies in *E. coli* by Atsumi *et al.* (2008) and other mesophiles (Atsumi *et al.,* 2009a; Smith *et al.,* 2010; Higashide *et al.,* 2011; Lee *et al.,* 2012) have shown that the production of isobutanol can be metabolically engineered by redirecting the keto acid intermediate in the valine biosynthesis pathway toward isobutanol production. Two important non-native enzymes, KIVD from *L. lactis* and ALS from *B. subtilis*, were employed to demonstrate this heterologous pathway. Furthermore, these same key enzymes have been used for isobutanol production in thermophilic bacteria, *P. thermoglucosidasius* and *C. thermocellum* (Lin *et al.,* 2014; Lin *et al.,* 2015).

Although the *P. thermoglucosidasius* organism has been engineered for isobutanol production (3.3 $g \cdot l^{-1}$) at elevated temperatures using glucose (Lin *et al.,* 2014), no cellulosic isobutanol production was reported using CBP. *P. thermoglucosidasius* is a promising CBP host because it can utilise various sugars derived from lignocellulose and grow at high temperatures. Furthermore, a recent study in *P. thermoglucosidasius* demonstrates cellulose hydrolysis and simultaneous fermentation of pretreated wheat straw for ethanol production (Bashir *et al.,* 2019). More recently, *P. thermoglucosidasius* was engineered for the heterologous production of terpenes in a consolidated bioprocess using waste bread (Styles *et al.,* 2021). Moreover, previous studies have also shown isobutanol production from cellulose, the mesophilic *C. cellulolyticum* (Higashide *et al.,* 2011) and the thermophilic *C. thermocellum* (Lin *et al.,* 2015). Thus, further highlighting the prospect of *P. thermoglucosidasius* as a CBP organism for isobutanol production. The study described here represents a small step in this direction, but much work must be done before meaning levels of isobutanol are achieved.

6.4 Conclusions

In this study, the construction of an isobutanol production system from glucose was attempted through the metabolic engineering of *P. thermoglucosidasius* NCIMB 11955. The titre of isobutanol achieved was, however, relatively low. Still, the prospect of operon modification by overexpressing the *ilvC* gene on a multicopy plasmid and driving the expression of the operon genes efficiently under the control of different promoters can potentially improve isobutanol titres. This chapter is merely a small step of many to allow the potential of *P. thermoglucosidasius* NCIMB 11955 as a CBP organism for isobutanol production to be achieved.

CHAPTER 7

7.0 GENERAL DISCUSSION

7.1 Overview

The continued reliance on fossil resources to produce fuels and chemicals is not sustainable. In addition to being finite resources, their extraction, processing and usage are associated with environmental damage and pollution, principally the release of greenhouse gases, the leading cause of climate change (Owusu and Asumadu-Sarkodie, 2016). Thus, alternative, more sustainable, and environmentally friendly processes are required. One of the most promising and attractive options is the use of microbial fermentative processes to produce chemicals and fuels from renewable feedstocks, mainly lignocellulosic biomass (Sheng *et al.,* 2017). Bioethanol, the predominant biofuel produced and used today, is produced by yeast (Mattanovich *et al.,* 2014). However, there is a growing interest in developing processes for alternative biofuels that are superior to ethanol, such as isobutanol, which has higher energy density (Smith *et al.,* 2010), reduced vapour pressure and lower hygroscopicity (Lee *et al.,* 2012). Furthermore, other microorganisms are being investigated to produce biofuels, including thermophilic bacteria belonging to the genus *Parageobacillus*.

The thermophile *P. thermoglucosidasius* has significant attractions as a chassis for producing fuels and chemicals from cheap, renewable lignocellulosic feedstock (Cripps *et al.,* 2009; Xio *et al.,* 2012; Lin *et al.,* 2014; Raita *et al.,* 2016; Zhou *et al.,* 2020; Styles *et al.,* 2021). Its growth at elevated temperatures improves feed conversion rate and reduces fermentation cooling costs between successive fermentation cycles, rendering the process more effective (Zeldes *et al.,* 2015). Elevated temperatures also confer desirable properties on the growth medium, such as reduced viscosity, reduced energy requirements for mixing, increased diffusion rates, and substrate solubility (Zeldes *et al.,* 2015). High temperatures additionally reduce the risk of contamination by other microorganisms (Taylor *et al.,* 2009). *P. thermoglucosidasius* is also

characterised by its ability to metabolise a wide range of pentose and hexose sugar monomers (xylose, glucose, cellobiose, and mixtures of glucose, xylose, and arabinose), typical of those derived from lignocellulosic biomass (Hussein *et al.,* 2015). Therefore, the work presented in this thesis focused on the metabolic engineering of *P. thermoglucosidasius* NCIMB 11955 for the production of biofuels (ethanol and isobutanol) and the platform chemical 3-HP.

7.2 General discussion and future prospects

7.2.1 Characterisation of pMTL60000 modular vector series

This study focused on adding to and characterising the pMTL60000 modular vector series to add to the vectors available for strain engineering in *Parageobacillus* and *Geobacillus* species, using the modular plasmid format previously used for *Clostridium* (pMTL80000) (Heap *et al.,* 2009) and *C. necator* (pMTL70000) (Ehsaan *et al.,* 2021). As described in chapter 3, this initially involved the construction of a pMTL66331 modular vector based on the *repBST1* replicon. After that, the characterisation of the new vector alongside the others based on other available replicons (Sheng *et al.,* 2017; Spencer, 2018) was successfully achieved. The transfer efficiencies in *P. thermoglucosidasius* of all the modular plasmids were good and supported previous studies (Reeve *et al.,* 2016). Plasmid segregational stability is often an essential determinant of the practicality of genetic tools; thus, the availability of a greater selection of replicons exhibiting varying stabilities, as shown in this work, would benefit various applications. Plasmids with high segregational stability will be more suited for gene expression purposes (Bartosiak-Jentys *et al.,* 2013, Lin *et al.,* 2014, Reeve *et al.,* 2016). On this basis, plasmids based on the pUB110.1, pNCI001, pNCI002, and pGEOTH02 replicons may represent the most appropriate to base an expression vector

on. In contrast, low-stability plasmids are desired for gene KO and KI, where plasmid curing is often a time-limiting step in genetic modification (Cripps *et al.*, 2009, Sheng *et al.*, 2017, Lau *et al.*, 2021). Therefore, plasmids based on the pUB110.2, pUB110.3 and pBST1 replicons could prove useful in applications where unstable plasmids are required.

After fully characterising the available replicon options, the nucleotide sequences of system modules were made available in the vector generator facility at the web www.plasmidvectors.com to allow users to download the desired plasmid sequence and order them.

The pMTL60000 modular vector series described here adds to the vectors available (Taylor *et al.*, 2008; Bartosiak-Jentys *et al.*, 2013; Reeve *et al.*, 2016; Sheng *et al.*, 2017) to further the strain engineering of *P. thermoglucosidasius* as a chassis for producing chemicals and fuels. These vectors could potentially be applied to advance strain engineering for various applications in many other related and industrially useful thermophiles and *Bacillus* species.

7.2.2 Implementation of theophylline-responsive CRISPR/Cas9 gene-editing tool in *P. thermoglucosidasius* NCIMB 11955

The engineering of microbial strains to produce value-added products is reliant on the availability of adequate and effective gene-editing tools that could be used to achieve the desired changes to metabolic pathways through gene knock-out and knock-in (Sheng *et al.*, 2017). Previous genetic tools developed for engineering *P. thermoglucosidasius* were limited, and this restricted the potential products that could be produced with this chassis.

The discovery of CRISPR/Cas9 systems in providing adaptive bacterial immunity (Barrangou *et al.,* 2007) and their subsequent application in genome editing (Jinek *et al.,* 2012; Xu *et al.,* 2015; Müller *et al.,* 2016; Bayer *et al.,* 2022) have led to a significant breakthrough in synthetic biology and metabolic engineering. This chapter detailed the successful implementation of an inducible genome-editing tool based on theophylline-responsive CRISPR/Cas9 to facilitate efficient knock-out and knock-in in *P. thermoglucosidasius* NCIMB 11955. The successful implementation of the tool described in this study is the first application of a genome-editing tool based on a theophylline-responsive CRISPR/Cas9 in *P. thermoglucosidasius*. This novel and efficient theophylline-responsive CRISPR/Cas9 system was successfully implemented for genome engineering in several species of Clostridia, including the pathogens *C. botulinum* and *C. difficile*, the solventogenic *C. pasteurianum* and the non-pathogen *C. sporogenes* (Cañadas *et al.,* 2019), and *C. autoethanogenum* (Seys *et al.,* 2020). Moreover, the system has been described in the model cyanobacterium *Synechocystis* sp. PCC 6803 for single and multiple genome editing (Cengic *et al.,* 2022). The exemplified theophylline-responsive CRISPR/Cas9 tool in this study addresses the constraints of the limited genome editing tools of *P. thermoglucosidasius* (Cripps *et al.,* 2009; Sheng *et al.,* 2017; Lau *et al.,* 2021).

Here, successful genome-editing was demonstrated by engineering multiple, independent ethanol producing strains equivalent to the industrial bioethanol production strain TM242 (Cripps *et al.,* 2009) from its progenitor strain TM89 (Δ*ldh*) to understand why the strain (TM242) produces ethanol from glucose with efficiencies that approach the theoretical maximum. This was achieved by the stepwise deletion of the *pflB* gene (encoding pyruvate formate lyase) and the upregulation of the *pdhA* gene (encoding pyruvate dehydrogenase) by placing it under the control of a strong P_{ldh} promoter from

G. stearothermophilus (Bartosiak-Jentys *et al.*, 2012). Significant differences in the fermentation profiles of the strains generated were noted. In comparison, the profile of the genetically equivalent strains AM242_2_5 and AM242_4_4 were comparable to strain TM242 and showed no significant differences. This is in contrast to a previous creation of an equivalent strain LS242 by Sheng *et al*. (2017) to demonstrate the utility of the developed ACE method, where unintended mutations arose and dramatically affected phenotype leading to the production of ethanol not matching that of strain TM242 (Cripps *et al.*, 2009). Therefore, the observation made in this study emphasises the need to routinely generate multiple, independent engineered strains combined with whole-genome sequencing to improve the screening and selection of strains. Furthermore, through the course of this study, it was demonstrated that the TM242 strain (Cripps *et al.*, 2009) produces high ethanol close to the theoretical maximum due to the SNPs/Indels it acquired during its generation.

Ethanol yields of greater than 90% of the theoretical maximum was achieved by the *P. thermoglucosidasius* AM242 strain from barley straw hydrolysate. Therefore, it could provide the basis for the development of an effective and economical process for bioethanol production from lignocellulosic feedstocks.

The described method represents a significant improvement and adds to the few existing genome-editing tools available (Cripps *et al.*, 2009; Zhou *et al.*, 2016; Sheng *et al.*, 2017; Lau *et al.*, 2021) for the metabolic engineering of *P. thermoglucosidasius* to produce chemicals and fuels. In addition, the tool described could speed up the development of future applications, such as the production of more complex compounds or engineering strains for other purposes. Furthermore, the theophylline-responsive CRISPR/Cas9 genome-editing tool described in this study could potentially be ported

to other thermophilic bacteria for specific and efficient metabolic engineering to generate products with biotechnological applications.

7.2.3 Engineering 3-hydroxypropionic acid production in *P. thermoglucosidasius* NCIMB 11955

7.2.3.1 Development of recombinant *P. thermoglucosidasius* NCIMB 11955 strains for 3-HP production

The development of bio-based technologies for the production of the platform chemical, 3-HP from cheaper and renewable biomass feedstocks has attracted much attention in the last two decades; when in 2004, the US Department of Energy listed 3-HP among the top 12 high value-added chemicals that can be derived from biomass (Werpy and Petersen, 2004). In the present study, various recombinant *P. thermoglucosidasius* strains were generated that produce 3-HP via the malonyl-CoA pathway from glucose. This was achieved by heterologously overexpressing five different MCR–MSR operons on autonomous plasmids and using the native ACC present in the wild type *P. thermoglucosidasius* NCIMB 11955.

The recombinant *P. thermoglucosidasius* strains (KT1, KT2, KT3, KT4, and KT5) carried different heterologous combinations of MCR–MSR operons, but with the same native ACC present in the genome, produced 3-HP in the range of 0.4–1.4 mM. Efforts to improve the production of 3-HP by blocking lactate formation by deleting the *ldhA* gene were unsuccessful. The low level of production may be due to the limited amount of malonyl-CoA. It has been reported that the level of intracellular malonyl-CoA, which is also involved in fatty acid biosynthesis, is the limiting factor in the production of 3-HP through the MCR pathway (Rathnasingh *et al*., 2012). Thus, in future studies, the upregulation of the ACC gene by placing it under the control of a strong promoter such

as the P_{ldhA} promoter from *G. stearothermophilus* (Bartosiak-Jentys *et al.*, 2012) could increase the intracellular malonyl-CoA concentration, which is the immediate substrate for MCR. To date, 3-HP production has only been attempted via the malonyl-CoA pathway in *P. thermoglucosidasius* NCIMB 11955 by heterologous expression of the ACC under the control of P_{ldh} promoter in the genome, and different MCR–MSR combinations as operons on autonomous plasmids (Bashir, 2018), with the recombinant strains producing 3-HP ranging from 3.0–3.8 mM from glucose.

3-HP production has been reported using engineered organisms such as *E. coli* (Rathnasingh *et al.,* 2012; Lim *et al.,* 2016), *K. pneumoniae* (Jiang *et al.*, 2018), *B. subtilis* (Kalantari *et al.*, 2017), *M. extorquens* (Yang *et al.*, 2017), and *C. necator* H16 (Salinas *et al.,* 2022). Moreover, 3-HP has been produced using an engineered hyperthermophile, *Pyrococcus furiosus* from hydrogen gas and carbon dioxide (Keller *et al.*, 2013).

To further increase the titre and productivity of 3-HP, blocking the acetate production pathway by deleting the *ack* gene (encoding acetate kinase) could increase the pool of intracellular acetyl-CoA required for the synthesis of malonyl-CoA. Additionally, the production of acetyl-CoA from pyruvate is catalysed by the *pdhA* gene (encoding pyruvate dehydrogenase); thus, upregulation of the *pdhA* gene could enhance carbon flux towards pyruvate and subsequently acetyl-CoA and, in turn, improve 3-HP production. Furthermore, increasing the concentrations of malonyl-CoA by overexpressing the native ACC on multicopy plasmid, and NADPH by expressing nicotinamide nucleotide transhydrogenase (PntAB) could enhance the production of 3-HP (Rathnasingh *et al.,* 2012).

7.2.3.2 Investigating 3-HP catabolism in *P. thermoglucosidasius* NCIMB 11955

Product utilisation is an undesirable characteristic in any biological process. Thus, it is essential to ensure that the generated compounds are not re-assimilated by the producing organisms (Arenas-Lopez *et al.,* 2019). Several microorganisms, including *P. denitrificans*, *M. extorquens*, *C. necator* H16, and *R. sphaeroides,* can degrade 3-HP (Zhou *et al.,* 2014; Yang *et al.,* 2017; Arenas-Lopez *et al.,* 2019; Schneider *et al.,* 2012). However, 3-HP degradation in *P. thermoglucosidasius* has not been described to date. For this reason, the native catabolism of 3-HP in *P. thermoglucosidasius* NCIMB 11955 was investigated and manipulated in this study to generate a strain incapable of 3-HP degradation to be used as a chassis for its production. As described in chapter 5, the wild type *P. thermoglucosidasius* NCIMB 11955 can utilise 3-HP as the sole source of carbon for growth, which requires abolishing.

Two principal pathways for the assimilation of 3-HP have been proposed in different organisms, a reductive pathway (Schneider *et al.,* 2012) and an oxidative pathway (Arenas-Lopez *et al.,* 2019). Putative genes for the oxidative route, namely two (methyl) malonate semialdehyde dehydrogenases (*mmsA1* and *mmsA2*) and a 3-hydroxyisobutyrate dehydrogenase (*hbdH*), are present in *P. thermoglucosidasius* NCIMB 11955. Whilst deletion of the *mmsA2* gene could only delay growth and 3-HP consumption up to 84 h, the combined inactivation of the two *mmsA* genes abolished growth and 3-HP consumption for the duration employed in this study. Furthermore, while the aim of this study was achieved by obtaining a *P. thermoglucosidasius* NCIMB 11955 strain incapable of 3-HP assimilation, the deletion of the *hbdH* gene for mutant strain construction and subsequent testing to determine its effect in 3-HP catabolism was not possible due to time constraints. However, its inactivation is a priority for further engineering and future work to be undertaken.

Two different (methyl) malonate semialdehyde dehydrogenases contribute to 3-HP degradation in *P. thermoglucosidasius* NCIMB 11955. The double Δ*mmsA2*Δ*mmsA1* mutant strain represents an ideal chassis for future production of 3-HP in *P. thermoglucosidasius* NCIMB 11955.

7.2.4 Towards metabolic engineering of *P. thermoglucosidasius* NCIMB 11955 for isobutanol production

The use of lignocellulosic biomass instead of sugar as the substrate for biofuel and bio-based chemicals production could lower the cost of production, provide the quantity required to make a remarkable impact, prevent the food versus fuel impasse, and improve carbon neutral and energy balances (Lynd *et al.,* 2005; Lin *et al.,* 2015). However, the use of lignocellulose is limited due to biomass resistance to degradation. A potential solution is CBP, where cellulose hydrolysis and simultaneous fermentation occur without the addition of expensive cellulases. *P. thermoglucosidasius* is a promising CBP host because it can utilise various sugars derived from lignocellulose and grow at high temperatures (Hussein *et al.,* 2015). Although this organism has been engineered for isobutanol production from glucose at elevated temperatures (Lin *et al.,* 2014), no cellulosic isobutanol production was reported using CBP. A recent study in *P. thermoglucosidasius* demonstrates cellulose hydrolysis and simultaneous fermentation of pretreated wheat straw for ethanol production (Bashir *et al.,* 2019). More recently, *P. thermoglucosidasius* was engineered for heterologous production of terpenes in a consolidated bioprocess using waste bread (Styles *et al.,* 2021). Moreover, previous studies have shown isobutanol production from cellulose, the mesophilic *C. cellulolyticum* (Higashide *et al.,* 2011) and the thermophilic *C. thermocellum* (Lin *et al.,* 2015).

This work aimed to engineer *P. thermoglucosidasius* for cellulosic isobutanol production via CBP approach. This initially involved engineering a recombinant strain of *P. thermoglucosidasius* NCIMB 11955 to produce isobutanol from glucose as a proof of concept. With the best combination of enzymes (*L. lactis* KIVD, *B. coagulans* ALS, *P. thermoglucosidasius* KARI, DHAD and ADH) under the control of the P_{ldh} promoter, a titre of 1 mM isobutanol from glucose at 50 °C after 48 h was achieved. Although a low concentration of isobutanol was produced, the system does demonstrate the production of isobutanol from glucose. The low isobutanol titre observed could be due to *ilvC* being present as a single copy gene or the fact that the P_{ldh} promoter is insufficient to drive the expression of the entire operon. Therefore, further work is needed to increase the isobutanol titre before looking at the potential of *P. thermoglucosidasius* as a CBP organism, where both cellulose hydrolysis and the fermentation of the sugars to isobutanol within a single process could be explored.

In future studies, increased isobutanol production titres may be achieved by overexpressing the *ilvC* on a multicopy plasmid (Lee *et al.,* 2012; Lin *et al.,* 2014; Lin *et al.,* 2015) or by overexpressing the operon genes under the control of different promoters of relatively higher strength (Lin *et al.*, 2015). In addition, high titres and yields of isobutanol could be achieved by knocking out competing pathways. This study highlighted the prospect of *P. thermoglucosidasius* as a CBP organism for isobutanol production. However, a lot of work is needed to achieve this and therefore, this study is only a small step in that direction.

CHAPTER 8

8.0 CONCLUSIONS

8.1 Conclusions

This study has achieved the aim of the project by engineering *P. thermoglucosidasius* NCIMB 11955 to produce sustainable fuels (ethanol and isobutanol) and chemicals (3-HP). The first part of the study added to and characterised the pMTL60000 modular vector series exhibiting varying properties to allow their use for various applications. The plasmids and their sequences are available at www.plasmidvectors.com.

The study presented in this thesis has contributed to expanding the tools available for the metabolic engineering of *P. thermoglucosidasius* NCIMB 11955 by successfully applying a theophylline-responsive CRISPR/Cas9 genome-editing tool to produce strains capable of producing ethanol efficiently as a primary fermentation product from barley straw hydrolysate. Furthermore, by systematically re-engineering *P. thermoglucosidasius* TM242 (Cripps *et al.*, 2009) using this tool, it was demonstrated that the TM242 strain produces ethanol from glucose with efficiencies that approach the theoretical maximum for a fermentative organism due to the SNPs/Indels it acquired during its generation.

The study also successfully demonstrated the engineering of *P. thermoglucosidasius* NCIMB 11955 to make the platform chemical 3-HP via the malonyl-CoA pathway using the native acetyl-CoA carboxylase and heterologous malonyl-CoA reductase and malonate semialdehyde reductase. Additionally, genes responsible for 3-HP degradation were identified, and a strain incapable of 3-HP assimilation was generated.

Another part of the study focused on isobutanol production with the ultimate aim of using a CBP approach for its production. While attempts and some progress were made in engineering *P. thermoglucosidasius* NCIMB 11955 to produce isobutanol from glucose, the significant titres needed to pave the way for isobutanol production through

CBP still needed to be achieved. Hence, further research will be required to achieve this goal.

The present study further demonstrates the potential of producing platform chemicals and biofuels from lignocellulosic feedstocks using *P. thermoglucosidasius* NCIMB 11955 as a chassis.

BIBLIOGRAPHY

Abas, N., Kalair, A. and Khan, N. (2015). Review of fossil fuels and future energy technologies. *Futures*, **69:** 31–49.

Alber, B.E. and Fuchs, G. (2002). Propionyl-coenzyme A synthase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO_2 fixation. *Journal of Biological Chemistry*, **277:** 12137-43.

Al-Hasan, M.I. and Al-Momany, M. (2008). The Effect of Isobutanol-Diesel Blends on Engine Performance. *Transport*, **23(4):** 306-310.

Aliyu, H., Lebre, P., Blom, J., Cowan, D. and De Maayer, P. (2016). Phylogenomic re-assessment of the thermophilic genus *Geobacillus*. *Systematic and Applied Microbiol*ogy, **39:** 527−533.

Arenas-López, C., Locker, J., Orol, D., Walter, F., Busche, T., Kalinowski, J., Minton, N.P., Kovács, K. and Winzer, K. (2019). The genetic basis of 3-hydroxypropanoate metabolism in *Cupriavidus necator* H16. *Biotechnology for Biofuels*, **12:** 150.

Argyros, D.A., Tripathi, S.A., Barrett, T.F., Rogers, S.R., Feinberg, L.F., Olson, D.G., Foden, J.M., Miller, B.B., Lynd, L.R. and Hogsett, D.A. (2011). High ethanol titers from cellulose using metabolically engineered thermophilic, anaerobic microbes. *Applied and Environmental Microbiology*, **77:** 8288–8294.

Asao, M. and Alber, B.E. (2013). Acrylyl-coenzyme A reductase, an enzyme involved in the assimilation of 3-hydroxypropionate by *Rhodobacter sphaeroides*. *Journal of Bacteriology*, **195:** 4716-25.

Ash, C., Farrow, J.A.E., Wallbanks, S. and Collins, M.D. (1991). Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. *Letters in Applied Microbiology*, **13:** 202-206.

Ashok, S., Mohan Raj, S., Ko, Y., Sankaranarayanan, M., Zhou, S., Kumar, V. and Park, S. (2013). Effect of *puuC* overexpression and nitrate addition on glycerol metabolism and anaerobic 3-hydroxypropionic acid production in recombinant *Klebsiella pneumoniae* Δ*glpK*Δ*dhaT*. *Metabolic Engineering*, **15:** 10-24.

Asumadu-Sarkodie, S. and Owusu, P.A. (2016). Feasibility of biomass heating system in Middle East Technical University, Northern Cyprus campus. *Cogent Engineering*, **3(1):** 1-15.

Atsumi, S., Hanai, T. and Liao, J.C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*, **451:** 86–89.

Atsumi, S., Higashide, W. and Liao, J.C. (2009a). Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nature Biotechnology*, **27:** 1177–80.

Atsumi, S., Li, Z. and Liao, J.C. (2009b). Acetolactate synthase from *Bacillus subtilis* serves as a 2-ketoisovalerate decarboxylase for isobutanol biosynthesis in *Escherichia coli*. *Applied and Environmental Microbiology*, **75:** 6306-11.

Aziz, Z. and Gozan, M. (2017). Production of gasohol from isobutanol. *Earth and Environmental Science*, **65:** 012055.

Barbirato, F., Grivet, J.P., Soucaille, P. and Bories, A. (1996). 3-Hydroxypropionaldehyde, an inhibitory metabolite of glycerol fermentation to 1,3-propanediol by enterobacterial species. *Applied Environmental Microbiology,* **62:** 1448–1451.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science,* **315:** 1709–1712.

Bartosiak-Jentys, J., Eley, K. and Leak, D.J. (2012). Application of *pheB* as a reporter gene for *Geobacillus spp*., enabling qualitative colony screening and quantitative analysis of promoter strength. *Applied and Environmental Microbiology,* **78:** 5945-7.

Bartosiak-Jentys, J., Hussein, A.H., Lewis, C.J. and Leak, D.J. (2013). Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*. *Microbiology (Reading),* **159:** 1267-1275.

Bashir, Z. (2018). Metabolic engineering of *Geobacillus thermoglucosidasius* for utilisation of biomass and production of 3-hydroxy propionic acid. *PhD Thesis* University of Nottingham.

Bashir, Z., Sheng, L., Anil, A., Lali, A., Minton, N.P. & Zhang, Y. (2019). Engineering *Geobacillus thermoglucosidasius* for direct utilisation of holocellulose from wheat straw. *Biotechnology for Biofuels,* **12:** 199.

Bayer, C.N., Sepulchro, A.G.V., Rennig, M. and Nørholm, M.H.H. (2022). Efficient Bacterial Genome Engineering throughout the Central Dogma Using the Dual-Selection Marker *tetA*^{OPT}. *ACS Synthetic Biology*, **11:** 3440-3450.

Berg, I.A., Kockelkorn, D., Buckel, W. and Fuchs, G. (2007). A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science,* **318:** 1782-6.

Bhatia, S.K., Kim, S., Yoon, J. and Yang, Y. (2017). Current status and strategies for second-generation biofuel production using microbial systems. *Energy Conversion and Management*, **148:** 1142–1156.

Bhatti, H.N., Hanif, M.A., Qasim, M. and Ata-ur-Rehman. (2008). Biodiesel production from waste tallow. *Fuel,* **87:** 296-2966.

Brekke, K. (2007). Butanol: an energy alternative? *Ethanol Today*, 36–39.

Campos-Fernandez, J., Arnal, J.M., Gomez, J. and Porado, M.P. (2012). A comparison of higher alcohols/diesel fuel blends in a diesel engine. *Applied Energy,* **95:** 267–275.

Cañadas, I.C., Groothuis, D., Zygouropoulou, M., Rodrigues, R. and Minton, N.P. (2019). RiboCas: A Universal CRISPR-Based Editing Tool for *Clostridium*. *ACS Synthetic Biology*, **8:** 1379-1390.

Cengic, I., Cañadas, I.C., Minton, N.P. and Hudson, E.P. (2022). Inducible CRISPR/Cas9 Allows for Multiplexed and Rapidly Segregated Single-Target Genome Editing in *Synechocystis* Sp. PCC 6803. *ACS Synthetic Biology,* **11:** 3100-3113.

Chen, C. and Liao, J.C. (2016). Frontiers in microbial 1-butanol and isobutanol production. *FEMS Microbiology Letters*, **363(5):** 1-13.

Chen, Z., Huang, J., Wu, Y., Wu, W., Zhang, Y. and Liu, D. (2017). Metabolic engineering of *Corynebacterium glutamicum* for the production of 3-hydroxypropionic acid from glucose and xylose. *Metabolic Engineering,* **39:** 151-158.

Choi, K.Y., Wernick, D.G., Tat, C.A. and Liao, J.C. (2014). Consolidated conversion of protein waste into biofuels and ammonia using *Bacillus subtilis*. *Metabolic Engineering,* **23:** 53–61.

Chylinski, K., Makarova, K.S., Charpentier, E., and Koonin, E.V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Research,* **42:** 6091–6105.

Cripps, R.E., Eley, K., Leak, D.J., Rudd, B., Taylor, M., Todd, M., Boakes, S., Martin, S. and Atkinson, T. (2009). Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metabolic Engineering,* **11:** 398–408.

Cui, W., Han, L., Cheng, J., Liu, Z., Zhou, L., Guo, J. and Zhou, Z. (2016). Engineering an inducible gene expression system for *Bacillus subtilis* from a strong constitutive promoter and a theophylline-activated synthetic riboswitch. *Microbial Cell Factories,* **15:** 199.

Dave, H., Ramakrishna, C. and Desai, J.D. (1996). Degradation of acrylic acid by fungi from petrochemical activated sludge. *Biotechnology Letters,* **18:** 963-964.

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

Demirbas A. (2008). Comparison of transesterification methods for the production of biodiesel from vegetable oils and fats. *Energy Conversion and Management*, **49:** 125-130.

Dennis, J.S., Scott, S.A. and Stephenson, A.L. (2008). Improving the sustainability of the production of biodiesel from oilseed rape in the UK. *Process Safety and Environmental Protection*, **86:** 427440.

Dickinson, J. (2000). Pathways of leucine and valine catabolism in yeast. *Method Enzymology,* **324:** 80–92.

Dietrich, C., Nato, A., Bost, B., Le Maréchal, P. and Guyonvarch, A. (2009). Regulation of *ldh* expression during biotin-limited growth of *Corynebacterium glutamicum*. *Microbiology (Reading),* **155:** 1360-1375.

Durre, P. (2007). Biobutanol: an attractive biofuel, *Biotechnology Journal*, **2:** 1525–1534.

Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Seyboth, K., Matschoss, P., Kadner, S. and von Stechow, C. (2011). *Renewable Energy Sources and Climate Change Mitigation*. Cambridge: Cambridge University Press.

Ehsaan, M., Baker, J., Kovács, K., Malys, N. and Minton, N.P. (2021). The pMTL70000 modular, plasmid vector series for strain engineering in *Cupriavidus necator* H16. *Journal of Microbiological Methods,* **189:** 106323.

Elisáková, V., Pátek, M., Holátko, J., Nesvera, J., Leyval, D., Goergen, J.L. and Delaunay, S. (2005). Feedback-resistant acetohydroxy acid synthase increases valine production in *Corynebacterium glutamicum*. *Applied and Environmental Microbiology,* **71:** 207-13.

Escobar, J.C., Lora, E.S., Venturini, O.J., Yanez, E.E., Castillo, E.F. and Almazan, O. (2009). Biofuels: environment, technology and food security. *Renewable and Sustainable Energy Reviews,* **13:** 1275-1287.

Garai-Ibabe, G., Ibarburu, I., Berregi, I., Claisse, O., Lonvaud-Funel, A., Irastorza, A. and Dueñas, M.T. (2008). Glycerol metabolism and bitterness producing lactic acid bacteria in cidermaking. *International Journal of Food Microbiology,* **121:** 253-61.

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

Gheshlaghi, R., Scharer, J.M., Moo-Young, M. and Chou, C.P. (2009). Metabolic pathways of clostridia for producing butanol. *Biotechnology Advances,* **27:** 764-781.

Grana, R., Frassoldati, A., Faravelli, T., Niemann, U., Ranzi, E., Seiser, R., Cattolica, R. and Seshadri, K. (2010). An experimental and kinetic modelling study of combustion of isomers of butanol. *Combustion Flame,* **157:** 2137–2154.

Guettler, M.V., Rumler, D. and Jain, M.K. (1999). *Actinobacillus succinogenes* sp. nov., a novel succinic-acid-producing strain from the bovine rumen. *International Journal of Systematic Bacteriology,* **49(1):** 207-16.

Haghighi Mood, S., Hossein Golfeshan, A., Tabatabaei, M., Salehi Jouzani, G., Najafi, G.H., Gholami, M. and Ardjmand, M. (2013). Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment. *Renewable and Sustainable Energy Reviews,* **27:** 77-93.

Hao, M., Cui, Y. and Qu, X. (2018). Analysis of CRISPR-Cas System in *Streptococcus thermophilus* and Its Application. *Frontiers in Microbiology,* **9:** 257.

Hazelwood, L.A., Daran, J.M., van Maris, A.J.A., Pronk, J.T. and Dickinson, J.R. (2008). The Ehrlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* metabolism. *Applied and Environmental Microbiology,* **74:** 2259–2266.

Heap, J.T., Pennington, O.J., Cartman, S.T. and Minton, N.P. (2009). A modular system for *Clostridium* shuttle plasmids. *Journal of Microbiological Methods,* **78:** 79-85.

Heap, J.T., Ehsaan, M., Cooksley, C.M., Ng, Y.K., Cartman, S.T., Winzer, K. and Minton, N.P. (2012). Integration of DNA into bacterial chromosomes from

plasmids without a counter-selection marker. *Nucleic Acids Research,* **40(8):** e59.

Herrmann, G., Selmer, T., Jessen, H.J., Gokarn, R.R., Selifonova, O., Gort, S.J. and Buckel, W. (2005). Two beta-alanyl-CoA: ammonia lyases in *Clostridium propionicum*. *FEBS Journal*, **272:** 813-821.

Hewitt, J. and Lokare, K.S. (2016). Isobutanol: A New Product from Revamped Ethanol Plants. Accessed 4th May 2019.

Higashide, W., Li, Y., Yang, Y. and Liao, J.C. (2011). Metabolic engineering of *Clostridium cellulolyticum* for production of isobutanol from cellulose. *Applied and Environmental Microbiology*, **77:** 2727-33.

Hoekman, S.K. (2009). Biofuels in the U.S. - challenges and opportunities. *Renewable Energy*, **34:** 14-22.

Holo, H. (1989). *Chloroflexus aurantiacus* secretes 3-hydroxypropionate, a possible intermediate in the assimilation of CO_2 and acetate. *Archives of Microbiology,* **151:** 252-256.

Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. and Pease, L.R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, **77:** 61-8.

Hryhorowicz, M., Lipinski, D., Zeyland, J. and Słomski, R. (2017). CRISPR/Cas9 Immune System as a Tool for Genome Engineering. *Archivum Immunologiae et Therapiae Experimentalis,* **65:** 233–240.

Hügler, M., Huber, H., Stetter, K.O. and Fuchs, G. (2003). Autotrophic CO_2 fixation pathways in archaea (Crenarchaeota). *Archives of Microbiology,* **179:** 160-73.

Hügler, M., Menendez, C., Schägger, H. and Fuchs, G. (2002). Malonyl-coenzyme A reductase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic $CO_{(2)}$ fixation. *Journal of Bacteriology,* **184:** 2404-10.

Hussein, A.H., Lisowska, B.K. and Leak, D.J. (2015). The genus *Geobacillus* and their biotechnological potential. *Advances in Applied Microbiology*, **92:** 1-48.

International Energy Agency (IEA) (2021), Oil Market Report - October 2021, IEA, Paris.

International Energy Agency (IEA) (2022), Global Energy Review: CO2 Emissions in 2021, IEA, Paris.

Ishii, M., Chuakrut, S., Arai, H. and Igarashi, Y. (2004). Occurrence, biochemistry and possible biotechnological application of the 3-hydroxypropionate cycle. *Applied Microbiology and Biotechnology,* **64:** 605-10.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli* and identification of the gene product. *Journal of Bacteriology*, **169:** 5429-5433.

Jers, C., Kalantari, A., Garg, A. and Mijakovic, I. (2019). Production of 3-Hydroxypropanoic Acid From Glycerol by Metabolically Engineered Bacteria. *Frontiers in Bioengineering and Biotechnology*, **7:** 124.

Jiang, J., Huang, B., Wu, H., Li, Z. and Ye, Q. (2018). Efficient 3-hydroxypropionic acid production from glycerol by metabolically engineered *Klebsiella pneumoniae*. *Bioresources and Bioprocessing,* **5:** 34.

Jiang, X., Meng, X. and Xian, M. (2009). Biosynthetic pathways for 3-hydroxypropionic acid production. *Applied Microbiology and Biotechnology*, **82:** 995-1003.

Jiang, Y., Xin, F., Lu, J., Dong, W., Zhang, W., Zhang, M., Wua, H., Maa, J. and Jiang, M. (2017). State of the art review of biofuels production from lignocellulose by thermophilic bacteria. *Bioresource Technology*, **245:** 1498–1506.

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

Kalantari, A., Chen, T., Ji, B., Stancik, I.A., Ravikumar, V., Franjevic, D., Saulou-Bérion, C., Goelzer, A. and Mijakovic, I. (2017). Conversion of Glycerol to 3-Hydroxypropanoic Acid by Genetically Engineered *Bacillus subtilis*. *Frontiers in Microbiology*, **8:** 638.

Kananaviciute, R. and Citavicius, D. (2015). Genetic engineering of *Geobacillus* spp. *Journal of Microbiological Methods,* **111:** 31-39.

Keller, M.W., Schut, G.J., Lipscomb, G.L., Menon, A.L., Iwuchukwu, I.J., Leuko, T.T., Thorgersen, M.P., Nixon, W.J., Hawkins, A.S., Kelly, R.M. and Adams, M.W. (2013). Exploiting microbial hyperthermophilicity to produce an industrial chemical, using hydrogen and carbon dioxide. *Proceedings of the National Academy of Science U S A,* **110:** 5840-5.

Kildegaard, K.R., Jensen, N.B., Schneider, K., Czarnotta, E., Özdemir, E., Klein, T., Maury, J., Ebert, B.E., Christensen, H.B., Chen, Y., Kim, I.K., Herrgård, M. J., Blank, L.M., Forster, J., Nielsen, J. and Borodina, I. (2016). Engineering and systems-level analysis of *Saccharomyces cerevisiae* for production of 3-hydroxypropionic acid via malonyl-CoA reductase-dependent pathway. *Microbial Cell Factories*, **15:** 53.

Ko, Y., Seol, E., Sundara Sekar, B., Kwon, S., Lee, J. and Park, S. (2017). Metabolic engineering of *Klebsiella pneumoniae* J2B for co-production of 3-hydroxypropionic acid and 1,3-propanediol from glycerol: Reduction of acetate and other by-products. *Bioresource Technology*, **244:** 1096-1103.

Konig, S. (1998). Subunit structure, function and organisation of pyruvate decarboxylases from various organisms. *Biochimica et Biophysica Acta*, **1385:** 271–286.

Kruyt, B., van Vuuren, D.P., de Vries, H. and Groenenberg, H. (2009). Indicators for energy security. *Energy Policy,* **37:** 2166–2181.

Kumar, R.R. and Prasad, S. (2011). Metabolic Engineering of Bacteria. *Indian Journal of Microbiology*, **51(3):** 403-409.

Kumar, V., Ashok, S. and Park, S. (2013). Recent advances in biological production of 3-hydroxypropionic acid. *Biotechnology Advances,* **31:** 945-61.

la Grange, D.C., den Haan, R. and van Zyl, W.H. **(**2010). Engineering cellulolytic ability into bioprocessing organisms. *Applied Microbiology and Biotechnology*, **87:** 1195–1208.

Lan, E.I., Chuang, D.S., Shen, C.R., Lee, A.M., Ro, S.Y. and Liao, J.C. (2015). Metabolic engineering of cyanobacteria for photosynthetic 3-hydroxypropionic acid production from CO_2 using *Synechococcus elongatus* PCC 7942. *Metabolic Engineering,* **31:** 163-70.

Larson, E.D. (2008). Biofuel production technologies: status, prospects and implications for trade and development. Report No. UNCTAD/DITC/TED/2007/10. United Nations Conference on Trade and Development, New York and Geneva.

Lau, M.S.H. (2018). The Expansion of Available Bioparts and the Development of CRISPR/Cas9 for Genetic Engineering of *Geobacillus thermoglucosidasius*. *PhD Thesis*, University of Nottingham.

Lau, M.S.H., Sheng, L., Zhang, Y. and Minton, N.P. (2021). Development of a Suite of Tools for Genome Editing in *Parageobacillus thermoglucosidasius* and Their Use to Identify the Potential of a Native Plasmid in the Generation of Stable Engineered Strains. *ACS Synthetic Biology,* **10:** 1739-1749.

Lee, P., Raj, S.M., Zhou, S., Ashok, S., Edwardraja, S. and Park, S. (2014). 3-hydroxyisobutyrate dehydrogenase-I from *Pseudomonas denitrificans* ATCC 13867 degrades 3-hydroxypropionic acid. *Biotechnology and Bioprocess Engineering,* **19:** 1-7.

Lee, S.Y., Park, J.H., Jang, S.H., Nielsen, L.K., Kim, J. and Jung, K.S. (2008). Fermentative butanol production by *Clostridia*. *Biotechnology Bioengineering,* **101:** 209–228.

Lee, W.H., Seo, S.O., Bae, Y.H., Nan, H., Jin, Y.S. and Seo, J.H. (2012). Isobutanol production n engineered *Saccharomyces cerevisiae* by overexpression of 2-ketoisovalerate decarboxylase and valine biosynthetic enzymes. *Bioprocessing Biosystem Engineering,* **35:** 1467–75.

Li, H., Opgenorth, P.H., Wernick, D.G., Rogers, S., Wu, T.Y., Higashide, W., Malati, P., Huo, Y.X., Cho, K.M. and Liao, J.C. (2012). Integrated electro microbial conversion of CO_2 to higher alcohols. *Science,* **335:** 1596.

Li, Y., Wang, X., Ge, X. and Tian, P. (2016). High Production of 3-Hydroxypropionic Acid in *Klebsiella pneumoniae* by Systematic Optimization of Glycerol Metabolism. *Scientific Reports,* **6:** 26932.

Liang, J., Roberts, A., Van Kranenburg, R., Bolhuis, A. and Leak, D. (2022). Relaxed control of sugar utilization in *Parageobacillus thermoglucosidasius* DSM 2542. *Microbiological Research***,** 126957.

Liao, H.H. and Kanikula, A.M. (1990). Increased efficiency of transformation of *Bacillus stearothermophilus* by a plasmid carrying a thermostable kanamycin resistance marker. *Current Microbiology,* **21:** 301-306.

Liao, H.H., Gokarn, R.R., Gort, S.J., Jessen, H.J. and Selifonova, O.V. (2005). Alanine 2,3-aminomutase. Patent application no. EP1575881.

Liao, J.C., Mi, L., Pontrelli, S. and Luo, S. (2016). Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nature Reviews Microbiology,* **14:** 288-304.

Lim, H.G., Noh, M.H., Jeong, J.H., Park, S. and Jung, G.Y. (2016). Optimum Rebalancing of the 3-Hydroxypropionic Acid Production Pathway from Glycerol in *Escherichia coli*. *ACS Synthetic Biology,* **5:** 1247-1255.

Lin, P.P., Mi, L., Morioka, A.H., Yoshino, K.M., Konishi, S., Xu, S.C., Papanek, B. A., Riley, L.A., Guss, A.M. and Liao, J.C. (2015). Consolidated bioprocessing of cellulose to isobutanol using *Clostridium thermocellum*. *Metabolic Engineering,* **31:** 44-52.

Lin, P.P., Rabe, K.S., Takasumi, J.L., Kadisch, M., Arnold, F.H. and Liao, J.C. (2014). Isobutanol production at elevated temperatures in thermophilic *Geobacillus thermoglucosidasius*. *Metabolic Engineering,* **24:** 1-8.

Liu, Z. and Fei, B. (2013). 'Characteristics of Moso Bamboo with Chemical Pretreatment', in A. K. Chandel, S. S. d. Silva (eds.), Sustainable Degradation of Lignocellulosic Biomass -Techniques, Applications and Commercialization, IntechOpen, London.

Love, J. (2022). Microbial pathways for advanced biofuel production. *Biochemical Society Transactions,* **50:** 987-1001.

Lynd, L.R., Van Zyl, W.H., Mcbride, J.E. and Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: an update. *Current Opinion in Biotechnology,* **16:** 577-83.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Terns, R.M., Terns, M.P., White, M.F., Yakunin, A.F., Garrett, R.A., van der Oost, J., Backofen, R. and Koonin, E.V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology,* **13:** 722–736.

Matsakas, L., Hrůzová, K., Rova, U. and Christakopoulos, P. (2018). Biological Production of 3-Hydroxypropionic Acid: An Update on the Current Status. *Fermentation,* **4:** 13.

Mattanovich, D., Sauer, M. and Gasser, B. (2014). Yeast biotechnology: teaching the old dog new tricks. *Microbial Cell Factories,* **13:** 34.

Maurya, D.P., Singla, A. and Negi, S. (2015). An overview of key pretreatment processes for biological conversion of lignocellulosic biomass to bioethanol. *3. Biotech,* **5:** 597–609.

Miah, R., Siddiqa, A., Chakraborty, U., Tuli, J.F., Barman, N.K., Uddin, A., Aziz, T., Sharif, N., Dey, S.K., Yamada, M. and Talukder, A. A. (2022). Development of high temperature simultaneous saccharification and fermentation by thermosensitive *Saccharomyces cerevisiae* and *Bacillus amyloliquefaciens*. *Scientific Reports,* **12:** 3630.

Minton, N.P., Ehsaan, M., Humphreys, C.M., Little, G.T., Baker, J., Henstra, A.M., Liew, F., Kelly, M.L., Sheng, L., Schwarz, K. and Zhang, Y. (2016). A roadmap for gene system development in *Clostridium*. *Anaerobe*, **41:** 104-112.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, **155:** 733-740.

Morone, A. and Pandey, R.A. (2014). Lignocellulosic biobutanol production: gridlocks and potential remedies, *Renewable and Sustainable Energy Reviews*, **37:** 21–35.

Mougiakos, I., Bosma, E.F., de Vos, W.M., van Kranenburg, R. and van der Oost, J. (2016). Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. *Trends in Biotechnology*, **34(7):** 575-587.

Mougiakos, I., Bosma, E.F., Ganguly, J., van der Oost, J. and van Kranenburg, R. (2018). Hijacking CRISPR-Cas for high-throughput bacterial metabolic engineering: advances and prospects. *Current Opinion in Biotechnology*, **50:** 146–157.

Müller, M., Lee, C.M., Gasiunas, G., Davis, T.H., Cradick, T.J., Siksnys, V., Bao, G., Cathomen, T. and Mussolino, C. (2016). *Streptococcus thermophilus* CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. *Molecular Therapy*, **24:** 636-44.

Nazina, T.N., Tourova, T.P., Poltaraus, A.B., Novikova, E.V., Grigoryan, A.A., Ivanova, A.E., Lysenko, A.M., Petrunyaka, V.V., Osipov, G.A., Belyaev, S.S. and Ivanov, M.V. (2001). Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus, Bacillus thermo-catenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus, G. thermocatenulatus, G. thermoleovorans, G. kaustophilus, G. thermoglucosidasius* and *G. thermodenitrificans*. *International Journal of Systematic Evolutionary Microbiology*, **51:** 433–446.

Ndaba, B., Chiyanzu, I. and Marx, S. (2015). *n*-Butanol derived from biochemical and chemical routes: A review. *Biotechnology Reports*, **8:** 1-9.

Nigam, P.S. and Singh, A. (2011). Production of liquid biofuels from renewable resources. *Progress in Energy and Combustion Science*, **37:** 52-68.

Otzen, C., Bardl, B., Jacobsen, I.D., Nett, M. and Brock, M. (2014). *Candida albicans* utilizes a modified β-oxidation pathway for the degradation of toxic propionyl-CoA. *Journal of Biological Chemistry*, **289:** 8151-69.

Owusu, P.A. and Asumadu-Sarkodie, S. (2016). A review of renewable energy sources, sustainability issues and climate change mitigation. *Cogent Engineering* **3:** 1167990.

Patil, V., Tran, K.Q. and Giselrod, H.R. (2008). Towards sustainable production of biofuels from Microalgae. *International Journal of Molecular Sciences*, **9:** 1188-95.

Pennington, O.J. (2006). The development of molecular tools for the expression of prodrug converting enzymes in *Clostridium sporogenes*. *PhD Thesis* University of Nottingham.

Pitera, D.J., Paddon, C.J., Newman, J.D. and Keasling, J.D. (2007). Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metabolic Engineering,* **9:** 193–207.

Qin, L., Zhao, X., Li, W.C., Zhu, J.Q., Liu, L., Li, B. Z. and Yuan, Y.J. (2018). Process analysis and optimization of simultaneous saccharification and co-fermentation of ethylenediamine-pretreated corn stover for ethanol production. *Biotechnology for Biofuels,* **11:** 118.

Raita, M., Ibenegbu, C., Champreda, V. and Leak, D.J. (2016). Production of ethanol by thermophilic oligosaccharide utilising *Geobacillus thermoglucosidasius* TM242 using palm kernel cake as a renewable feedstock. *Biomass and Bioenergy,* **95:** 45-54.

Ramey D. (2004). Butanol advances in biofuels. The Light Party, lightparty.com/Energy/Butanol.html.

Rathnasingh, C., Raj, S.M., Lee, Y., Catherine, C., Ashok, S. and Park, S. (2012). Production of 3-hydroxypropionic acid via malonyl-CoA pathway using recombinant *Escherichia coli* strains. *Journal of Biotechnology,* **157:** 633-40.

Reeve, B., Martinez-Klimova, E., De Jonghe, J., Leak, D.J. and Ellis, T. (2016). The *Geobacillus* Plasmid Set: A Modular Toolkit for Thermophile Engineering. *ACS Synthetic Biology,* **5:** 1342-1347.

Ritchie, H. and Roser, M. (2020) – "Energy". Published online at OurWorldInData.org.

Salinas, A., Mcgregor, C., Irorere, V., Arenas-López, C., Bommareddy, R.R., Winzer, K., Minton, N.P. and Kovács, K. (2022). Metabolic engineering of *Cupriavidus necator* H16 for heterotrophic and autotrophic production of 3-hydroxypropionic acid. *Metabolic Engineering,* **74:** 178-190.

Sarathy, S., Vranckx, S., Yasunanga, K., Mehl, M., Osswald, P., Metcalfe, W.K., Westbrook, C.K., Pitz, W.J., Kohse-Hoingaus, K., Fernandes, R.X. and Curran, H.J. (2012). A comprehensive chemical kinetic combustion model for the four butanol isomers. *Combustion Flame,* **159:** 2028–2055.

Schneider, K., Asao, M., Carter, M.S. and Alber, B.E. (2012). *Rhodobacter sphaeroides* uses a reductive route via propionyl coenzyme A to assimilate 3-hydroxypropionate. *Journal of Bacteriology,* **194:** 225-32.

Schwarz, M., Köpcke, B., Weber, R.W., Sterner, O. and Anke, H. (2004). 3-hydroxypropionic acid as a nematicidal principle in endophytic fungi. *Phytochemistry,* **65:** 2239-45.

Sentheshanmuganathan, S. and Elsden, S.R. (1958). The mechanism of the formation of tyrosol by *Saccharomyces cerevisiae*. *Biochemical Journal,* **69:** 10–218.

Seys, F.M., Rowe, P., Bolt, E.L., Humphreys, C.M. and Minton, N.P. (2020). A Gold Standard, CRISPR/Cas9-Based Complementation Strategy Reliant on 24 Nucleotide Bookmark Sequences. *Genes,* **11:** 458.

Shaw, A.J., Podkaminer, K.K., Desai, S.G., Bardsley, J.S., Rogers, S.R., Thorne, P.G., Hogsett, D.A. and Lynd, L.R., (2008). Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *Proceedings of the National Academy of Science,* **105:** 13769–13774.

Sheng, L., Kovács, K., Winzer, K., Zhang, Y. and Minton, N. P. (2017). Development and implementation of rapid metabolic engineering tools for chemical and fuel production in *Geobacillus thermoglucosidasius* NCIMB 11955. *Biotechnology for Biofuels*, **10:** 5.

Sheng, L., Zhang, Y. and Minton, N.P. (2016). Complete Genome Sequence of *Geobacillus thermoglucosidasius* NCIMB 11955, the Progenitor of a Bioethanol Production Strain. *Genome Announcements*, **4(5):** e01065-16.

Singh, Y.D. and Satapathy, K.B. (2018). Conversion of Lignocellulosic Biomass to Bioethanol: An Overview with a Focus on Pretreatment. *International Journal of Engineering and Technologies,* **15:** 17-43.

Smith, K.M., Cho, K.M, and Liao, J.C. (2010). Engineering *Corynebacterium glutamicum* for isobutanol production. *Applied Microbiology and Biotechnology*, **87:** 1045–1055.

Sobolov, M. and Smiley, K.L. (1960). Metabolism of glycerol by an acrolein-forming *Lactobacillus*. *Journal of Bacteriology,* **79:** 261-6.

Soh, L. M.J., Mak, W.S., Lin, P.P., Mi, L., Chen, F.Y.H., Damoiseaux, R., Siegel, J. B. and Liao, J.C. (2017). Engineering a Thermostable Keto Acid Decarboxylase Using Directed Evolution and Computationally Directed Protein Design. *ACS Synthetic Biology,* **6:** 610-618.

Song, C.W., Kim, J.W., Cho, I.J. and Lee, S.Y. (2016). Metabolic Engineering of *Escherichia coli* for the Production of 3-Hydroxypropionic Acid and Malonic Acid through β-Alanine Route. *ACS Synthetic Biology*, **5:** 1256-1263.

Sorek, R., Kunin, V. and Hugenholtz, P. (2008). CRISPR-a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology*, **6:** 181–186.

Spencer, J. (2018). Metabolic engineering of the thermophile *Geobacillus* to produce the advanced biofuel *n*-butanol. *PhD Thesis* University of Nottingham.

Stevens, D.J., Worgetten, M. and Saddler, J. (2004). Biofuels for transportation: an examination of policy and technical issues. IEA Bioenergy Task 39, Liquid Biofuels Final Report 2001e2003. Canada.

Straathof, A.J., Sie, S., Franco, T.T. and Van Der Wielen, L.A. (2005). Feasibility of acrylic acid production by fermentation. *Applied Microbiology and Biotechnology,* **67:** 727-34.

Styles, M.Q., Nesbitt, E.A., Hoffmann, T.D., Queen, J., Ortenzi, M.V. and Leak, D. J. (2021). The heterologous production of terpenes by the thermophile *Parageobacillus thermoglucosidasius* in a consolidated bioprocess using waste bread. *Metabolic Engineering*, **65:** 146-155.

Suyama, A., Higuchi, Y., Urushihara, M., Maeda, Y. and Takegawa, K. (2017). Production of 3-hydroxypropionic acid via the malonyl-CoA pathway using recombinant fission yeast strains. *Journal of Bioscience and Bioengineering,* **124:** 392-399.

Suzuki, Y., Kishigami, T., Inoue, K., Mizoguchi, Y., Eto, N., Takagi, M. and Abe, S. (1983). *Bacillus thermoglucosidasius* sp. nov., a New Species of obligately Thermophilic Bacilli. *Systematic and Applied Microbiology,* **4:** 487-495.

Szambelan, K., Nowak, J., Szwengiel, A., Jeleń, H. and Łukaszewski, G. (2018). Separate hydrolysis and fermentation and simultaneous saccharification and fermentation methods in bioethanol production and formation of volatile by-products from selected corn cultivars. *Industrial Crops and Products*, **118:** 355-361.

Takamizawa, K., Horitsu, H., Ichikawa, T., Kawai, K. and Suzuki, T. (1993). β-hydroxypropionic acid production by *Byssochlamys* sp. grown on acrylic acid. *Applied Microbiology and Biotechnology*, **40:** 196-200.

Talarico, T.L., Casas, I.A., Chung, T.C. and Dobrogosz, W.J. (1988). Production and isolation of reuterin, a growth inhibitor produced by *Lactobacillus reuteri*. *Antimicrobial Agents and Chemotherapy,* **32:** 1854-8.

Tang, Y.J., Sapra, R., Joyner, D., Hazen, T.C., Myers, S., Reichmuth, D., Blanch, H. and Keasling, J.D. (2009). Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant *Geobacillus* strain. *Biotechnology and Bioengineering,* **102:** 1377-1386.

Taylor, M.P., Esteban, C.D. and Leak, D.J. (2008). Development of a versatile shuttle vector for gene expression in *Geobacillus* spp. *Plasmid*, **60:** 45-52.

Taylor, M.P., Eley, K.L., Martin, S., Tuffin, M.I., Burton, S.G. and Cowan, D.A. (2009). Thermophilic ethanologenesis: future prospects for second-generation bioethanol production. *Trends in Biotechnology*, **27:** 398-405.

Taylor, M.P., van Zyl, L., Tuffin, M.I., Leak, D.J., and Cowan, D.A. (2011). Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts. *Microbial Biotechnology*, **4:** 438-448.

The state of food and agriculture (2008). BIOFUELS: prospects, risks and opportunities. Food and Agriculture Organization of the United Nations. p.11

Topp, S., Reynoso, C.M.K., Seeliger, J.C., Goldlust, I.S., Desai, S.K., Murat, D., Shen, A., Puri, A.W., Komeili, A., Bertozzi, C.R., Scott, J.R. and Gallivan, J.P. (2010). Synthetic Riboswitches That Induce Gene Expression in Diverse Bacterial Species. *Applied and Environmental Microbiology,* **76:** 7881-7884.

Tucker, B.J. and Breaker, R.R. (2005). Riboswitches as versatile gene control elements. *Current Opinion in Structural Biology,* **15:** 342-348.

United Nations Framework Convention on Climate Change (UNFCCC) (2015). *Adoption of the Paris agreement*.

Wackett, L.P. (2008). Biomass to fuels via microbial transformations. *Current Opinion in Chemical Biology*, **12:** 187–193.

Wang, Y., Sun, T., Gao, X., Shi, M., Wu, L., Chen, L. and Zhang, W. (2016). Biosynthesis of platform chemical 3-hydroxypropionic acid (3-HP) directly from CO_2 in cyanobacterium *Synechocystis* sp. PCC 6803. *Metabolic Engineering*, **34:** 60-70.

Werpy, T. and Petersen, G. (2004). Top Value Added Chemicals from Biomass: Volume I -- Results of Screening for Potential Candidates from Sugars and Synthesis Gas. United States.

Wiedenheft, B., Sternberg, S.H. and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, **482:** 331–338.

Winkler, W.C. and Breaker, R.R. (2005). Regulation of bacterial gene expression by riboswitches. *Annual Review of Microbiology*, **59:** 487-517.

Xiao, Z., Wang, X., Huang, Y., Huo, F., Zhu, X., Xi, L. and Lu, J.R. (2012). Thermophilic fermentation of acetoin and 2,3-butanediol by a novel *Geobacillus* strain. *Biotechnology for biofuels*, **5:** 88-88.

Xu, K., Ren, C., Liu, Z., Zhang, T., Zhang, T., Li, D., Wang, L., Yan, Q., Guo, L., Shen, J. and Zhang, Z. (2015). Efficient genome engineering in eukaryotes using Cas9 from *Streptococcus thermophilus*. *Cellular and Molecular Life Sciences*, **72:** 383-99.

Yang, Y.M., Chen, W.J., Yang, J., Zhou, Y.M., Hu, B., Zhang, M., Zhu, L.P., Wang, G.Y. and Yang, S. (2017). Production of 3-hydroxypropionic acid in engineered *Methylobacterium extorquens* AM1 and its reassimilation through a reductive route. *Microbial Cell Factories*, **16:** 179.

Yao, T., Xu, L., Ying, H., Huang, H. and Yan, M. (2010). The catalytic property of 3-hydroxyisobutyrate dehydrogenase from *Bacillus cereus* on 3-hydroxypropionate. *Applied Biochemistry and Biotechnology*, **160:** 694-703.

Zaushitsyna, O., Dishisha, T., Hatti-Kaul, R. and Mattiasson, B. (2017). Crosslinked, cryostructured *Lactobacillus reuteri* monoliths for production of 3-hydroxypropionaldehyde, 3-hydroxypropionic acid and 1,3-propanediol from glycerol. *Journal of Biotechnology*, **241:** 22-32.

Zeldes, B.M., Keller, M.W., Loder, A.J., Straub, C.T., Adams, M.W.W. and Kelly, R.M.(2015). Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. *Frontiers in Microbiology*, **6:** 1209.

Zhang, Y.H. and Lynd, L.R. (2004). Toward an aggregated understanding of enzymatic hydrolysis of cellulose: non-complexed cellulose systems. *Biotechnology Bioengineering*, **88:** 797–824.

Zhou, J., Lian, J. and Rao, C.V. (2020). Metabolic engineering of *Parageobacillus thermoglucosidasius* for the efficient production of (2R, 3R)-butanediol. *Applied Microbiology and Biotechnology*, **104:** 4303-4311.

Zhou, J., Wu, K. and Rao, C.V. (2016). Evolutionary engineering of *Geobacillus thermoglucosidasius* for improved ethanol production. *Biotechnology and Bioengineering*, **113:** 2156-67.

Zhou, S., Ainala, S.K., Seol, E., Nguyen, T.T. and Park, S. (2015). Inducible gene expression system by 3-hydroxypropionic acid. *Biotechnology for Biofuels*, **8:** 169.

Zhou, S., Ashok, S., Ko, Y., Kim, D.M. and Park, S. (2014). Development of a deletion mutant of *Pseudomonas denitrificans* that does not degrade 3-hydroxypropionic acid. *Applied Microbiology and Biotechnology*, **98:** 4389-98.

Zhou, S., Catherine, C., Rathnasingh, C., Somasundar, A. and Park, S. (2013a). Production of 3-hydroxypropionic acid from glycerol by recombinant *Pseudomonas denitrificans*. *Biotechnology and Bioengineering*, **110:** 3177-87.

Zhou, S., Mohan Raj, S., Ashok, S., Edwardraja, S., Lee, S.G. and Park, S. (2013b). Cloning, expression and characterization of 3-hydroxyisobutyrate dehydrogenase from *Pseudomonas denitrificans* ATCC 13867. *PLoS One*, **8:** e62666.

APPENDICES

Appendix I: DNA sequence of Gram-positive replicon-*repBST1*

repBST1 (AscI and FseI in red)

ggcgcgccTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGC
TCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAA
GCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAA
CTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGTTCCTTAA
GGAACGTACAGACGGCTTAAAAGCCTTTAAAAACGTTTTTAAGGGGTTTG
TAGACAAGGTAAAGGATAAAACAGCACAATTCCAAGAAAAACACGATTT
AGAACCTAAAAAGAACGAATTTGAACTAACTCATAACCGAGAGGTAAAA
AAAGAACGAAGTCGATTCGAGCACTACAGGAACAAAAAAGAGCGATGAT
TGAAGGAACGCAAAAACAAGAGAAGGTCACATTGGAAATAGCAGCGACA
ACGGAGGACAAGACAACAGAGGACAAGAGGACAAAAAAACAGGGATGG
CTGTCGAGAATTTTTGGAAAATAGACCTTGTCGCGACTCGAAGATTAGTAG
TACAATAGAGACAACATAACAGAACGGGACAAAATAGAAGAGCTTCCGA
AGCGTGTGTAAGTGGTGCTCCCAACACCGTTCGCCGTACTCGCCCTACGGC
AAACACCTGCTCGAAAGCTCTTGTGTTAACATATTTATGCCTTGTCTTCATT
TTAATGCATAGATATGTTGAATGCAAGAGCGAAACGCTTGGAAATTTTCGT
TTTTCCTCCGATGAGAGCGTGTTTGCCTGCGTATTAATGTACGGCAACGCG
CTCTTTTGTTTGTCCCGAAAAATTCGGGAGAAACGGGAGGAGAACGGAAA
TGTACATAATCACTAGCGAACAATTCACGCAACTGAAAGAAATCGAGAGA
GTCACAGATTTCGTTTATAAACGCCGTAAGATTAAAAATGAGCTGATGGA
GCGTATCGCCGACGCATTAGGCCGCAAGTTCCGCAAAGCGCAAAAAACAA
CGCGCGTCATCGACGAAATTGTTTGGCTCGCAACAGACCGCGGTTTTTCGT
TCCCAGGGCGCGAAACGCTTGCCAAAAAGTTAGGCGTGTCCTTGCCGACT
GTTGATCGTGCTACTCGCATGCTCAAAGACTCGGGCGAAGTCGTCGTTTGC
TATCGTGAGAATCCGAATAGTAACGGACCTAAAACGCCTGTATTCATTTTC
CGGAGCCATGCAAATTTCGAGCGTATTGCAGCCGTTTTAAACTTGCGTGAC
AACGAAGCTGATAAAGTAGAAAACGCCGAAAAGCCTACGGAATCAAGCG
ATTTAGCGCGTAAAACGGACGCTACCATTAGTTCACCAGTTGAAAAACAT
GACGATATTAATAAATTAATAGACAATACACATCCGTTGGATAAAATTGT
AAAATACGTCACTATTAAAGTGAACGAAGTGCAGAAGCGTCATCCTTACG
GCATTCGATATCTATCCGCATACATAAACAAAACACTCGACGACCTAATG
CAACGTGCAAAAGCGAAGGTCAAAGCCGAAAAAGCTCGCGCCCAACAAC
GCAAACAAGCCGACTATCAACCGACCGCGACCGGCTTAATTTGGTACGAC
TTCACGGCGGACACAAGCGCCAGCCAGCCGACAAGCCAGCGCAAGCCAG
CGTTTGACAGCGAAGACATGTTTGCGCCGTTATTGCAGGCTGTTGCACGAA
TTAAGGGCGTTGAGGTGTGACACACTCGTTTTAAGGCGTCTAACAGGCGC
GTAGAGACGTCTTTAAGTCGGGGGAATATTAAAAATTCCACGCCTCCTAA
AACGCGATATACGCCGATTATGACGGTCGTTTGCTTACGTACTCGTGCAAC
TACCTTCGCTCACATCGTCCAAAAATCGTCATAGTCAACGTCTTTCTTCGT
GAGCTGGCGCAAAGCCTTAATGATTTTCGATGCCGTTTTTATTGTCGGCGA
AAATTTATCTGACTGGCAAAGGCGGCTGATCGTGCTTTTGCTGACGCCTGC
TTTGTTTGCGACGTCTTGCTGCGTGATGCCGTGTTTGTCTAAAAATCGCGC
AAATTTTGATCGAGATCAGGGAATGAGTTTATAAAATAAAAAAAGCACCT
GAAAAGGTGTCTTTTTTTGATGGTTTTGAACTTGTTCTTTCTTATCTTGATA
CATATAGAAATAACGTCATTTTTATTTTAGTTGCTGAAAGGTGCGTTGAAG
TGTTGGTATGTATGTGTTTTAAAGTATTGAAAACCCTTAAAATTGGTTGCA

CAGAAAAACCCCATCTGTTAAAGTTATAAGTGACTAAACAAATAACTAAA
TAGATGGGGGTTTCTTTTAATATTATGTGTCCTAATAGTAGCATTTATTCAG
ATGAAAAATCAAGGGTTTTAGTGGACAAGACAAAAAGTGGAAAAGTGAG
ACCATGGACTCTAGTGGATCGACTAGAGTCCGACGACTGCTCCGGCCGTC
CCAGAATGGAAAAGAAGAGGGGCGGATGTCTCTCCGGTTGGCGTGGCGAC
GATTAAATATTTCAAGCACGGGAAATCGTCGTACACCGAAGTGCTGGAAT
GGGCCGCTGCGCAAAAAATCGACCGAAAAATAAATATAAATCGGATATAC
AATCGGCAATTGACGAAACTGCAAAATATCCTGTAAAGGATACGGATTTT
ATGACCGATGATGAAGAAAAGAATTTGAAACGTTTGTCTGATTTGGAGGA
AGGTTTACACCGTAAAAGGTTAATCTCCTATGGTGGTTTGTTAAAAGAAAT
ACATAAAAAATTAAACCTTGATGACACAGAAGAAGGCGATTTGATTCATA
CAGATGATGACGAAAAAGCCGATGAAGATGGATTTTCTATTATTGCAATG
TGGAATTGGGAACGGAAAAATTATTTTATTAAAGAGTAGTTCAACAAACG
GGCCAGTTTGTTGAAGATTAGATGCTATAATTGTTATTAAAAGGATTGAAG
GATGCTTAGGAAGACGAGTTATTAATAGCTGAATAAGAACGGTGCTCTCC
AAATATTCTTATTTAGAAAAGCAAATCTAAAATTATCTGAAAAGGGAggccg
gcc

Appendix II: DNA sequences of heterologous genes for 3-HP production

***mcr_C. aurantiacus* (Caur_2614)**

ATGTCCGGCACAGGACGCTTAGCGGGAAAAATTGCGTTGATTACGGGTGG
CGCGGGCAATATTGGCTCCGAATTAACACGTCGCTTTCTTGCGGAGGGTGC
GACAGTCATTATTTCGGGACGCAATAGAGCGAAACTTACGGCGTTGGCGG
AACGTATGCAAGCGGAAGCGGGAGTGCCAGCGAAACGTATTGATCTTGAA
GTCATGGATGGCAGCGACCCTGTCGCGGTACGGGCGGGCATTGAAGCGAT
TGTCGCGAGACATGGCCAAATTGATATTTTAGTCAACAACGCGGGCTCCG
CGGGAGCGCAACGCCGGTTAGCGGAAATTCCTTTAACAGAAGCGGAACTT
GGACCAGGCGCGGAAGAAACACTTCATGCGTCCATTGCGAATCTTTTAGG
CATGGGCTGGCATCTTATGAGAATTGCGGCGCCACACATGCCAGTAGGCT
CCGCCGTCATTAATGTCTCGACGATTTTTAGCAGAGCGGAATATTATGGCC
GTATTCCCTATGTCACACCAAAAGCGGCGTTAAATGCGTTAAGCCAACTTG
CGGCGAGAGAACTGGGAGCGAGAGGCATTCGTGTAAATACGATTTTTCCT
GGACCTATTGAATCCGATCGCATTCGCACGGTCTTTCAACGTATGGATCAA
CTTAAAGGACGCCCTGAAGGCGATACGGCGCATCATTTTCTTAATACAATG
AGACTTTGCCGTGCGAATGATCAAGGTGCGTTAGAACGCCGGTTTCCATCG
GTAGGCGATGTCGCGGATGCGGCGGTGTTTCTTGCGTCCGCGGAATCCGC
GGCGTTATCGGGAGAAACAATTGAAGTGACACATGGCATGGAATTACCGG
CGTGCTCCGAAACGTCCCTTCTTGCGCGTACTGACCTTCGCACAATTGATG
CGTCGGGCCGCACGACGCTTATTTGTGCGGGAGATCAAATTGAAGAAGTG
ATGGCGTTAACAGGTATGCTTAGAACGTGTGGCTCCGAAGTCATTATTGGC
TTTCGTTCCGCGGCGGCGTTGGCGCAATTTGAACAAGCGGTCAACGAGAG
CAGACGCCTTGCGGGAGCGGATTTTACACCACCTATTGCGTTACCTTTAGA
TCCTCGTGATCCAGCGACGATTGATGCGGTGTTTGATTGGGCGGGAGAAA
ATACCGGCGGCATTCATGCGGCGGTCATTCTTCCTGCGACAAGCCATGAAC
CTGCGCCGTGTGTCATTGAAGTGGATGATGAAAGAGTGCTTAATTTTCTTG
CGGATGAAATTACAGGCACGATTGTCATTGCGTCCCGGTTGGCGAGATATT
GGCAATCCCAACGGTTAACACCTGGAGCGAGAGCGCGGGGCCCTAGAGTC
ATTTTTCTTTCCAACGGCGCGGATCAAAATGGCAATGTCTATGGCCGTATT
CAATCCGCGGCGATTGGACAACTTATTCGTGTGTGGAGACATGAAGCGGA
ACTTGATTATCAACGTGCGTCCGCGGCGGGAGATCATGTGCTTCCACCAGT
CTGGGCGAATCAAATTGTTCGGTTTGCGAATCGCAGCTTAGAAGGCCTTGA
ATTTGCGTGTGCGTGGACGGCGCAACTTTTACATAGCCAAAGACATATTAA
TGAGATTACGTTAAATATTCCTGCGAATATTTCCGCGACGACGGGGGCGC
GCTCGGCGTCGGTGGGCTGGGCGGAATCCCTTATTGGACTTCATCTTGGCA
AAGTCGCGTTAATTACGGGTGGCTCCGCGGGCATTGGAGGGCAGATAGGG
CGCTTACTTGCGCTTTCGGGAGCGCGGGTAATGTTAGCGGCGCGTGATCGC
CATAAACTTGAACAAATGCAAGCGATGATTCAATCCGAATTAGCGGAAGT
AGGCTATACAGATGTCGAAGATCGTGTACATATTGCGCCAGGCTGTGATG
TCAGCTCCGAAGCGCAACTTGCGGATTTGGTCGAACGCACGTTGTCCGCGT
TTGGCACAGTGGATTATCTTATTAACAACGCGGGCATTGCGGGAGTCGAA
GAAATGGTTATTGATATGCCTGTAGAAGGCTGGAGACATACGCTTTTTGCG
AATCTTATTAGCAATTATAGCCTTATGAGAAAACTTGCGCCTCTTATGAAA
AAACAAGGCTCGGGCTATATTCTTAATGTAAGCAGCTATTTTGGGGGCGA
AAAAGATGCGGCGATTCCATATCCAAATAGAGCGGATTATGCGGTCTCCA
AAGCGGGCCAACGTGCGATGGCGGAAGTGTTTGCGAGATTTTTAGGACCT

GAAATTCAAATTAATGCGATTGCGCCAGGCCCAGTAGAAGGGGATAGACT
TCGTGGCACTGGTGAAAGACCTGGCTTGTTTGCGCGTAGAGCGCGCCTTAT
TCTTGAAAACAAAAGACTTAACGAACTTCATGCGGCGCTTATTGCGGCGG
CGCGCACAGATGAAAGATCGATGCATGAACTTGTAGAACTTTTGTTGCCTA
ATGATGTCGCGGCGTTGGAACAAAATCCTGCGGCGCCGACGGCGTTGCGG
GAATTAGCGCGTCGCTTTCGTTCCGAGGGCGATCCTGCGGCGAGCAGCTCC
TCCGCGCTTCTTAATCGTAGCATTGCGGCGAAACTTTTAGCGCGTCTTCAT
AATGGTGGCTATGTGTTACCAGCGGATATTTTTGCGAATCTTCCTAATCCA
CCAGATCCGTTCTTCACCCGAGCGCAAATTGATAGAGAAGCGCGTAAAGT
AAGAGATGGCATTATGGGCATGTTATATCTTCAACGTATGCCTACAGAATT
TGATGTAGCGATGGCGACAGTCTATTATCTTGCGGATCGCAATGTCTCGGG
AGAAACGTTTCATCCATCGGGAGGCCTTCGCTATGAACGCACACCTACAG
GTGGAGAATTGTTTGGACTTCCATCCCCTGAACGGTTAGCGGAATTAGTGG
GCAGCACAGTATATCTTATTGGAGAACACTTAACAGAACATCTTAATCTTT
TAGCGAGAGCGTATCTTGAACGCTATGGAGCGAGACAAGTGGTAATGATT
GTAGAAACAGAAACAGGCGCGGAAACAATGCGCCGGTTACTTCATGATCA
TGTCGAAGCGGGACGCCTTATGACGATTGTCGCGGGAGATCAAATTGAAG
CGGCGATTGATCAAGCGATTACACGCTATGGCAGACCAGGCCCAGTGGTC
TGCACACCGTTTAGACCGTTACCTACGGTCCCGTTGGTAGGCCGTAAAGAT
TCCGATTGGTCCACGGTGTTGTCCGAAGCGGAATTTGCGGAACTTTGTGAA
CATCAACTTACACATCATTTTCGTGTCGCGCGGAAGATTGCGTTGTCCGAT
GGAGCGTCCTTAGCGTTGGTCACACCAGAAACGACAGCGACATCCACGAC
AGAACAATTTGCGTTGGCGAATTTTATTAAAACGACACTTCATGCGTTTAC
AGCGACGATTGGAGTCGAATCGGAACGTACGGCGCAACGCATTCTTATTA
ATCAAGTAGATTTGACACGTCGGGCGAGAGCGGAAGAACCAAGAGATCCT
CATGAACGCCAACAAGAACTTGAACGGTTTATTGAAGCGGTGCTTTTAGTC
ACAGCGCCTTTACCACCTGAAGCGGATACAAGATATGCGGGCCGTATTCA
TAGAGGACGTGCGATTACGGTGCATCATCATCATCATCATTAA

***mcr_M. sedula* (MSED_0709)**

ATGCGCCGGACACTTAAAGCGGCGATTCTTGGCGCGACAGGCTTAGTTGG
CATTGAATATGTTCGTATGCTTGCGGATCATCCGTATATTAAACCTACATA
TCTTGCGGGCAAAGGCTCGGTGGGCAAACCGTATGGAGAAATTGTACGCT
GGCAAACAGTAGGCAATGTACCTAAAGAAGTAGCGAATCAAGAAGTTAA
ACCAACAGACCCGAAACTTATGGATGATGTCGATATTATTTTTAGCCCACT
TCCTCAAGGGGCGGCGGGACCAGTCGAAGAACAATTTGCGAAATTAGGCT
TTAATGTAATTTCCAATAGCCCTGATCATCGGTTTGATATGGATGTACCAA
TGATTATTCCAGAAGTGAATCCACATACGGTCACGTTAATTGATGAACAAC
GTAAACGCAGAGATTGGAAAGGCTTTATTGTGACGACGCCGTTGTGCACA
GCGCAAGGCGCGGCGATTCCACTTACACCTATCTATCAAAATTTTAAAATG
TCCGGAGTAATGATTACGACGATGCAAAGCCTTTCGGGCGCGGGCTATCC
TGGCATTGCGTCCTTGGATATTGTAGATAATGCGTTACCGTTGGGAGATGG
CTATGATGCGAAAACAGTCAAAGAAATTACACGCATTCTTTCCGAAGTCA
AACGGAATGTGCAAGAACCTGGGGTGAACGAGATTACGTTAGATGCGACA
ACACATCGCATTGCGACGATTCATGGCCATTATGAAGTCGCGTATGTGACG
TTTAAAGAAGATACAGATGTTCGTAAAGTTATGGAAAGCATGGAATCGTT
TAAGGGAGAACCTCAGGATCTTAAACTTCCTACAGCGCCTGAAAAACCGA
TTATTGTCACGACACAAGATGCGAGACCTCAAGTGTTTTTCGATCGCTGGG

CGGGCAATCCGCCAGGCATGAGCGTGGTGGTAGGCCGCTTGAAACAAGTC
AATCCTCGCACGATTCGGTTTGTCTCCCTTATTCATAATACGGTTCGCGGT
GCGGCGGGGGGGGGAGTGTTGACAGCGGAACTTCTTGTCGAAAAAGGCTA
TATTGATAAACGCCATCATCATCATCATCATTGA

***msr_S. solfataricus* (SSO_0647)**

ATGTTTATTCACATGAAATCCATTAACAAGGTCGCGGTCATTGGCGCGGGA
GTGATTGGAGTCGGCTGGACAACACTTTTGTTAGCGAAAGGCTACAAGGT
CAATTTGTATACGGAAAAGAAAGAAACGTTGGAAAAAGCGCTTGCGAAA
GTGTCCGCGTATCTTGTCAATCTTAAAAATCTGGGCATGATTAACGAAGAA
CCGGAATCCTATATTACGAATCTTACAGGCATTACGAAAATTGATGATGCG
ATTCATAATGTGGATTTTGTCATTGAAGCGATTATTGAAGATTATACAGCG
AAAAAGAACCTTTTTAAACTTCTTGATACACAACTTCCTCAAGATATTATT
ATTGCGTCCAGCACGTCCGGCCTTCTTATGACAGAAATTCAAAAAGCGAT
GATTCGCCATCCTGAACGGGGAGTAATTGCGCATCCGTGGAATCCACCAC
ATTTGTTACCGTTAGTCGAAATTGTGCCTGGTGAAAAAACGAGCAAAGAA
ACAGTTGATCTTACGAGAGAGTTTATGGAAAAACTTGATCGGGTGGTAGT
GTTGCTTCGTAAAGAAGTACCTGGCTTTATTGGCAATAGACTTGCGTTTGC
GTTGTTTCGGGAAGCGGTTAATTTGGTAGATGAAGGTGTCGCGACGGTCG
AAGATATTGATAAAGTAATGACGGCGGCGATTGGCCTTCGCTGGGCGTTT
ATGGGCCCGTTTCTCACTTATCATTTGGGCGGCGGTGAGGGGGGCATTGAA
TATTTCTTTAGCAAAGGCTTTGGCTATGGCGCGAATGAGTGGATGCATACA
CTTGCGAAATGGGATAAATTTCCGTATACAGCGGTCAAAAAAGTGTTAGA
ACAGATGAAAGAATATGAGTTCATTAAAGGCAAATCGTTTCAAGACCTTA
GCCGCTGGCGTGATGAAAAACTTGTGACGATCTACAAAGCGGTCTACAAA
GAAAAATGCAATAGCCATCATCATCATCATCATTAA

***msr_M. sedula* (MSED_1993)**

ATGACGGAAAAAGTGTCCGTGGTCGGCGCGGGAGTCATTGGAGTAGGCTG
GGCGACGTTGTTTGCGTCCAAAGGCTATTCCGTCTCGTTGTATACGGAGAA
GAAGGAAACATTAGATAAAGGCATTGAAAAACTTCGCAATTATGTGCAAG
TGATGAAAAATAACAGCCAAATTACGGAAGATGTCAATACAGTGATTAGC
CGGGTCAGCCCGACAACGAATCTTGATGAAGCGGTACGCGGCGCGAATTT
TGTAATTGAAGCAGTAATTGAAGATTATGATGCGAAGAAGAAGATTTTTG
GCTATCTTGATTCGGTGTTAGATAAAGAAGTCATTCTTGCGTCGAGCACGT
CCGGCCTTCTTATTACAGAAGTTCAAAAAGCGATGAGCAAACATCCTGAA
CGTGCGGTTATTGCGCATCCGTGGAATCCACCACATTTGTTACCGTTAGTC
GAAATTGTGCCTGGCGAAAAAACGAGCATGGAAGTGGTAGAACGCACGA
AATCCCTTATGGAAAAACTTGATCGTATTGTAGTGGTGCTTAAAAAAGAA
ATTCCTGGCTTTATTGGCAATCGCTTAGCGTTTGCGTTGTTTCGGGAAGCG
GTCTATCTTGTCGATGAGGGTGTCGCGACAGTAGAAGATATTGATAAAGT
AATGACAGCGGCGATTGGCCTTCGCTGGGCGTTTATGGGACCCTTTTTAAC
CTATCACCTGGGCGGAGGCGAAGGGGGCTTAGAATACTTTTTCAATCGGG
GCTTTGGCTATGGCGCGAATGAGTGGATGCATACACTTGCGAAATATGAT
AAATTTCCATATACGGGCGTGACGAAAGCGATTCAACAGATGAAAGAATA
TAGCTTTATTAAAGGCAAAACGTTTCAAGAAATTTCCAAGTGGAGAGATG

AAAAACTTCTTAAAGTGTACAAACTTGTCTGGGAAAAACATCATCATCATC
ATCATTAA

mcr_S. tokodaii **(ST_2171)**

GTGATTCTTATGAGACGTACACTTAAAGCGGCGATTCTTGGCGCGACAGG
ACTTGTGGGCATTGAATATGTTCGTATGTTGAGCAATCATCCGTATATTAA
ACCTGCGTATCTTGCGGGCAAAGGCTCGGTGGGCAAACCATACGGAGAAG
TGGTACGCTGGCAAACAGTTGGCCAAGTACCTAAAGAAATTGCGGATATG
GAAATTAAACCGACAGATCCTAAACTGATGGATGATGTAGATATTATTTTT
AGCCCGTTACCTCAAGGCGCGGCGGGACCAGTCGAAGAACAATTTGCGAA
AGAAGGCTTTCCAGTGATTTCCAATAGCCCGGATCATCGGTTTGATCCTGA
TGTGCCGTTACTTGTGCCGGAACTTAATCCACATACAATTAGCCTTATTGA
TGAACAACGCAAACGTCGTGAGTGGAAAGGCTTTATTGTGACGACACCGT
TGTGCACAGCGCAAGGTGCGGCGATTCCTTTGGGAGCGATTTTTAAAGATT
ACAAGATGGATGGCGCGTTTATTACGACGATTCAATCGTTGTCCGGAGCG
GGCTATCCTGGCATTCCGAGCTTGGATGTAGTAGATAATATTTTGCCGCTC
GGAGATGGCTATGATGCGAAAACGATTAAAGAAATTTTTCGCATTCTTTCC
GAAGTCAAACGTAATGTCGATGAACCGAAACTTGAAGATGTCTCGCTTGC
GGCGACAACACATCGCATTGCGACGATTCATGGCCATTATGAAGTGTTGT
ATGTGAGCTTTAAAGAAGAAACAGCGGCGGAAAAAGTCAAAGAAACGTT
AGAAAATTTTCGGGGAGAACCACAGGACCTTAAATTACCTACAGCGCCTA
GCAAACCAATTATTGTGATGAATGAGGATACACGCCCTCAAGTCTATTTTG
ATCGCTGGGCGGGAGATATTCCAGGCATGTCGGTGGTAGTAGGACGCCTT
AAACAAGTCAACAAGCGCATGATTCGGTTAGTGTCCCTTATTCATAATACG
GTCAGAGGCGCGGCGGGGGGTGGCATTCTTGCGGCGGAATTGTTGGTCGA
AAAAGGCTATATTGAAAAACATCATCATCATCATCATTAA

DNA sequences of *P. thermoglucosidasius* **NCIMB 11955 3-HP utilisation genes**

mmsA1 **(BCV53_10115)**

GTGACAGTGATAAAAAATGAAACAACTATTTTGAAAAACTTCATTAATGG
ACAATGGGTAGAATCAAGCGGGACGGAGACGCTGGAAGTGCCAAATCCG
GCGACAGGGGAAATATTGGCGCGGGTTCCGATTTCGACAAAGGAAGATGT
TGACAGGGCTGTCCAAGCGGCCAAAAAGGCATTTGAAACATGGAAAAATA
CACCGGTTCCTAAGCGGGCGCGGATTATGTTTGCCTTTCACCATTTATTGA
ACGAGCATCATGAAGAATTGGCAACGCTTGTCGTTCAAGAGAACGGGAAA
GCATATAAAGAAGCGTACGGCGAAATTCAAAGGGGAATTGAGTGCGTCGA
ATTTGCCGCAGGCGCTCCAACATTAATGATGGGGGAGTCGCTATCGGGAA
TTGCGGAAGGTATTGATTCGGAAATGTTCCGTTATCCTTTAGGGGTGGTCG
CCGGAATTACGCCGTTTAATTTTCCGATGATGGTGCCTCTCTGGATGTTTCC
GCTGGCAATTGCTTGCGGGAATACGTTCGTGTTAAAACCATCTGAACGAA
CGCCGATTTTAGCCAATAAACTAGCTGAGTTGTTTACAAAGGCCGGGGCG
CCAGAAGGAGTACTCAATATCGTTCATGGGGCGCATGATGTCGTTAATGC
CCTCATCGATCACGAGGATATTCAAGCCATTTCGTTTGTTGGTTCTCAGCC

TGTCGCCAAATATGTATATGAGCGTGCAGCGGCACGAGGCAAACGGGTAC
AGGCGCTTTCCGGAGCAAAAAACCATCATATCGTTATGCCTGATGCGGAT
ATCGAGACAGCGGTGCAACATGTGATAAGTTCCGCTTTTGGCAGTGCGGG
CCAGCGTTGTATGGCTTGCAGCGCGGTGGTTGTGGTCGGCGACAATGAAC
AGTTTGTTAAACGGTTAAAGCAAAAGGCAGACGAACTTGTTATCGGAAAT
GGCATGGATCCAGAAGTGCTTTTAACTCCGGTGATCCGTCAGTCCCATCGC
GAAAAAGTATTAGGCTACATTCAAAAAGGGATTGAAGAAGGGGCGACAC
TCATTCGCGACGGGCGCAAAGAAATCGAGGAAATGCCAGAAGGCAACTTC
CTTGGTCCGACGATTTTTGATTACGTCACACCAGATATGACGATTGCCAAA
GAAGAAATCTTTGCGCCTGTTCTTAGCTTGCTGCGTGCCAACGATTTAGAC
GAAGCGCTTGAATATATTCGCAAATCTCGATATGGAAACGGTGCGACCAT
TTATACGAAAGACGCGAAAGCGGTCCGGAAATTCCGTGAAGAAGCGGATG
CGGGAATGTTAGGAATCAATGTCGGCGTGCCGGCAACGATGGCGTTCTTC
CCATTCTCCGGCTGGAAGGATTCGTTTTATGGCGACCTCCATGTTAACGGA
AAAGATGGAGTGAATTTCTATACACGCAAAAAAATGATTACTTCTCGCTTC
GATTTTTAA

***mmsA2* (BCV53_11155)**

ATGTCTTCAGCGACTACTGTACAAACATTGAAAAACTTTATAGGTGGACA
ATGGGTAGAATCTTCTTCAGGAAAGGTAGAACTCGTTCCTAATCCGGCTAC
TGGAGAAGCGTTGGCTTATGTTCCAATTTCTTCACGGGAAGAATTAGATCA
AGCGGTTGCAGCAGCAAAAGAAGCATTTAAAACATGGAGCAAAGTGCCTG
TACCGCGCCGCGCCCGCATTTTGTTCCGCTACCAGCAGCTTCTTGTTGAGC
ATTGGGAAGAGTTAGCTCGCCTTGTTACATTAGAAAACGGCAAAAGCTAT
GATGAAGCATATGGTGAAGTTCAGCGTGGGATCGAATGCGTAGAGTTTGC
TGCGGGAGTTCCTACTTTAATGATGGGGCAGCAGCTTCCAGACATTGCGAC
GGGGATCGAATCAGGCATGTACCGCTATCCGCTTGGCGTCGTTGCCGGAA
TTACGCCTTTTAATTTCCCGATGATGGTCCCATGCTGGATGTTTCCTTTGGC
GATTGCCTGCGGAAATACGTTTGTGCTGAAACCGTCTGAACGCACGCCGA
TGTTAGCCAACCGTTTAGCTGAACTATTTACGGAAGCCGGTCTTCCTGCAG
GCGTCCTAAATATTGTTCATGGAGCGCATGATGTCGTTAATGGAATTTTAG
AGCATAAAGATATTAAAGCGGTTTCTTTCGTCGGTTCACAGCCGGTCGCCG
AGTATGTATATAAAACGGCGGCCGCACATGGAAAACGGGTGCAAGCTTTG
GCAGGTGCAAAAAACCATTCCATTGTTATGCCAGATGCTGATTTGGATTAT
GCCGTTACCAACATTATTAATGCCGCGTTTGGCTCGGCAGGGGAACGCTGC
ATGGCATGCGCAGTAGTAGTGGCCGTCGGCGATATTGCGGATGAACTTGT
GGAACGTTTAAAGAAAGCAGCCGATGAGATACAAATCGGCAACGGTTTAC
AGAAGGACGTCTTTTTAGGTCCTGTCATTCGCCAGTCGCATAAAGAACGG
ACGATTAAATATATCGAAATCGGCGAAAAAGAAGGAGCGCTGCTTGTTCG
CGACGGCCGCCGCGATTCTGCTACAAGCGGAGAAGGCTACTTTATCGGGC
CAACGATTTTTGACCATGTGAAGCCGGGCATGACGATTTGGACAGATGAA
ATTTTCGCTCCTGTCTTGTCGGTTGTTCGGGCTCGAGATTTAGATGAAGCC
ATTCAAATTGCCAACAAGTCTGAATTTGCCAACGGCGCATGCATTTACACG
GACAGCGCGAAAGCGATCCGTCAATTCCGTGAAGAAATCGATGCCGGCAT
GCTCGGTGTCAATGTGGCAGTCCCGGCGCCAATGGCGTTTTTCCCATTCTC
TGGTTATAAAAATTCTTTCTATGGAGACCTTCATGCAAATGGCCGTGACGG
CGTTGAATTTTACACGCGCAAAAAAATGGTGACAGCAAGATACTGA

***hbdH* (BCV53_14255)**

ATGAATATCGGATTTATTGGCACAGGTGTGATGGGAGCAAGAATGGCTAG
ACGTCTCCTTGGCGCAGGTTATCATGTCATCGTGTACAATCGTTCGCCCGA
AAAAACAGAAGCGTTAGTACGGCTGGGGGCAGACAGAGCCAGCACGATC
GCAGAACTGGCCCGGCATTCGCATGTCATTTGCACTTGTTTATCGATGCCG
GACGATGTTATCGACGTTTATTTGCGGAAAGAAGGGGTGATCGAAAGCGC
CCGGCCGGGGACGATTTGCCTTGATTTTACGACAGTGGGGCCGGATACGA
GCAAAGCCGTTGCGAAACGGGCCAAAGAGAAGAAAATCGATTTTTTGGAT
GCTCCGGTAAGCGGAGGGCCAGAAGGAGCGGAGCAAGGCACGTTAACGA
TAATGGTCGGCGGAGCGAAATCGGCTTGGGAGCGGGTGGTGCCGCTATTG
AACGTCCTTGGTGAAACGGTGGAGTATTTGGGGCCGAGCGGTTCGGGAAG
TGTGGCAAAACTAATCAATCAATATCTTGTCGCCGTCCATTCTGTTGCTGC
CGCAGAAGCAATGGTCGCTGGGGTGGCGCTTGGGCTGGATGCCAAGCAGC
TTTACCGGATTTTAAAAGCGAGCTACGGAGACAGCCGCATGCTTCGCAGG
CATATGGAACAATATGTTTTTCCGCGCAAGTTTGCAGCAGGCGGAGCAGT
CAAATATGTATATAAAGACGTTTGTTTGGCGAATCGGTTGATGAATGAAA
CGGGGGTGAGGCAATTTTTAGGGCAAATTGCCGAAGAGACGTTCGCGGAT
GCGATGAAACAAGGGCTAGCGGATTTAGATATGTCTGCGATCATTCAGCC
GTTAGAGAAGCGGTGCAATGTGATGGTAACAAAGGAAGAATCATAA

Appendix III: DNA sequences of Isobutanol pathway genes

P_{gapd}

CGAAGAAACGAAACAACAGCTTTGGCAAATTTTCAAAAGATGGAACGCCC
TTTAAAAAGGGCGCTTTTTTTATGATTTTACTGAATTTTTTCTTCACATATC
ACCGAAATATGATATACTATTTTTGTGAAATGCTGTTAATTAGTATAGCAT

P_{ldh}

GCGGGACGGGGAGCTGAGTGCTCCCGTTGTTTGCCGCGGCGTCTGTCATG
AAATGGACAAACAATAGTCAAACAATCGCCACAATCGCGCATGCATTGCG
GTGCGCCTTTCGCGTAAAATATTTATATGAAAGTGTTCGCATTATATTGAG
GGAGGATGAATGCA

kivD (native *L. lactis*)

ATGTATACAGTAGGAGATTACCTATTAGACCGATTACACGAGTTAGGAAT
TGAAGAAATTTTTGGAGTCCCTGGAGACTATAACTTACAATTTTTAGATCA
AATTATTTCCCGCAAGGATATGAAATGGGTCGGAAATGCTAATGAATTAA
ATGCTTCATATATGGCTGATGGCTATGCTCGTACTAAAAAAGTTGCCGCAT
TTCTTACAACCTTTGGAGTAGGTGAATTGAGTGCAGTTAATGGATTAGCAG
GAAGTTACGCCGAAAATTTACCAGTAGTAGAAATAGTGGGATCACCTACA
TCAAAAGTTCAAAATGAAGGAAAATTTGTTCATCATACGCTGGCTGACGG
TGATTTTAAACACTTTATGAAAATGCACGAACCTGTTACAGCAGCTCGAAC
TTTACTGACAGCAGAAAATGCAACCGTTGAAATTGATCGAGTACTTTCTGC
ACTACTAAAAGAAAGAAAACCTGTCTATATCAACTTACCAGTTGATGTTGC
TGCTGCAAAAGCAGAGAAACCCTCACTCCCTTTGAAAAAAGAAAATCCAA
CTTCAAATACAAGTGACCAAGAAATTTTGAACAAAATTCAAGAAAGCTTG
AAAAATGCCAAAAAACCAATCGTGATTACAGGACATGAAATAATTAGTTT
TGGCTTAGAAAAAACAGTCACTCAATTTATTTCAAAGACAAAACTCCCTAT
TACGACATTAAACTTTGGAAAAAGTTCAGTTGATGAAACTCTCCCTTCATT
TTTAGGAATCTATAATGGTAAACTCTCAGAACCTAATCTTAAAGAATTCGT
GGAATCAGCCGACTTCATCCTGATGCTTGGAGTTAAACTCACAGACTCTTC
AACAGGAGCCTTCACTCATCATTTAAATGAAAATAAAATGATTTCACTGA
ATATAAATGAAGGAAAAATATTTAACGAAAGAATCCAAAATTTTGATTTT
GAATCCCTCATCTCCTCTCTCTTAGACCTAAGCGGAATAGAATACAAAGGA
AAATATATCGATAAAAAGCAAGAAGACTTTGTTCCATCAAATGCGCTTTTA
TCACAAGACCGCCTATGGCAAGCAGTTGAAAACCTAACTCAAAGCAATGA
AACAATCGTTGCTGAACAAGGGACATCATTCTTTGGCGCTTCATCAATTTT
CTTAAAACCAAAGAGTCATTTTATTGGTCAACCCTTATGGGGATCAATTGG
ATATACATTCCCAGCAGCATTAGGAAGCCAAATTGCAGATAAAGAAAGCA
GACACCTTTTATTTATTGGTGATGGTTCACTTCAACTTACAGTGCAAGAAT
TAGGATTAGCAATCAGAGAAAAAATTAATCCAATTTGCTTTATTATCAATA
ATGATGGTTATACAGTTGAAAGAGAAATTCATGGACCAAATCAAAGCTAC
AATGATATTCCAATGTGGAATTACTCAAAATTACCAGAATCGTTTGGAGCA
ACAGAAGATCGAGTAGTCTCAAAAATCGTTAGAACTGAAAATGAATTTGT

GTCTGTCATGAAAGAAGCTCAAGCAGATCCAAATAGAATGTACTGGATTG
AGTTAATTTTGGCAAAAGAAGATGCACCAAAAGTACTGAAAAAAATGGGC
AAACTATTTGCTGAACAAAATAAA

***kivD* (LLKF_1386) (Codon optimised)**

ATGTATACGGTAGGCGATTATTTATTAGATCGCTTACATGAATTAGGCATT
GAAGAAATTTTTGGAGTCCCGGGAGATTATAATTTACAATTTTTAGATCAA
ATTATTTCGCGCAAAGATATGAAATGGGTAGGAAATGCGAATGAATTAAA
TGCGTCGTATATGGCGGATGGATATGCGCGCACAAAAAAAGCGGCGGCGT
TTTTAACGACGTTTGGCGTGGGAGAATTATCGGCGGTCAATGGATTAGCG
GGCTCGTATGCGGAAAATTTACCGGTAGTGGAAATTGTTGGAAGCCCGAC
GTCGAAAGTTCAAAATGAAGGCAAATTTGTGCATCATACATTAGCGGATG
GAGATTTTAAACATTTTATGAAAATGCATGAACCGGTAACGGCGGCGCGC
ACGTTATTAACAGCGGAAAATGCGACGGTTGAAATTGATCGCGTGTTAAG
CGCGTTATTAAAAGAACGCAAACCGGTGTATATTAATTTACCGGTTGATGT
TGCGGCGGCGAAAGCGGAAAAACCGTCGTTACCGTTAAAAAAAGAAAAT
CCGACGTCGAATACATCGGATCAAGAAATTTTAAATAAAATTCAAGAATC
GTTAAAAAATGCGAAAAAACCGATTGTGATTACGGGCCATGAAATTATTT
CGTTTGGATTAGAAAATACAGTAACGCAATTTATTAGCAAAACGAAATTA
CCGATTACGACATTAAATTTTGGCAAAAGCAGCGTTGATGAAACATTACC
GTCGTTTTTAGGAATTTATAATGGAAAATTATCGGAACCGAATTTAAAAGA
ATTTGTGGAATCGGCGGATTTTATTTTAATGTTAGGAGTTAAATTAACGGA
TTCGTCGACGGGAGCGTTTACGCATCATTTAAATGAAAATAAAATGATTTC
GTTAAATATTGATGAAGGCAAAATTTTTAATGAATCGATTCAAAATTTTGA
TTTTGAATCGTTAATTTCGTCGTTATTAGATTTATCGGGAATTGAATATAA
AGGAAAATATATTGATAAAAAACAAGAAGATTTTGTGCCGTCGAATGCGT
TATTAAGCCAAGATCGCTTATGGCAAGCGGTCGAAAATTTAACACAAAGC
AATGAAACAATTGTAGCGGAACAAGGCACATCGTTTTTTGGCGCGAGCAG
CATTTTTTTAAAACCGAAAAGCCATTTTATTGGACAACCGTTATGGGGAAG
CATTGGCTATACATTTCCGGCGGCGTTAGGCAGCCAAATTGCGGATAAAG
AAAGCCGCCATTTATTATTTATTGGCGATGGCAGCTTACAATTAACAGTGC
AAGAATTAGGCTTAGCGATTCGCGAAAAAATTAATCCGATTTGTTTTATTA
TTAATAATGATGGCTATACAGTTGAACGCGAAATTCATGGCCCGAATCAA
AGCTATAATGATATTCCGATGTGGAATTATAGCAAATTACCGGAATCGTTT
GGCGCGACAGAAGAACGCGTCGTAAGCAAAATTGTACGCACAGAAAATG
AATTTGTCAGCGTCATGAAAGAAGCGCAAGCGGATCCGAATCGCATGTAT
TGGATTGAATTAGTCTTAGCGAAAGAAGATGCGCCGAAAGTTTTAAAAAA
AATGGGCAAATTATTTGCGGAACAAAATAAAAGCTAA

***kivD* (LLKF[Mut.]) (Codon optimised)** [LLKF_1386 with 20 bp swapped with wild-type sequence in yellow]

ATGTATACGGTAGGCGATTATTTATTAGATCGCTTACATGAATTAGGCATT
GAAGAAATTTTTGGAGTCCCGGGAGATTATAATTTACAATTTTTAGATCAA
ATTATTTCGCGCAAAGATATGAAATGGGTAGGAAATGCGAATGAATTAAA
TGCGTCGTATATGGCGGATGGATATGCGCGCACAAAAAAAGCGGCGGCGT

TTTTAACGACGTTTGGCGTGGGAGAATTATCGGCGGTCAATGGATTAGCG
GGCTCGTATGCGGAAAATTTACCGGTAGTGGAAATTGTTGGAAGCCCGAC
GTCGAAAGTTCAAAATGAAGGCAAATTTGTGCATCATACATTAGCGGATG
GAGATTTTAAACATTTTATGAAAATGCATGAACCGGTAACGGCGGCGCGC
ACGTTATTAACAGCGGAAAATGCGACGGTTGAAATTGATCGCGTGTTAAG
CGCGTTATTAAAAGAACGCAAACCGGTGTATATTAATTTACCGGTTGATGT
TGCGGCGGCGAAAGCGGAAAAACCGTCGTTACCGTTAAAAAAAGAAAAT
CCGACGTCGAATACATCGGATCAAGAAATTTTAAATAAAATTCAAGAATC
GTTAAAAAATGCGAAAAAACCGATTGTGATTACGGGCCATGAAATTATTT
CGTTTGGATTAGAAAATACAGTAACGCAATTTATTAGCAAAACGAAATTA
CCGATTACGACATTAAATTTTGGCAAAAGCAGCGTTGATGAAACATTACC
GTCGTTTTTAGGAATTTATAATGGAAAATTATCGGAACCGAATTTAAAAGA
ATTTGTGGAATCGGCGGATTTTATTTTAATGTTAGGAGTTAAATTAACGGA
TTCGTCGACGGGAGCGTTTACGCATCATTTAAATGAAAATAAAATGATTTC
GTTAAATATTGATGAAGGCAAAATTTTTAATGAATCGATTCAAAATTTTGA
TTTTGAATCGTTAATTTCGTCGTTATTAGATTTATCGGGAATTGAATATAA
AGGAAAATATATTGATAAAAAACAAGAAGATTTTGTGCCGTCGAATGCGT
TATTAAGCCAAGATCGCTTATGGCAAGCGGTCGAAAATTTAACACAAAGC
AATGAAACAATTGTAGCGGAACAAGGCACATCGTTTTTTGGCGCGAGCAG
CATTTTTTTAAAACCGAAAAGCCATTTTATTGGACAACCGTTATGGGGAAG
CATTGGCTATACATTTCCGGCGGCGTTAGGCAGCCAAATTGCGGATAAAG
AAAGCCGCCATTTATTATTTATTGGCGATGGCAGCTTACAATTAACAGTGC
AAGAATTAGGCTTAGCGATTCGCGAAAAAATTAATCCGATTTGTTTTATTA
TTAATAATGATGGCTATACAGTTGAACGCGAAATTCATGGCCCGAATCAA
AGCTATAATGATATTCCGATGTGGAATTATAGCAAATTACAGAATCGTTTG
GAGCAACAGAAAGAACGCGTCGTAAGCAAAATTGTACGCACAGAAAATG
AATTTGTCAGCGTCATGAAAGAAGCGCAAGCGGATCCGAATCGCATGTAT
TGGATTGAATTAGTCTTAGCGAAAGAAGATGCGCCGAAAGTTTTAAAAAA
AATGGGCAAATTATTTGCGGAACAAAATAAAAGCTAA

***kivD* (LLM4) (Codon optimised)**

ATGTACACGGTTGGCGACTACTTGTTGGACCGCTTACACGAGTTGGGAATC
GAAGAAATCTTCGGCGTGCCGGGCGACTATAACTTGCAGTTTTTGGACCAC
ATCATCAGCCGCAAGGATATGAAGTGGGTGGGAAACGCTAATGAATTAAA
CGCGTCGTACATGGCTGACGGCTATGCGAGAACGAAAAAGGTGGCGGCGT
TCTTGACAACGTTCGGCGTGGGCGAGTTAAGCGCGGTGAACGGATTAGCG
GGCTCGTATGCGGAGAATTTGCCAGTGGTGGAGATTGTTGGAAGCCCAAC
AAGCAAGGTGCAGAACGAAGGCAAGTTTGTGCACCACACGTTGGCGGACG
GCGACTTCAAACATTTCATGAAGATGCATGAGCCGATTACGGCGGCTAGA
ACGTTGTTAACGGCGGAAAACGCGACAGTGGAGATTGACCGCGTGTTGTC
GGCTTTGTTAAAAGAGCGCAAACCGGTGTACATCAACTTACCAGTGGACG
TGGCTGCGGCTAAAGCTGAGAAACCGTCGTTGCCGTTGAAGAAGGAAAAC
CCGACGTCGAACACGAGCGACCAAGAAATCTTAAATAAAATTCAAGAATC
GTTGAAAAACGCGAAAAAACCGATCGTGATTACGGGCCACGAGATTATCA
GCTTCGGATTGGAGAAGACAGTGACGCAATTTATTTCGAAAACAAAATTA
CCGATCACAACGTTGAACTTTGGCAAGTCGTCGGTGGACGAAACGTTGCC
GAGCTTCTTAGGAATCTACAACGGCAAGTTATCGGAGCCGAACTTAAAGG
AGTTCGTTGAGAGCGCTGATTTTATTTTGATGTTAGGAGTGAAGTTGACAG

ATTCGTCGACGGGAGTGTTCACGCATCACTTGAATGAGAACAAAATGATC
AGCTTGAACATTAACGAAGGCAAAATTTTCAATGAACGCATTCAGAATTT
CGATTTTGAAAGCTTAATTTCGTCGTTGTTGGATTTGTCGGGCATCGAGTA
TAAGGGCAAGTACATCGACAAGAAGCAAGAAGACTTCGTGCCGTCGAATG
CGTTATTGAGCCAAGATAGATTGTGGCAAGCGGTGGAGAATTTGACGCAG
TCGAACGAGACGATTGTTGCGGAACAAGGCACGTCGTTTTTCGGCGCGAG
CCCGATCTTTTTAAAGCCGAAGTCGCATTTCATCGGACAGCCATTGTGGGG
ATCGATTGGCTATACGTTCCCGGCTGCGTTAGGCTCGCAGATCGCTGATAA
AGAATCGCGCCACTTGTTATTCATTGGCGACGGCTCGTTGCAATTGACAGT
TCAAGAATTGGGCTTGGCGATTCGCGAGAAGATCAATCCGATCTGCTTCAT
CATCAACAACGACGGCTACACGGTGGAAAGAGAAATCCACGGACCGAAC
CAAAGCTACAACGACATCCCGATGTGGAACTACTCGAAGTTGCCAGAATC
GTTTGGAGCAACAGAAGACCGCGTGGTGAGCAAGATCGTGCGCACAGAG
AACGAGTTTGTGAGCGTGATGAAGGAAGCGCAAGCGGACCCGAACAGAA
TGTATTGGATCGAGTTAATTTTGGCGAAGGAGGACGCTCCGAAGGTTTTGA
AGAAAATGGGCAAATTGTTCGCTGAACAAAATAAAAGCTAA

***kivD* (WT_LLM4) (Codon optimised)** [LLKF sequence containing four amino acid mutations (Q34H, V130I, A290V, S386P) of LLM4 in blue]

ATGTATACAGTAGGAGATTACCTATTAGACCGATTACACGAGTTAGGAAT
TGAAGAAATTTTTGGAGTCCCTGGAGACTATAACTTACAATTTTTAGATCA
CATTATTTCCCGCAAGGATATGAAATGGGTCGGAAATGCTAATGAATTAA
ATGCTTCATATATGGCTGATGGCTATGCTCGTACTAAAAAAGTTGCCGCAT
TTCTTACAACCTTTGGAGTAGGTGAATTGAGTGCAGTTAATGGATTAGCAG
GAAGTTACGCCGAAAATTTACCAGTAGTAGAAATAGTGGGATCACCTACA
TCAAAAGTTCAAAATGAAGGAAAATTTGTTCATCATACGCTGGCTGACGG
TGATTTTAAACACTTTATGAAAATGCACGAACCTATTACAGCAGCTCGAAC
TTTACTGACAGCAGAAAATGCAACCGTTGAAATTGATCGAGTACTTTCTGC
ACTACTAAAAGAAAGAAAACCTGTCTATATCAACTTACCAGTTGATGTTGC
TGCTGCAAAAGCAGAGAAACCCTCACTCCCTTTGAAAAAAGAAAATCCAA
CTTCAAATACAAGTGACCAAGAAATTTTGAACAAAATTCAAGAAAGCTTG
AAAAATGCCAAAAAACCAATCGTGATTACAGGACATGAAATAATTAGTTT
TGGCTTAGAAAAAACAGTCACTCAATTTATTTCAAAGACAAAACTCCCTAT
TACGACATTAAACTTTGGAAAAAGTTCAGTTGATGAAACTCTCCCTTCATT
TTTAGGAATCTATAATGGTAAACTCTCAGAACCTAATCTTAAAGAATTCGT
GGAATCAGCCGACTTCATCCTGATGCTTGGAGTTAAACTCACAGACTCTTC
AACAGGAGTGTTCACTCATCATTTAAATGAAAATAAAATGATTTCACTGA
ATATAAATGAAGGAAAAATATTTAACGAAAGAATCCAAAATTTTGATTTT
GAATCCCTCATCTCCTCTCTCTTAGACCTAAGCGGAATAGAATACAAAGGA
AAATATATCGATAAAAAGCAAGAAGACTTTGTTCCATCAAATGCGCTTTTA
TCACAAGACCGCCTATGGCAAGCAGTTGAAAACCTAACTCAAAGCAATGA
AACAATCGTTGCTGAACAAGGGACATCATTCTTTGGCGCTTCACCGATTTT
CTTAAAACCAAAGAGTCATTTTATTGGTCAACCCTTATGGGGATCAATTGG
ATATACATTCCCAGCAGCATTAGGAAGCCAAATTGCAGATAAAGAAAGCA
GACACCTTTTATTTATTGGTGATGGTTCACTTCAACTTACAGTGCAAGAAT
TAGGATTAGCAATCAGAGAAAAAATTAATCCAATTTGCTTTATTATCAATA
ATGATGGTTATACAGTTGAAAGAGAAATTCATGGACCAAATCAAAGCTAC

AATGATATTCCAATGTGGAATTACTCAAAATTACCAGAATCGTTTGGAGCA
ACAGAAGATCGAGTAGTCTCAAAAATCGTTAGAACTGAAAATGAATTTGT
GTCTGTCATGAAAGAAGCTCAAGCAGATCCAAATAGAATGTACTGGATTG
AGTTAATTTTGGCAAAAGAAGATGCACCAAAAGTACTGAAAAAAATGGGC
AAACTATTTGCTGAACAAAATAAAAGCTAA

B. subtilis alsS **(Codon optimised)**

ATGACGAAAGCGACGAAAGAACAAAAAAGCCTTGTCAAAAATCGCGGAG
CGGAATTAGTAGTAGATTGTTTAGTAGAACAAGGCGTTACACATGTGTTTG
GCATTCCGGGAGCGAAAATTGATGCGGTATTTGATGCGTTACAAGATAAA
GGACCGGAAATTATTGTTGCGCGCCATGAACAAAATGCGGCGTTTATGGC
GCAAGCGGTGGGCCGCTTAACGGGAAAACCGGGAGTGGTTTTAGTTACGA
GCGGACCGGGCGCGTCGAATTTAGCGACGGGATTATTAACAGCGAATACA
GAAGGAGATCCGGTAGTGGCGTTAGCGGGAAATGTGATTCGCGCGGATCG
CTTAAAACGCACACATCAAAGCCTTGATAATGCGGCGTTATTTCAACCGAT
TACGAAATATTCGGTCGAAGTACAAGATGTTAAAAATATTCCGGAAGCGG
TCACAAATGCGTTTCGCATTGCGAGCGCGGGCCAAGCGGGAGCGGCGTTT
GTTAGCTTTCCGCAAGATGTTGTCAATGAAGTGACAAATACGAAAAATGT
CCGCGCGGTTGCGGCGCCGAAATTAGGACCGGCGGCGGATGATGCGATTT
CGGCGGCGATTGCGAAAATTCAAACGGCGAAATTACCGGTGGTTTTAGTG
GGCATGAAAGGAGGCCGCCCGGAAGCGATTAAAGCGGTACGCAAATTATT
AAAAAAAGTGCAATTACCGTTTGTTGAAACATATCAAGCGGCGGGAACAT
TAAGCCGCGATTTAGAAGATCAATATTTTGGCCGCATTGGATTATTTCGCA
ATCAACCGGGCGATTTATTATTAGAACAAGCGGATGTCGTGTTAACGATTG
GCTATGATCCGATTGAATATGATCCGAAATTTTGGAATATTAATGGAGATC
GCACGATTATTCATTTAGATGAAATTATTGCGGATATTGATCATGCGTATC
AACCGGATTTAGAATTAATTGGCGATATTCCGTCGACGATTAATCATATTG
AACATGATGCGGTTAAAGTGGAATTTGCGGAACGCGAACAAAAAATTTTA
AGCGATTTAAAACAATATATGCATGAAGGCGAACAAGTTCCGGCGGATTG
GAAATCGGATCGCGCGCATCCGTTAGAAATTGTGAAAGAATTACGCAATG
CGGTGGATGATCATGTTACGGTAACATGCGATATTGGATCGCATGCGATTT
GGATGTCGCGCTATTTTCGCAGCTATGAACCGTTAACGTTAATGATTAGCA
ATGGAATGCAAACGTTAGGCGTAGCGTTACCGTGGGCGATTGGCGCGTCG
TTAGTTAAACCGGGAGAAAAAGTGGTATCGGTAAGCGGCGATGGAGGCTT
TTTATTTTCGGCGATGGAATTAGAAACAGCGGTTCGCTTAAAAGCGCCGAT
TGTGCATATTGTATGGAATGATAGCACATATGATATGGTAGCGTTTCAACA
ATTAAAAAAATATAATCGCACAAGCGCGGTCGATTTTGGCAATATTGATA
TTGTTAAATATGCGGAATCGTTTGGCGCGACAGGCTTACGCGTCGAATCGC
CGGATCAATTAGCGGATGTCTTACGCCAAGGCATGAATGCGGAAGGCCCG
GTGATTATTGATGTCCCGGTCGATTATTCGGATAATATTAATTTAGCGTCG
GATAAATTACCGAAAGAATTTGGCGAATTAATGAAAACAAAAGCGTTATA
A

B. coagulans alsS **(Codon optimised)**

ATGGGAGTAGGCACAGTCGAGAAAAAGAATAGCAATCCTACGAATACGA
CGGAAAAAACAGCGGCGGATCTTGTGGTAGATTGCCTTGAAAAACAAGAA
GTGCCGTATGTGTTTGGCATTCCGGGAGCGAAAATTGATGCGGTGTTTGAT
GTACTTAAAGAACGGGGACCGGAACTTATTGTGTGTAGACATGAACAAAA
TGCGGCGTTTATGGCGGCGGCGATTGGCCGCCTTACAGGCAAACCTGGGG
TCTGCTTAGTCACATCGGGCCCAGGCGCGTCCAATCTTGCGACAGGACTTG
CGACAGCGAATACAGAATGTGATCCTGTGGTCGCGATTGCGGGCAATGTC
CCTCGTGCGGATCGCTTGAAGAAAACTCATCAAAGTATGGATAATGTCTC
GTTGTTTCAACCTATTACGAAATATGCGGCGGAAGTGGTGCATCCTGATAC
AGTGCCGGAAGTAATGACGAATGCGTTTCGCTCCGCGGCGTCCGCGCAAG
CGGGAGCGGCGTTTATTAGCTTTCCACAAGATGTACTTAAAGAACCTGCGT
CGGTTAAAGCGTTGGGCCCTCTTAAATCCCCTAAACTTGGCAAAGCGAAC
GAAGAAGCGGTCAAAGAAGCGGTCAAAGCGATTCAACATGCGAAATTAC
CAGTCATTTTGGTAGGCATGAGAGCGAGCAGACCGGAAGTGGTCAAAGCG
GTACGCTCCCTTCTTAAAAAGATTGCGCTTCCAGTAGTGGAAACGTTTCAA
GCGGCGGGCCTTATTAGCCGTGATCTTGAAGATCGCTTCTTTGGACGCATT
GGCTTGTTTCGTAATCAACCGGGAGATATTCTTCTTGAACATGCGGATGTA
GTGTTAGCGATTGGCTATGATTCGGTAGAATATGATCCTAAATTTTGGAAC
TCCGAAGGTGAACGCAAGATTATTCATCTCGATGAAATTCGTGCGGATATT
GATCATGATTATCAACCTGAAATTGAACTTGTGGGGGATATTTCCGCGTCC
GTCGATAGCATTAAAGAACAACTTGCGCGCCTTAATATGAATGGCAAATC
GATGGAACTTCTTGAACGGTTACGCAATCAACTTAATTTGAGAGATGAAC
CGTCCGAAAAAGCGGATAAAAATCTTGTACATCCTTTACAGTTTATTCATG
ATCTTCGCTCCCTTATTGATGATCATGTGACAGTGACGTGTGATGTGGGCA
GCCATTATATTTGGATGGCACGCCATTTTCGGGTGTATGAACCTAATCGGT
TGTTGTTTAGCAATGGTATGCAAACGTTGGGAGTCGCGTTACCGTGGGCGA
TTGCGGCGACGTTAGTTAATCCTGGAGAAAAAGTAGTCTCCATTTCGGGA
GATGGAGGCTTTTTGTTTTCCGCGATGGAATTAGAAACGGCGGTCAGACTT
AAATCGCCGTTGGTTCATATTGTCTGGAGAGATGGAACTTATGATATGGTC
GCGTTTCAACAACAGATGAAATATGGCCGCACATCGGGCGCGGATTTTGG
ACCTGTCGATATTGTCAAACATGCGGAATCGTATGGAGCGAAAGGCTTGA
GAGTTAATAGCCCAGATGAATTAGTCTCGGTGCTTAAAGAAGCGTTAGAT
TCCGAAGGACCTGTCGTGGTAGATGTACCAGTCGATTATTCCGATAACCTG
GAACTTGCGAAAAAACTTTTACCTAATCAATTAGTTTAA

T. daqus alsS **(Codon optimised)**

ATGAATATTGCGGAAGAAATGGTTCAATCCCTTGAAGAAAAAGGCGTACG
GTATTTGTTTGGAGTACCGGGAGAAGAAAATATTGATTTTCTTGAAGCGGT
ACGAAATAGCTCCATTCAATTCATTCTTACACGCCATGAACAAACAGCGG
CGATGATGGCGGCGATGGTAGGCCGGTTGACAGGCATTCCTGGAGTCTGC
TTATCGACGTTGGGACCGGGAGCGGCGAATCTTGCGACGGGGGTCGCGAC
AGCGAATGCGGATGGACTTCCGTTAATTGCGATTGCGGGCCAAGCGGAAA
CGAATCGCCTTCATCATCCGAGCCATCAAATGGTTGACCTTCCTGAGTTCT
TCCGTCCTATTACGAAATGGTCCACGACAATTCGTCATAGCTCCGCGCCTC
GTGAAACGATTGAAAAAGCGTTTAATCTTGCGATGTCCGAACGCCCAGGC
GCGGTGTTTATTGAACTTCCTGAGGATCTTGCGGTCAAAGAAGGGGATTCG

GTGACAGGCATTGCGTCCGCGTTAGGACGGGTCGAAGCGGCGGAAATTGA
TATGAAACGTGCGGCGGAACAAATTAATGCGGCGGAAATGCCGCTTATTT
TAGCGGGCGCGGGCACACTTCGGGGCAAAGCGAGCAGAGAAGTCGCGGC
GTTGTCCGAAAAACTTTGTGCGTCCGTGTTAGAAACGTTTATGGGCAAAGG
CGCGGTGAGCTGGCGCTTGGATACACATCTCGCGGCGGCGGGGACTTGGG
GACGGGATTGGGCGGATGAAGCGATTGATAAAGCGGACCTCGTGTTGGCG
ATTGGCTATGATTATATTGAATATAGCCCAGTACGCTGGAACAAAGAACG
CAAACAAGTAGTTCATATTAATTGGCAAGAAGCGGAAGTCGATGCGCATT
ATCCAGTGGTCGCGTCCCTTATTGGTGACCTTTCGTGTAATCTTGCGAGAC
TTACGGAACTTGTCAGAGAACGCACAGCGTTACCTGCGCCGTTTGCGGAA
CTTCGCCGTAAGATTCGTCAAGAATGGGAAGAAGCGGGCCAAGATACAGC
GTTTCCACTTAAACCTAGCAAAATTTTGTATGATCTTCGTGCGGCGATGAA
TGATGATGATCTTGTAGTATGTGATGTAGGCGCGCATAAAGTCTGGTTAGG
CCGCATGTTTCCTGTCTTTGCGCCGGATACATGCTTTATTTCCAATGGCCTT
GCGACAATGGGAGTCGCGTTACCAGGCGCGATTGCGGCGAAATTAGTCAA
ACCTGAACGCAGAGTTGTCGCGGTCTGTGGAGATGGAGGCTTTCTTATGA
ATGTTCAAGAATTAGAAACAGCGGCACGCCTTAAATTACCGCTTGTCATTC
TTTTGTGGAGAGATGGAGGCTATGGCCTTATTGAGTGGAAACAACTTGAA
AAATTCAGACATAGCCATCATGTGACGTTTGGCAATCCAGATTTTGTACAT
CTAGCGCGCTCCTTTGGCATTGAAGGATGTAGAGTCGAAAAGGGCGAAGA
ACTTAGACCACTTTTGGAAAAAGCGTTACATGCGGAAGGACCAGTGTTAA
TTGATTGCCCAGTAGATTATCGGGAAAATCTTAAATTAGAAGAAAAAATT
CGCTTTGGCTGTGGAGATGAAACACGGTAA

Native *ilvC* (*P. thermoglucosidasius* NCIMB 11955) (BCV53_17785)

ATGGCAAAAGTTTACTATAACGGAGATGCAAATGAAAAATATTTACAAGG
AAAAACAGTGGCAATCATCGGTTACGGCTCACAAGGACACGCGCATGCGC
AAAACCTCCGCGACAGTGGAGTGAATGTGATTGTTGGGCTCCGGAAAGGG
AAATCATGGGAAAAAGCGGAACAAGACGGTTTTGCCGTATATAGTGTCCG
CGAAGCAGCAAAACAGGCGGATATCGTCATGATTCTTCTCCCGGATGAAA
AACAGCCGAATGTATACAAAGAAGAAATCGAACCAGAGCTTCAGCCGGG
GAACGCGTTAGTATTTGCCCATGGATTTAATATTCACTTTCATCAAATCGT
TCCTCCTGAACATGTCGATGTATTTTTAGTCGCACCAAAAGGCCCTGGACA
TTTAGTGCGCCGCACGTATGTGGAAGGAGCCGGCGTGCCGGCGTTAATTG
CGGTGCAGCAAGATGTGACGGGGGCGGCAAAAGAAACGGCGCTCGCGTA
CGCAAAAGCGATCGGCGCGACAAGAGCGGGTGTGCTTGAAACAACGTTTA
AAGAAGAGACAGAAACGGACTTATTTGGCGAGCAGGCAGTATTATGCGGA
GGCTTAACGTCGCTTATTAAAGCGGGGTTTGAAACGCTTGTGGAAGCAGG
GTATCAGCCGGAATTGGCCTATTTTGAATGTTTGCATGAAATGAAGCTGAT
CGTCGACCTTCTGTACGAAGGCGGGCTTTCTTGGATGCGCCATTCCATCTC
TGATACGGCGCAATGGGGAGATTTTATTTCTGGGCCGCGCATCATTAATGA
CGCAGTGAAAGCAGAGATGAAAAAAGTACTTCATGATATTCAGACAGGAA
AATTTGCAAAAAGCTGGATTTTGGAAAACCAGGCGAACCGCCCAGAGTTT
AATGCGATTAATCGCCGCGAAAATGAGCATCTCATTGAAATTGTCGGACG
CGAATTGCGCAGCATGATGCCATTTGTGAATGAAAAACAGAAAGAAGTGG
TGGTTACGAGTGAGAAAAATTAA

Native *ilvD* (*P. thermoglucosidasius* NCIMB 11955) (BCV53_15070)

ATGGAGGGATTGAGATTGGGAAAACTTCGCAGCGATATGATCAAAAAAGG
ATTTGACCGGGCGCCGCACCGCAGTTTGCTGCGCGCGGCAGGCGTTAAGG
AAGAAGATTTTGACAAGCCGTTTATTGCCGTGGTGAACTCATATATTGATA
TTATTCCGGGGCATGTCCATTTGCAGGAGTTCGGCAAAATTGTCAAAGAA
GCGATCCGCGAAGCGGGCGGAGTCCCGTTTGAAATGAATACGATCGGCGT
TGACGACGGAATCGCGATGGGGCATATCGGCATGCGCTATTCGCTGCCAA
GCCGCGAAATTATCGCAGATTCGATCGAAACGGTGGTTTCCGCGCATTGGT
TTGACGGAATGGTATGCATTCCAAATTGTGACAAAATAACGCCGGGCATG
ATGATGGCGGCGATGCGTCTGAATATCCCGACGATTTTCGTCAGCGGCGG
GCCGATGAAAGCTGGCGTGACGAGCGATGGAAGAAAAATTTCGCTTTCTT
CCGTATTTGAAGGGGTTGGCGCTTATCAAGCCGGAAAAATCGATGAAAAA
GGCTTAATGGAATTAGAACGATATGGCTGCCCGACATGCGGTTCCTGTTCG
GGAATGTTTACGGCGAACTCGATGAACTGTTTGGCGGAAGCGTTAGGGCT
CGCATTGCCAGGAAACGGCACGATTTTAGCGGTAGACCCGGCCCGGAAAG
AGCTCGTCCGCCAATCGGCAAAACAATTAATGTATTTGATCGAGCATGAC
ATTAAGCCGCGCGATATTGTCACCGAAAAGGCGATCGACAACGCGTTTGC
GCTCGACATGGCGCTTGGCGGCTCGACGAACACCGTATTACATACGCTGG
CGATTGCGAACGAAGCGGGCATCGATTATTCGCTGGAACGGATTAACGAA
ATCGCGGCAAAGGTGCCGCATCTTGCCAAACTGGCGCCAGCTTCCGATGT
GCATATTGAAGACTTGCATGAAGCGGGCGGCGTATCGGCAGTGTTGCATG
AGTTAGCGAAAAAAGAAGGAACGCTGCATCTCGATACATTGACGGTAACC
GGAAAAACACTTGGAGAAAACATCGATGGCTGTGAAGTCAAAGACTATAA
CGTCATCCGCCCAATTGATAACCCGTATTCTGAAACGGGCGGTCTTGCCGT
ACTGTTTGGAAACCTCGCCCCGGACGGAGCGATCATAAAAACAGGCGGCG
TGCAAGCCGGCATTACGCGCCATGAAGGGCCAGCGATCGTTTTTGATTCGC
AGGAAGAGGCGCTGGAAGGAATCGCAAACGGCAAAATCAAGCCAGGCCA
TGTCGTCGTCATCCGTTATGAAGGGCCAAAAGGAGGACCGGGAATGCCGG
AAATGCTCGCGCCAACTTCGCAAATTGTCGGCATGGGGCTTGGCACGAAA
GTTGCGCTCGTGACGGATGGGCGGTTCTCCGGGGCGTCGCGCGGCATATC
GATAGGACACGTTTCCCCAGAAGCGGCGGAAGGCGGCCCGATTGCTTTTA
TTGAAGACGGAGACATCATTGAGATCGATATTAAAAACAGAACGATTAAC
GCAAAGCTTTCTGATGAAGAATGGGAAAAACGAAAAGCGAACTGGAAAG
GATTTGAACCGAAAGTGAAAACCGGATACCTTGCGCGCTATTCGAAGCTC
GTTACTTCCGCAAGCACGGGCGGGATTATGAAAATTTAA

Native *adhA* (*P. thermoglucosidasius* NCIMB 11955) (BCV53_06450)

CAGGGAGGAACTTAGATGAAAGCACTTACATACCTAGGGCCAGGAAAAA
AAGAATTAATGGAAAAACCAAAGCCAAAAATTGAAAAGGAAACCGATGC
AATCGTCAAAATAATAAAAACAACGATTTGTGGAACGGATTTGCACATTC
TTTCAGGAGATGTTCCTACTGTTGAAGAAGGGCGGATTTTAGGACACGAA
GGCGTCGGAATTATTGAAGAAGTTGGTTCGGCCGTAAAGAATTTTAAAAA
AGGCGACAGAGTGTTGATTTCTTGCATTACCTCTTGTGGAAAATGCGAAAA
TTGCAAGAAAGGGTTATACGCCCATTGCGAAGATGGCGGCTGGATCTTGG
GCCACTTAATTGATGGAACTCAAGCAGAATATGTAAGAATTCCGCACGCA
GACAACAGCCTTTATCCTATTCCGGAAGGCGTGGATGAAGAGACTCTTGTC
ATGCTTAGTGACATTCTTCCAACAGGATTTGAAATCGGCGTGTTGAACGGC

AAAGTTCAGCCTGGACAAACCGTCGCCATTATCGGAGCTGGTCCCGTAGG
TATGGCAGCGCTATTAACAGCCCAATTTTATTCACCAGCAGAGATCATTAT
GGTTGATTTAGACGATAACCGTTTAGAAGTTGCGAAAAAATTTGGCGCGA
CCCAAGTGGTGAATAGCGCTGATGGCAAGGCAGTGGAAAAAATTATGGAA
TTAACCGGCGGGAAAGGTGTAGACGTCGCGATGGAAGCCGTCGGAATTCC
GGCAACATTTGATATTTGTCAAGAAATTGTCAAACCAGGTGGCTATATCGC
CAATATCGGTGTTCATGGAAAAAGCGTGGAATTTCACATTGAAAAATTAT
GGATACGCAACATTACGTTGACAACCGGTCTTGTCAACACGACTTCTACGC
CGATGTTATTAAAAACGGTGCAGTCGAAAAAATTGAAGCCGGAACAATTA
ATTACCCATCGTTTCGCCTTTTCAGACATTATGAAAGCGTATGAAGTATTT
GGAAATGCAGCAAAAGAAAAAGCGTTAAAAGTCATTATTTCCAACAGTTA
A

Appendix IV: Unique SNVs and Indels of TM89 and TM242 strains compared to NCIMB 11955 (Sheng *et al.*, 2017)

Position	Strains	SNP	Gene	Locus	Effect
1037973	TM89	A>G	L-rhamnose Isomerase	BCV53_05120	–
1466387	TM89	C>T	N-acetyl-gamma-glutamyl-phosphate reductase	BCV53_07260	H210Y
1819652	TM89	T>C	Transcriptional repressor codY	BCV53_09070	V143A
1842701	TM89	A>T	Chemotaxis Protein CheA	BCV53_09205	N322V
2844365	TM89	T>C	Cytosolic protein	BCV53_13935	V1A
3210549	TM89	T>–	30S ribosomal protein S1	BCV53_15740	N65fs
3417349	TM89	A>G	Hypothetical protein	BCV53_16880	–
3515110	TM89	C>T	Adenine phosphoribosyltransferase	BCV53_17395	D119N
3515272	TM89	T>G	Adenine phosphoribosyltransferase	BCV53_17395	T65P
34296*	TM242	A>–	–		–
2006731	TM242	G>A	Type III restriction modification system methylation subunit	BCV53_09980	E266K

Appendix V: Strains and SNPs verified

Position	SNPs verified	Locus	Strains
26985*	A>G	BCV53_19285	AM242_5_2, AM242_5_3
405673	T>-	BCV53_01955	AM242_4_3
609276	G>A	BCV53_03070	AM242_4_2, AM242_6_4
856889	C>T	BCV53_04235	AM242_1_3, AM242_3_1
1084854	A>T	BCV53_05335	AM242_4_6, AM242_6_2
1086291	T>C	BCV53_05335	AM242_3_4
1367189	A>G	BCV53_06805	AM242_4_3, AM242_5_1
1415815	A>T	BCV53_07015	AM242_1_1, AM242_1_2
1471195	G>A	BCV53_07290	AM242_1_1, AM242_1_2
1514453	G>A	BCV53_07525	AM242_1_6, AM242_3_1
1541623	C>T	BCV53_07715	AM242_4_3
1955303	C>G	BCV53_09715	AM242_2_2
1986844	G>A	-	AM242_5_1
2063731	C>T	BCV53_10245	AM242_5_2
2169641	C>T	BCV53_10705	AM242_1_1, AM242_1_2
2242729	T>C	-	AM242_4_2, AM242_6_4
2323810	G>A	BCV53_11450	AM242_1_2
2360630	C>T	BCV53_11600	AM242_5_1, AM242_5_5
2570676	A>G	BCV53_12610	AM242_1_1, AM242_1_2
2638542	C>T	BCV53_12920	AM242_4_4, AM242_6_1
2657967	G>A	BCV53_13005	AM242_1_3, AM242_3_1
2791650	C>T	BCV53_13695	AM242_5_1, AM242_5_2
2891366	G>A	BCV53_14215	AM242_2_3
2916351	G>A	BCV53_14315	AM242_4_3
3015609	C>T	BCV53_14775	AM242_3_4

3113091	G>A	BCV53_15235	AM242_1_3, AM242_3_2, AM242_4_5, AM242_5_2
3219412	G>A	BCV53_15805	AM242_4_6, AM242_6_2
3228940	T>C	-	AM242_1_1, AM242_1_2
3411590	G>A	BCV53_16860	AM242_1_1, AM242_1_2
3523149	C>T	BCV53_17445	AM242_5_2

Appendix VI: List of SNVs and Indels per individual strain compared to NCIMB 11955

Position	Strain	SNP	Gene	Locus	Effect
1415815	AM242_1_1	A>T	AraC family transcriptional regulator	BCV53_07015	Q158L
1471195	AM242_1_1	G>A	Carbamoyl phosphate synthase large subunit	BCV53_07290	V152I
2169641	AM242_1_1	C>T	Acyl-CoA dehydrogenase	BCV53_10705	A196T
2570676	AM242_1_1	A>G	Acyl-CoA dehydrogenase	BCV53_12610	M208V
3228940	AM242_1_1	T>C	-	-	-
3411590	AM242_1_1	G>A	4-hydroxy-3-methylbut-2-en-1-yl diphosphatesynthase	BCV53_16860	E114K
1415815	AM242_1_2	A>T	AraC family transcriptional regulator	BCV53_07015	Q158L
1471195	AM242_1_2	G>A	Carbamoyl phosphate synthase large subunit	BCV53_07290	V152I
2169641	AM242_1_2	C>T	Acyl-CoA dehydrogenase	BCV53_10705	A196T
2323810	AM242_1_2	G>A	Histidine kinase	BCV53_11450	G566E
2570676	AM242_1_2	A>G	Acyl-CoA dehydrogenase	BCV53_12610	M208V
3228940	AM242_1_2	T>C	-	-	-
3411590	AM242_1_2	G>A	4-hydroxy-3-methylbut-2-en-1-yl diphosphatesynthase	BCV53_16860	E114K
856889	AM242_1_3	C>T	Histidine kinase	BCV53_04235	-
2657967	AM242_1_3	G>A	Acetaldehyde dehydrogenase (acetylating)	BCV53_13005	-
3113091	AM242_1_3	G>A	Hypothetical protein	BCV53_15235	G159R
1415815	AM242_1_4	A>T	AraC family transcriptional regulator	BCV53_07015	Q158L
1471195	AM242_1_4	G>A	Carbamoyl phosphate synthase large subunit	BCV53_07290	V152I
2169641	AM242_1_4	C>T	Acyl-CoA dehydrogenase	BCV53_10705	A196T
2570676	AM242_1_4	A>G	Acyl-CoA dehydrogenase	BCV53_12610	M208V
3113091	AM242_1_4	G>A	Hypothetical protein	BCV53_15235	G159R
3228940	AM242_1_4	T>C	-	-	-
1415815	AM242_1_5	A>T	AraC family transcriptional regulator	BCV53_07015	Q158L
1471195	AM242_1_5	G>A	Carbamoyl phosphate synthase large subunit	BCV53_07290	V152I
2169641	AM242_1_5	C>T	Acyl-CoA dehydrogenase	BCV53_10705	A196T

2570676	AM242_1_5	A>G	Acyl-CoA dehydrogenase	BCV53_12610	M208V
3228940	AM242_1_5	T>C	-	-	-
3411590	AM242_1_5	G>A	4-hydroxy-3-methylbut-2-en-1-yl diphosphatesynthase	BCV53_16860	E114K
856889	AM242_1_6	C>T	Histidine kinase	BCV53_04235	-
1514453	AM242_1_6	G>A	Stage V sporulation protein AD	BCV53_07525	G166E
2337594	AM242_1_6	G>A	Arabinose-binding protein	BCV53_11500	G314E
2657967	AM242_1_6	G>A	Acetaldehyde dehydrogenase (acetylating)	BCV53_13005	-
1955303	AM242_2_2	C>G	Hypothetical protein	BCV53_09715	A157P
2891366	AM242_2_3	G>A	Hypothetical protein	BCV53_14215	R56W
856889	AM242_3_1	C>T	Histidine kinase	BCV53_04235	-
1514453	AM242_3_1	G>A	Stage V sporulation protein AD	BCV53_07525	G166E
2657967	AM242_3_1	G>A	Acetaldehyde dehydrogenase (acetylating)	BCV53_13005	-
856889	AM242_3_2	C>T	Histidine kinase	BCV53_04235	-
2657967	AM242_3_2	G>A	Acetaldehyde dehydrogenase (acetylating)	BCV53_13005	-
3113091	AM242_3_2	G>A	Hypothetical protein	BCV53_15235	G159R
856889	AM242_3_3	C>T	Histidine kinase	BCV53_04235	-
2657967	AM242_3_3	G>A	Acetaldehyde dehydrogenase (acetylating)	BCV53_13005	-
1086291	AM242_3_4	T>C	Hypothetical protein	BCV53_05335	E137G
3015609	AM242_3_4	C>T	Nucleotidyltransferase	BCV53_14775	-
856889	AM242_3_5	C>T	Histidine kinase	BCV53_04235	-
3113091	AM242_3_5	G>A	Hypothetical protein	BCV53_15235	G159R
3015609	AM242_3_6	C>T	Nucleotidyltransferase	BCV53_14775	-
609276	AM242_4_2	G>A	DNA gyrase subunit B	BCV53_03070	T353M
2242729	AM242_4_2	T>C	-	-	-
405673	AM242_4_3	T>-	MerR family transcriptional regulator	BCV53_01955	V65fs
1367189	AM242_4_3	A>G	Hypothetical protein	BCV53_06805	-
1541623	AM242_4_3	C>T	Hypothetical protein	BCV53_07715	L39F
2916351	AM242_4_3	G>A	MFS transporter	BCV53_14315	A155T

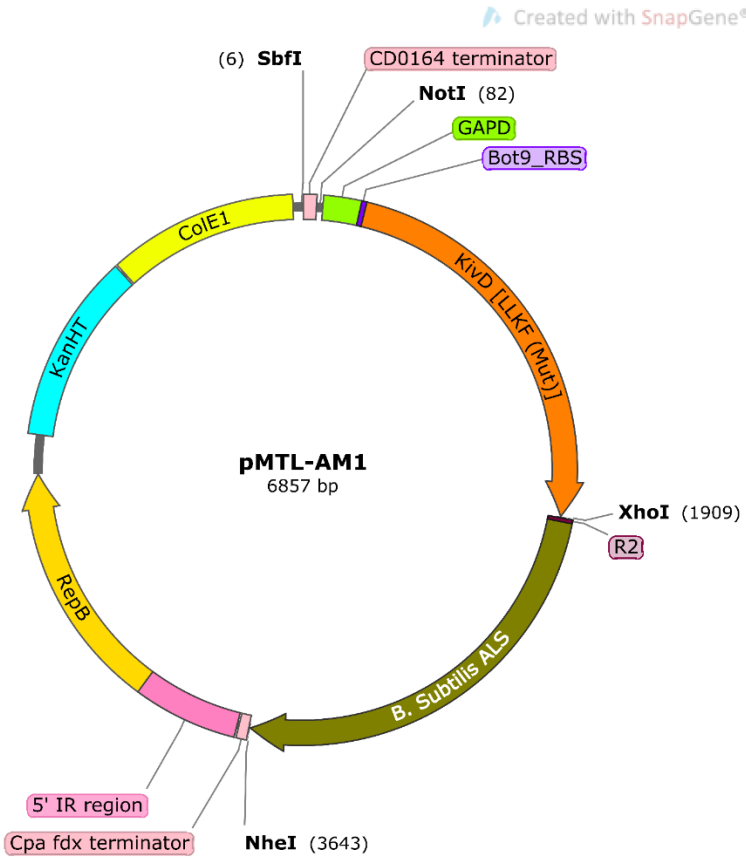
2638542	AM242_4_4	C>T	Permease	BCV53_12920	D231N
2638542	AM242_4_5	C>T	Permease	BCV53_12920	D231N
3113091	AM242_4_5	G>A	Hypothetical protein	BCV53_15235	G159R
1084854	AM242_4_6	A>T	Hypothetical protein	BCV53_05335	L616*
2638542	AM242_4_6	C>T	Permease	BCV53_12920	D231N
3219412	AM242_4_6	G>A	Glutamate dehydrogenase	BCV53_15805	Q148*
1367189	AM242_5_1	A>G	Hypothetical protein	BCV53_06805	-
1986844	AM242_5_1	G>A	-	-	-
2360630	AM242_5_1	C>T	Xylulokinase	BCV53_11600	T180M
2791650	AM242_5_1	C>T	ABC transporter permease	BCV53_13695	-
1367189	AM242_5_2	A>G	Hypothetical protein	BCV53_06805	-
1986844	AM242_5_2	G>A	-	-	-
2063731	AM242_5_2	C>T	PTS sugar transporter subunit IIA	BCV53_10245	-
2791650	AM242_5_2	C>T	ABC transporter permease	BCV53_13695	-
3113091	AM242_5_2	G>A	Hypothetical protein	BCV53_15235	G159R
3523149	AM242_5_2	C>T	tRNA guanosine(34) transglycosylase Tgt	BCV53_17445	G186D
26985	AM242_5_2	A>G	Hydroxyglutarate oxidase	BCV53_19285	K237R
1367189	AM242_5_3	A>G	Hypothetical protein	BCV53_06805	-
2791650	AM242_5_3	C>T	ABC transporter permease	BCV53_13695	-
3523149	AM242_5_3	C>T	tRNA guanosine(34) transglycosylase Tgt	BCV53_17445	G186D
26985	AM242_5_3	A>G	Hydroxyglutarate oxidase	BCV53_19285	K237R
1367189	AM242_5_4	A>G	Hypothetical protein	BCV53_06805	-
1986844	AM242_5_4	G>A	-	-	-
2791650	AM242_5_4	C>T	ABC transporter permease	BCV53_13695	-
1367189	AM242_5_5	A>G	Hypothetical protein	BCV53_06805	-
1986844	AM242_5_5	G>A	-	-	-
2360630	AM242_5_5	C>T	Xylulokinase	BCV53_11600	T180M
2791650	AM242_5_5	C>T	ABC transporter permease	BCV53_13695	-

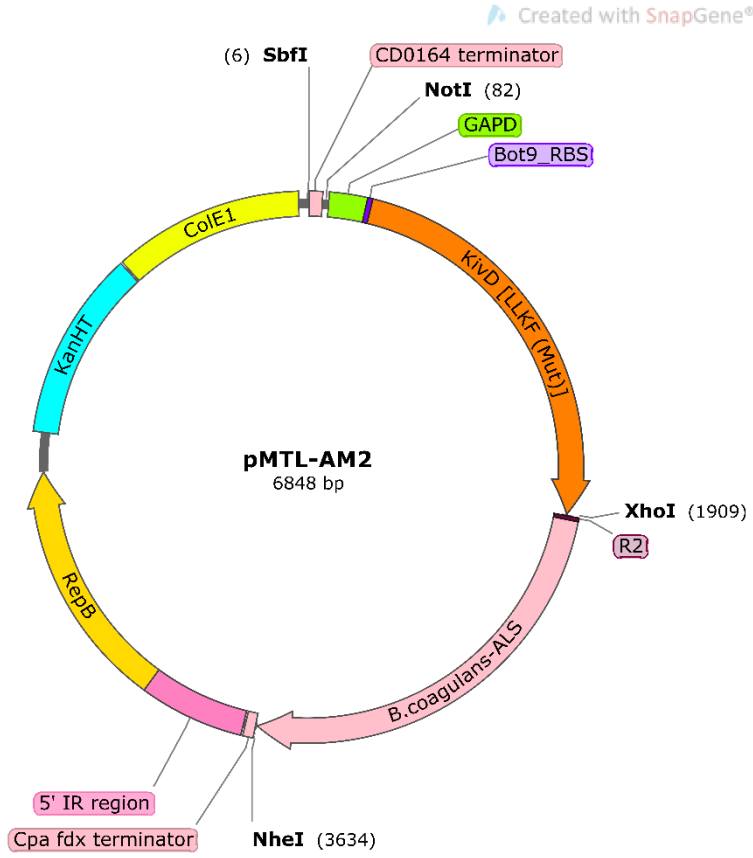
1367189	AM242_5_6	A>G	Hypothetical protein	BCV53_06805	-
2360630	AM242_5_6	C>T	Xylulokinase	BCV53_11600	T180M
2791650	AM242_5_6	C>T	ABC transporter permease	BCV53_13695	-
2638542	AM242_6_1	C>T	Permease	BCV53_12920	D231N
1084854	AM242_6_2	A>T	Hypothetical protein	BCV53_05335	L616*
2638542	AM242_6_2	C>T	Permease	BCV53_12920	D231N
3219412	AM242_6_2	G>A	Glutamate dehydrogenase	BCV53_15805	Q148*
609276	AM242_6_4	G>A	DNA gyrase subunit B	BCV53_03070	T353M
2242729	AM242_6_4	T>C	-	-	-
2638542	AM242_6_5	C>T	Permease	BCV53_12920	D231N
296868	AM242_6_6	A>G	Transposase	BCV53_01475	C72R

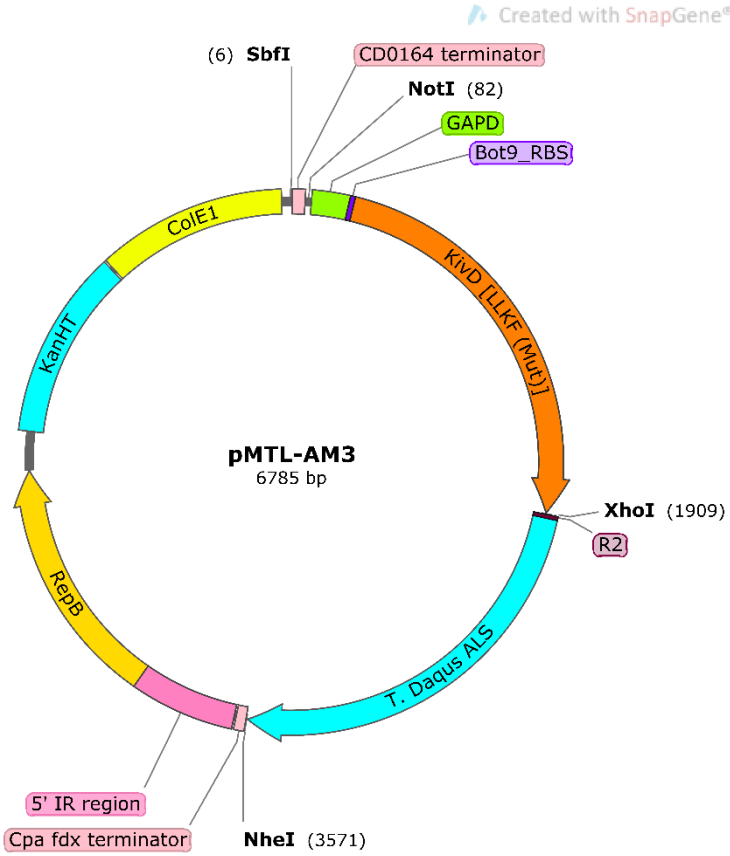
Appendix VII: SNVs and Indels of AM242_G>A and TM242_A>G strains

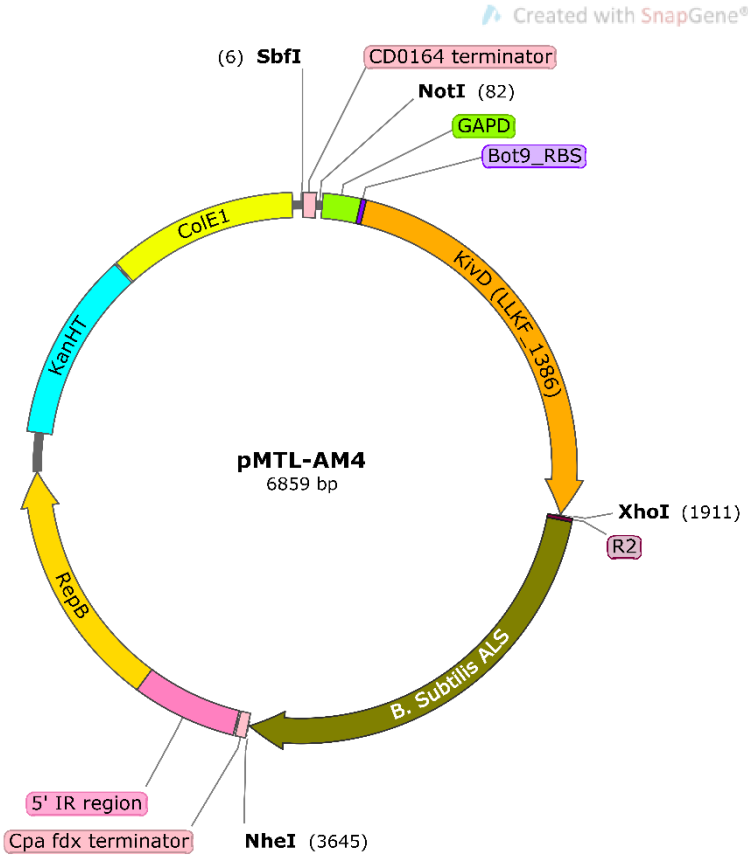
Position	Strain	SNP	Gene	Locus	Effect
525533	AM242_G>A_1	TT>CC	-	-	-
2046908	AM242_G>A_1	C>T	Quinone oxidoreductase	BCV53_10160	A218V
2915662	AM242_G>A_1	TTTA>-	-	-	-
3344819	AM242_G>A_1	G>A	RNA methyltransferase	BCV53_16485	Q82*
525533	AM242_G>A_2	TT>CC	-	-	-
2046908	AM242_G>A_2	C>T	Quinone oxidoreductase	BCV53_10160	A218V
2915662	AM242_G>A_2	TTTA>-	-	-	-
3344819	AM242_G>A_2	G>A	RNA methyltransferase	BCV53_16485	Q82*
525533	AM242_G>A_3	TT>CC	-	-	-
2046908	AM242_G>A_3	C>T	Quinone oxidoreductase	BCV53_10160	A218V
2847841	AM242_G>A_3	A>T	-	-	-
3344819	AM242_G>A_3	G>A	RNA methyltransferase	BCV53_16485	Q82*
105832	TM242_A>G_1	G>T	Transposase	BCV53_00590	-
557479	TM242_A>G_1	T>C	Aminoacyl-tRNA hydrolase	BCV53_02800	-
911418	TM242_A>G_1	A>-	Preprotein translocase subunit TatA	BCV53_04560	Lys47fs
3836489	TM242_A>G_1	A>C	-	-	-
105832	TM242_A>G_2	G>T	Transposase	BCV53_00590	-
209465	TM242_A>G_2	G>A	-	-	-
557479	TM242_A>G_2	T>C	Aminoacyl-tRNA hydrolase	BCV53_02800	-
105832	TM242_A>G_3	G>T	Transposase	BCV53_00590	-
557479	TM242_A>G_3	T>C	Aminoacyl-tRNA hydrolase	BCV53_02800	-
911418	TM242_A>G_3	A>-	Preprotein translocase subunit TatA	BCV53_04560	Lys47fs
3836489	TM242_A>G_3	A>C	-	-	-

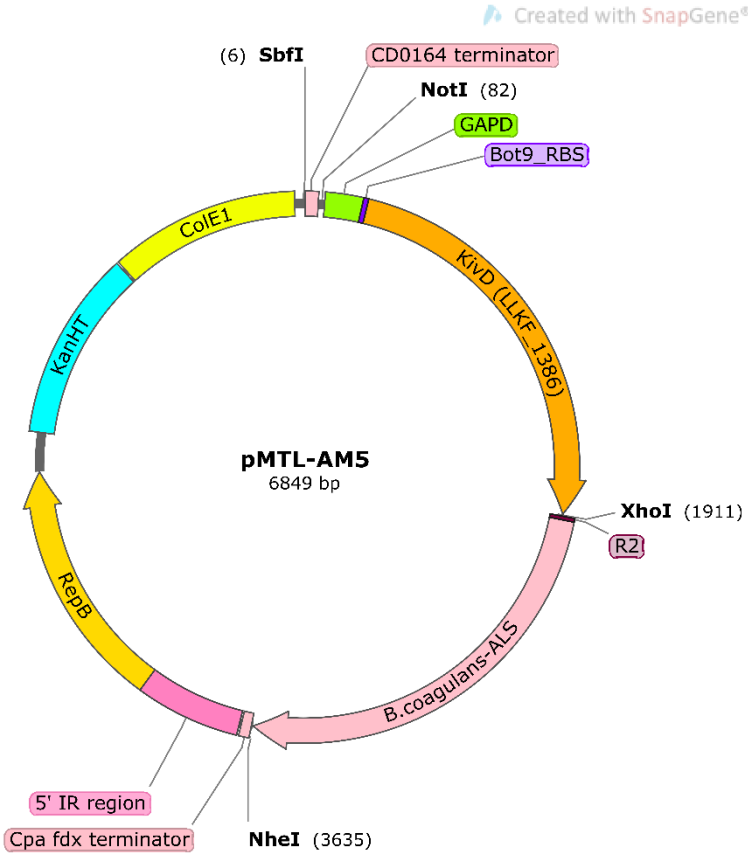
Appendix VIII: Plasmid maps of pMTL-AM1, pMTL-AM2, pMTL-AM3, pMTL-AM4, pMTL-AM5, pMTL-AM6, pMTL-AM7, pMTL-AM8 and pMTL-AM9. The plasmids consist of a Gram-positive replicon, RepB, with the 5′ incompatibility region (IR); a selectable marker, *kanHT*, conferring resistance to kanamycin; a Gram-negative replicon, ColE1; a transcriptional terminator isolated from downstream of the *Clostridium difficile* strain 630 CD0164 gene, CD0164 terminator; a transcriptional terminator of the ribosomal RNA gene (*rrnB*) of *E. coli*, T1T2 (pMTL-AM7, pMTL-AM8, pMTL-AM9); glyceraldehyde-3-phosphate dehydrogenase promoter, GAPD (pMTL-AM1, pMTL-AM2, pMTL-AM3, pMTL-AM4, pMTL-AM5, pMTL-AM6); ribosome binding site, Bot9_RBS (Lau *et al.*, 2021); lactate dehydrogenase promoter with native ribosome binding site from *Geobacillus stearothermophilus*, LDH_Native_RBS (pMTL-AM7, pMTL-AM8, pMTL-AM9); ketoisovalerate decarboxylase from *L. lactis*, kivD [LLKF(Mut.)] (pMTL-AM1, pMTL-AM2, pMTL-AM3); ketoisovalerate decarboxylase from *L. lactis*, kivD (LLKF_1386) (pMTL-AM4, pMTL-AM5, pMTL-AM6, pMTL-AM7, pMTL-AM8, pMTL-AM9); ribosome binding site, R2 (Lau *et al.*, 2021); acetolactate synthase (ALS) from *B. subtilis* (pMTL-AM1, pMTL-AM4, pMTL-AM7), *B. coagulans* (pMTL-AM2, pMTL-AM5, pMTL-AM8), and *T. daqus* (pMTL-AM3, pMTL-AM6, pMTL-AM9); and a transcriptional terminator of the ferredoxin gene of *Clostridium pasteurianum*, Cpa fdx terminator.

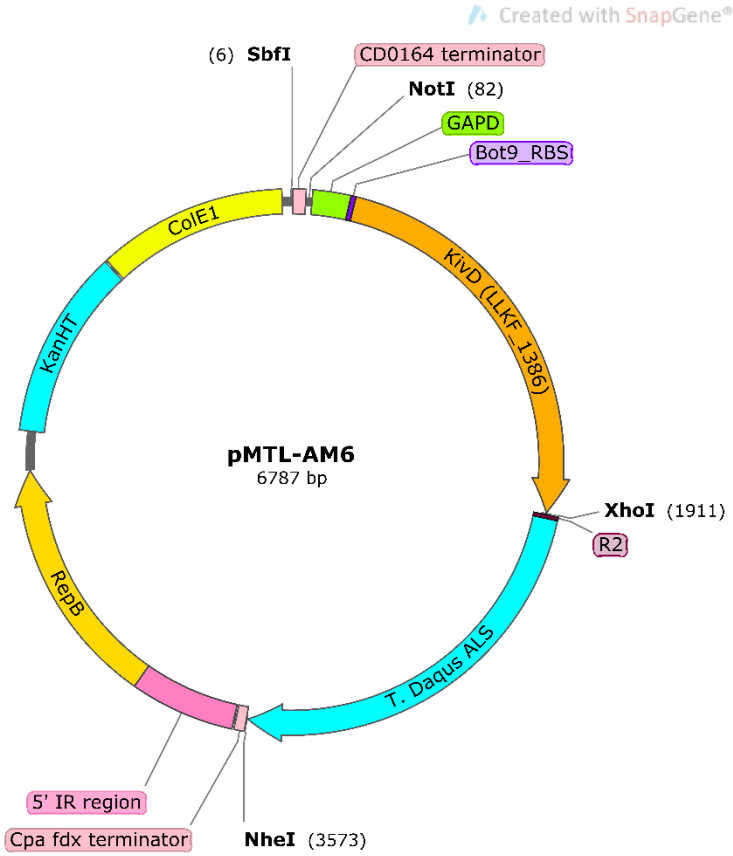

Created with SnapGene®
(6) SbfI
CD0164 terminator
NotI (82)
GAPD
Bot9_RBS
ColE1
KivD [LLKF (Mut)]
KanHT
pMTL-AM1
6857 bp
XhoI (1909)
R2
RepB
B. Subtilis ALS
5' IR region
Cpa fdx terminator
NheI (3643)

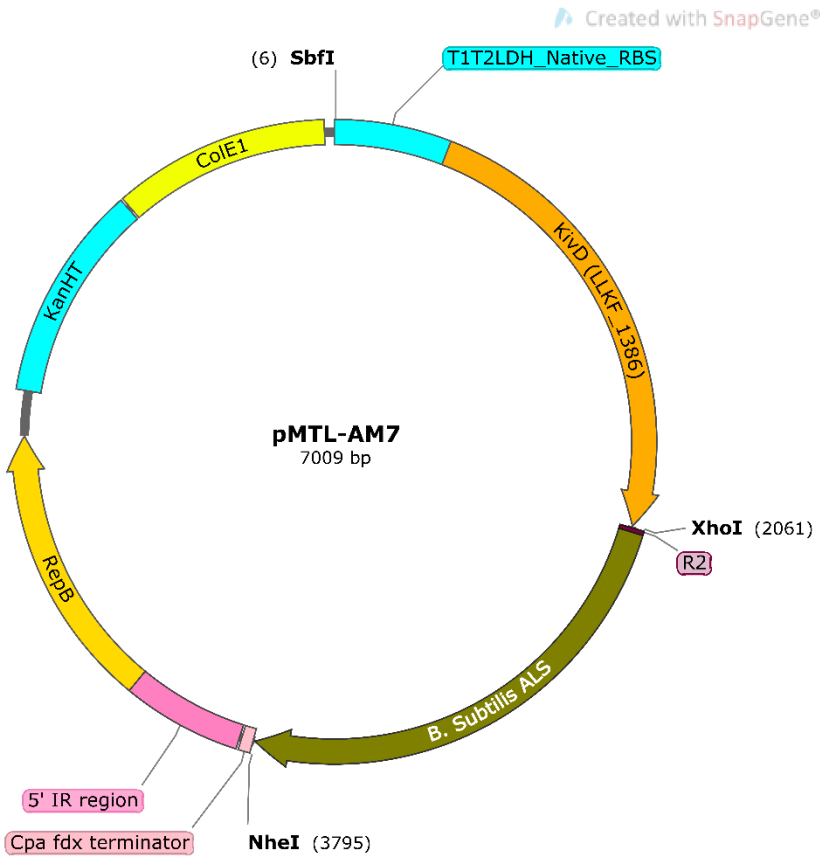

Created with SnapGene®
(6) SbfI
CD0164 terminator
NotI (82)
GAPD
Bot9_RBS
ColE1
KivD [LLKF (Mut)]
KanHT
pMTL-AM2
6848 bp
XhoI (1909)
R2
RepB
B.coagulans-ALS
5' IR region
Cpa fdx terminator
NheI (3634)

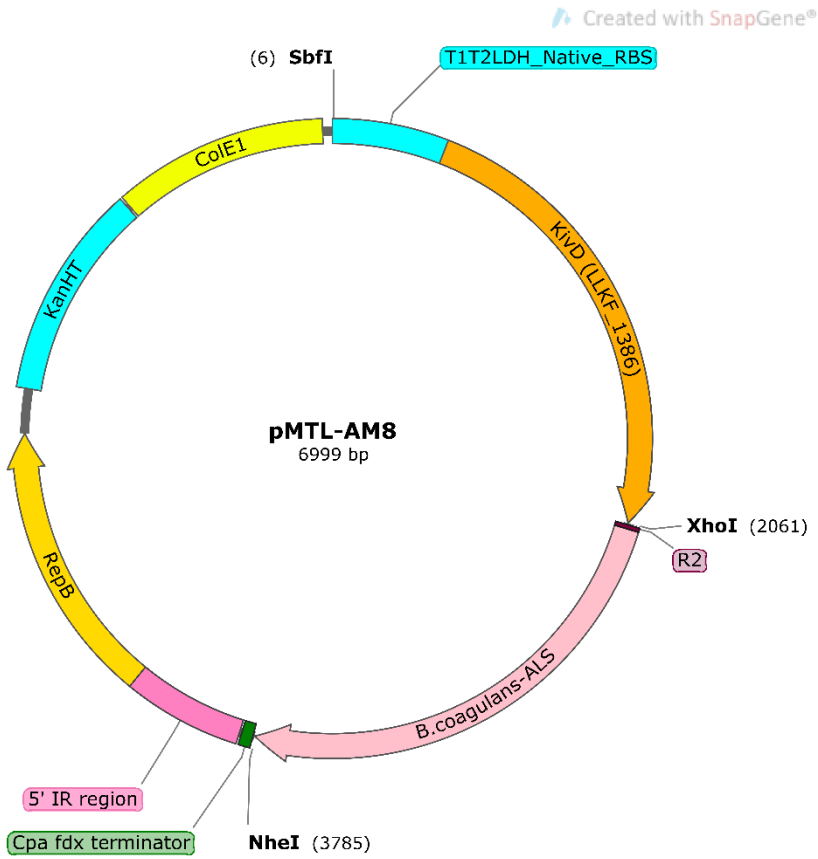

Created with SnapGene®
(6) SbfI
CD0164 terminator
NotI (82)
GAPD
Bot9_RBS
ColE1
KivD [LLKF (Mut)]
KanHT
pMTL-AM3
6785 bp
XhoI (1909)
R2
RepB
T. Daqus ALS
5' IR region
Cpa fdx terminator
NheI (3571)

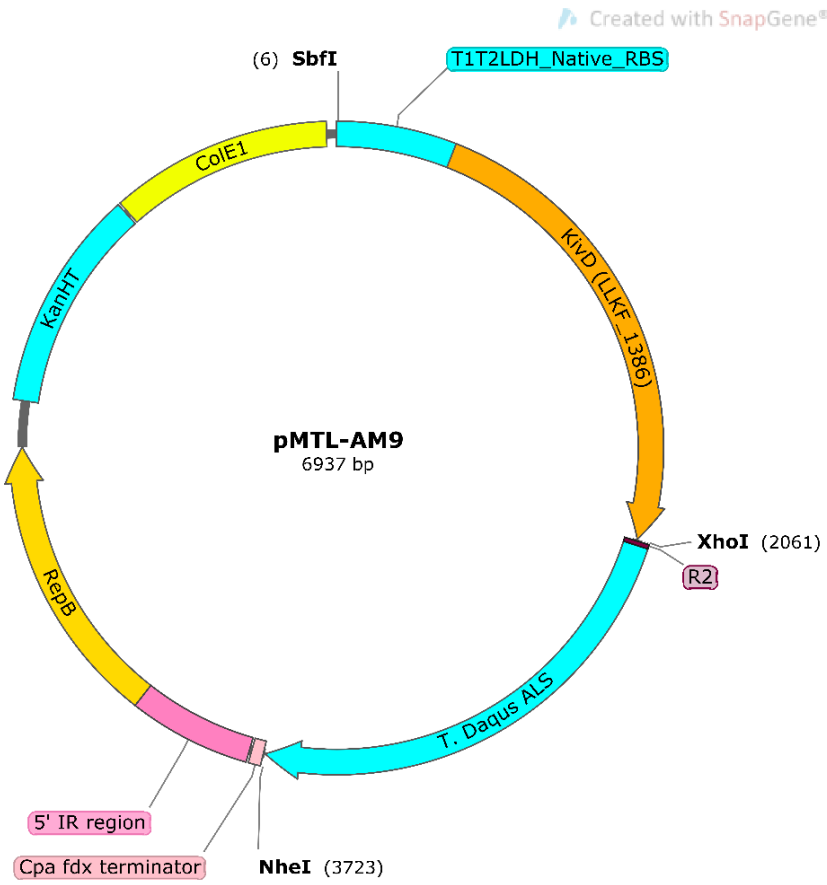

Created with SnapGene®
(6) SbfI
CD0164 terminator
NotI (82)
GAPD
Bot9_RBS
ColE1
KivD (LLKF_1386)
KanHT
pMTL-AM4
6859 bp
XhoI (1911)
R2
RepB
B. Subtilis ALS
5' IR region
Cpa fdx terminator
NheI (3645)


Created with SnapGene®
(6) SbfI
CD0164 terminator
NotI (82)
GAPD
Bot9_RBS
ColE1
KivD (LLKF_1386)
KanHT
pMTL-AM5
6849 bp
XhoI (1911)
R2
RepB
B.coagulans-ALS
5' IR region
Cpa fdx terminator
NheI (3635)

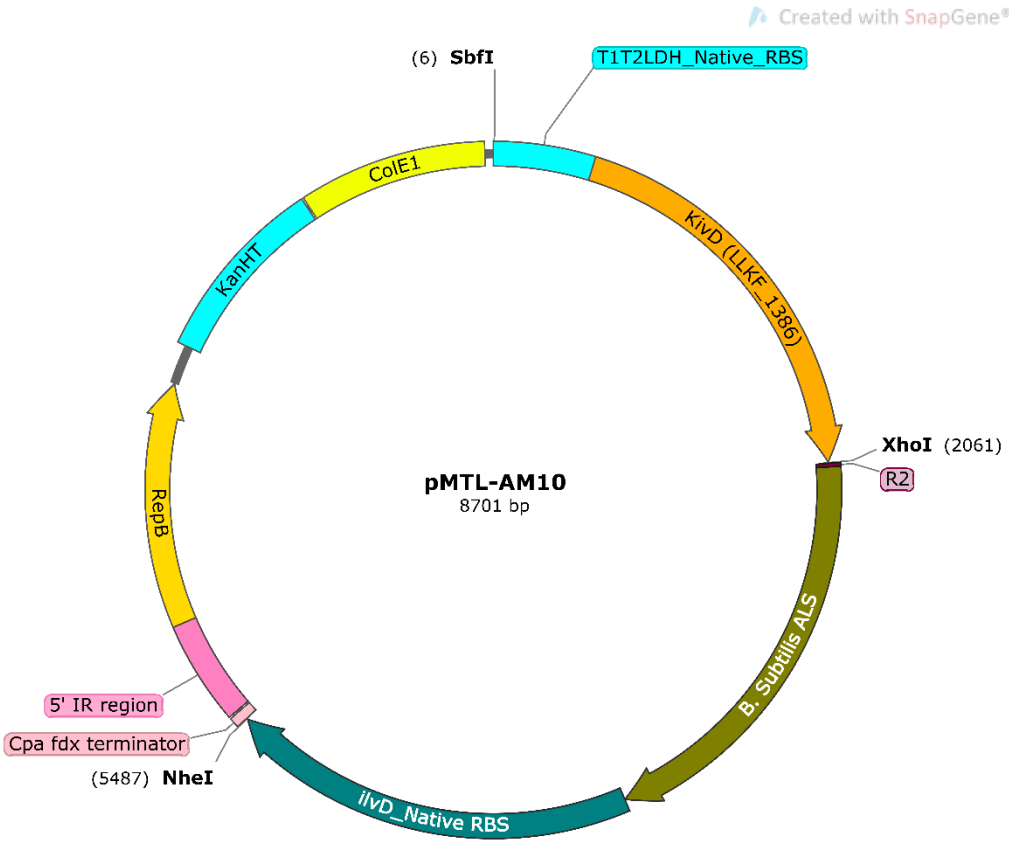

Created with SnapGene®
(6) SbfI
CD0164 terminator
NotI (82)
GAPD
Bot9_RBS
ColE1
KivD (LLKF_1386)
KanHT
pMTL-AM6
6787 bp
XhoI (1911)
R2
RepB
T. Daqus ALS
5' IR region
Cpa fdx terminator
NheI (3573)

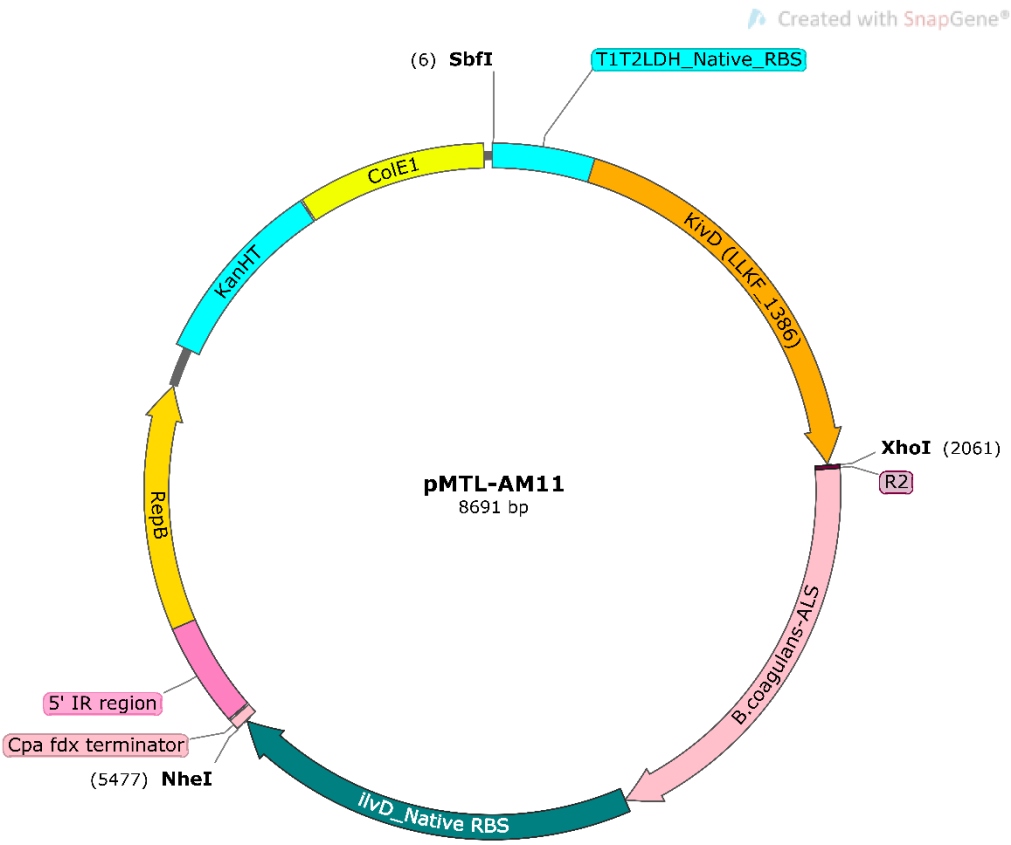

Created with SnapGene®
(6) SbfI
T1T2LDH_Native_RBS
ColE1
KivD (LLKF_1386)
KanHT
pMTL-AM7
7009 bp
XhoI (2061)
R2
RepB
B. Subtilis ALS
5' IR region
Cpa fdx terminator
NheI (3795)

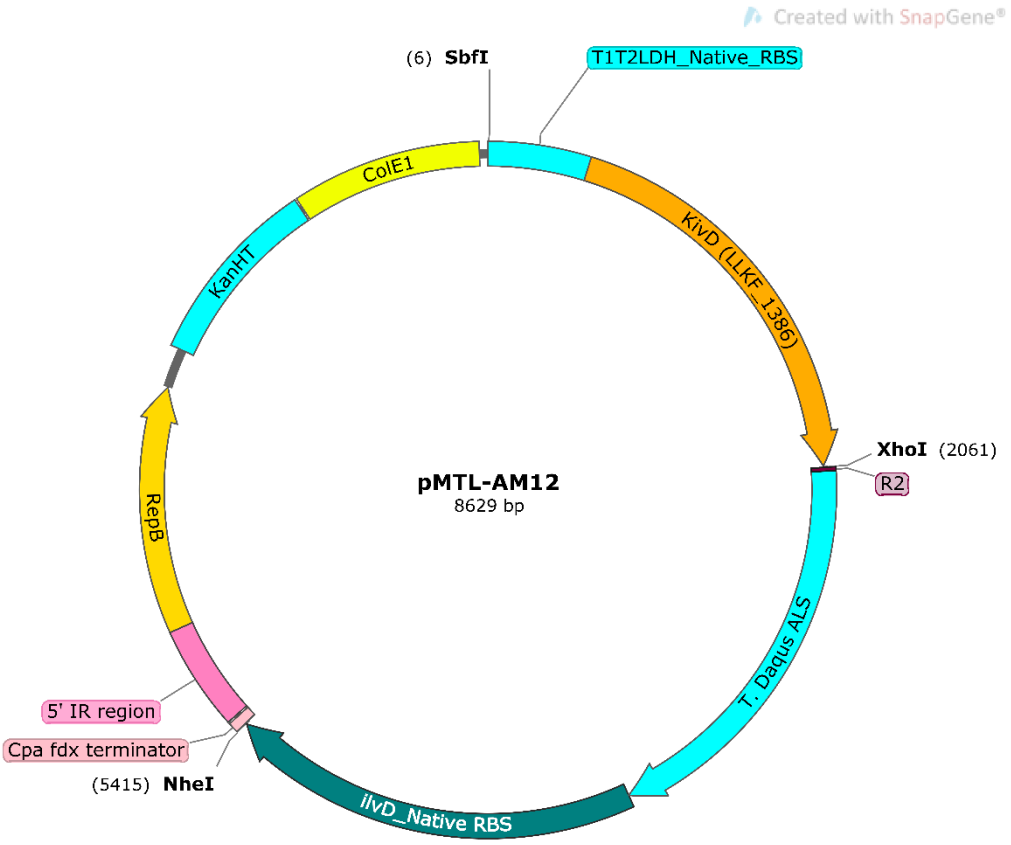

Created with SnapGene®
(6) SbfI
T1T2LDH_Native_RBS
ColE1
KivD (LLKF_1386)
KanHT
pMTL-AM8
6999 bp
XhoI (2061)
R2
RepB
B.coagulans-ALS
5' IR region
Cpa fdx terminator
NheI (3785)


Created with SnapGene®
(6) SbfI
T1T2LDH_Native_RBS
ColE1
KivD (LLKF_1386)
KanHT
pMTL-AM9
6937 bp
XhoI (2061)
R2
RepB
T. Daqus ALS
5' IR region
Cpa fdx terminator
NheI (3723)

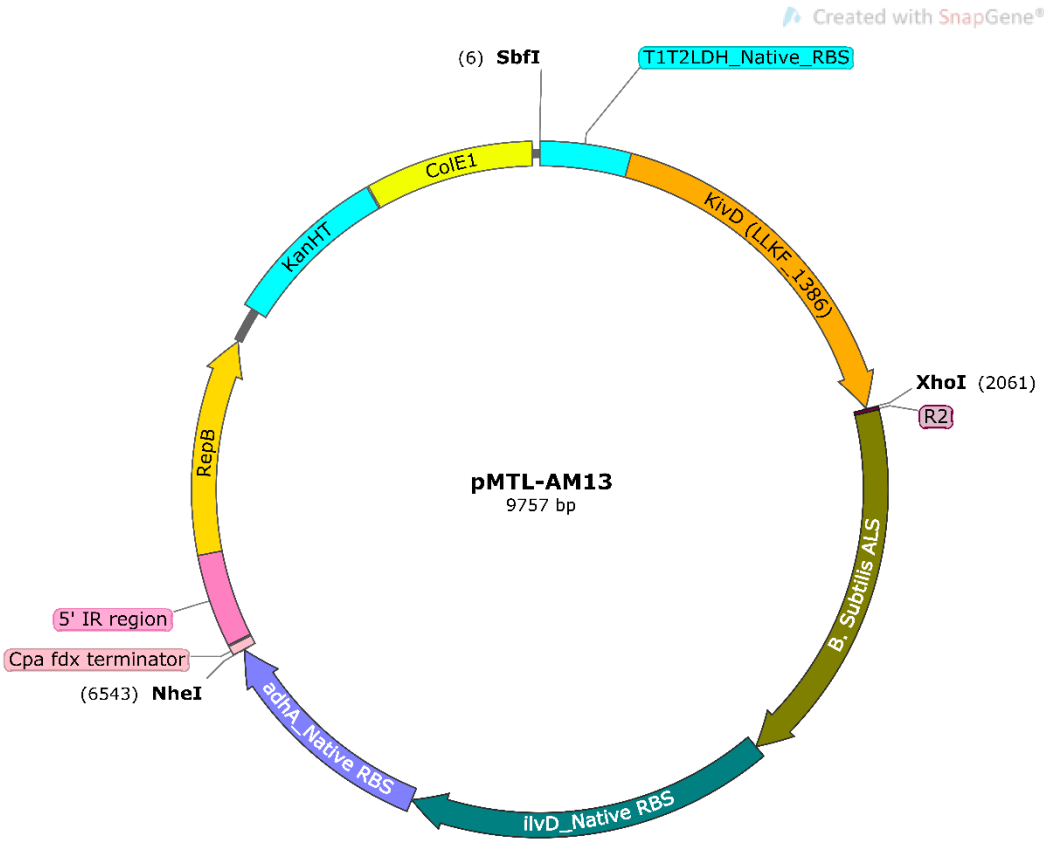
Appendix IX: Plasmid maps of pMTL-AM10, pMTL-AM11 and pMTL-AM12. The plasmids consist of a Gram-positive replicon, RepB, with the 5′ incompatibility region (IR); a selectable marker, *kanHT*, conferring resistance to kanamycin; a Gram-negative replicon, ColE1; a transcriptional terminator of the ribosomal RNA gene (*rrnB*) of *E. coli*, T1T2; lactate dehydrogenase promoter with native ribosome binding site from *Geobacillus stearothermophilus*, LDH_Native_RBS; ketoisovalerate decarboxylase from *L. lactis*, kivD (LLKF_1386); ribosome binding site, R2 (Lau *et al.*, 2021); acetolactate synthase (ALS) from *B. subtilis* (pMTL-AM10), *B. coagulans* (pMTL-AM11), and *T. daqus* (pMTL-AM12); dihydroxy acid dehydratase with native ribosome binding site from *P. thermoglucosidasius* NCIMB 11955, ilvD_Native RBS; and a transcriptional terminator of the ferredoxin gene of *Clostridium pasteurianum*, Cpa fdx terminator.

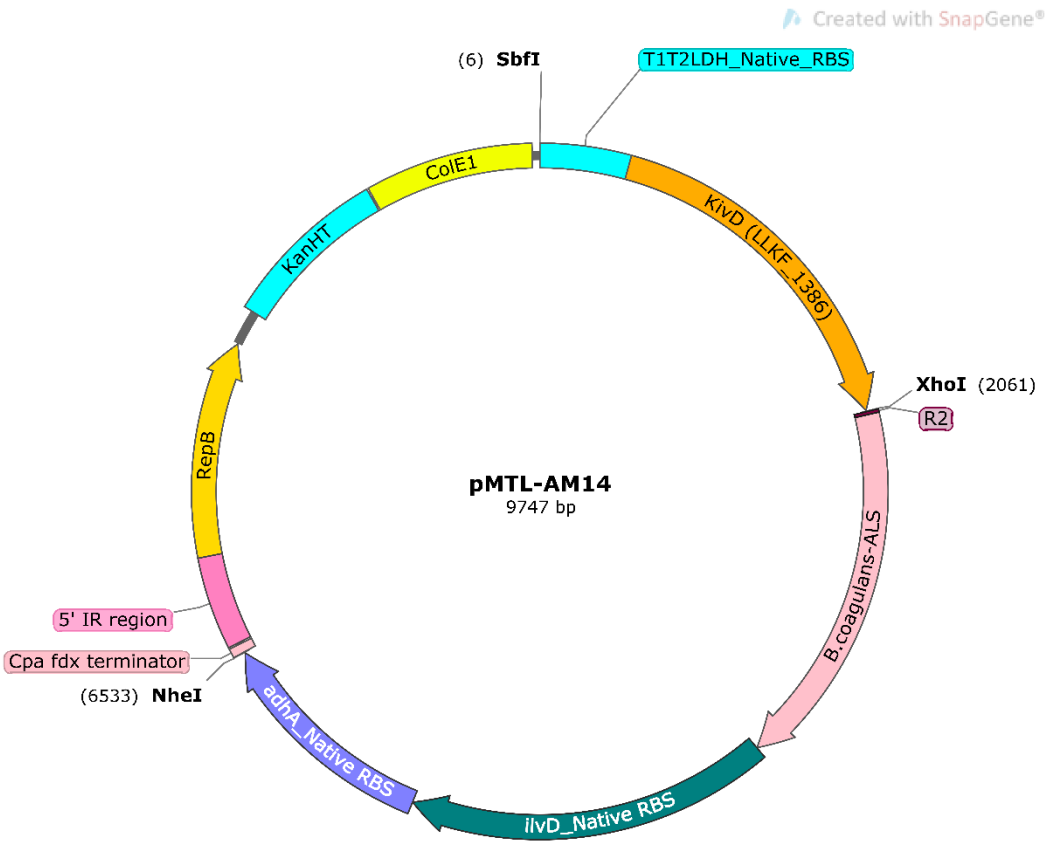


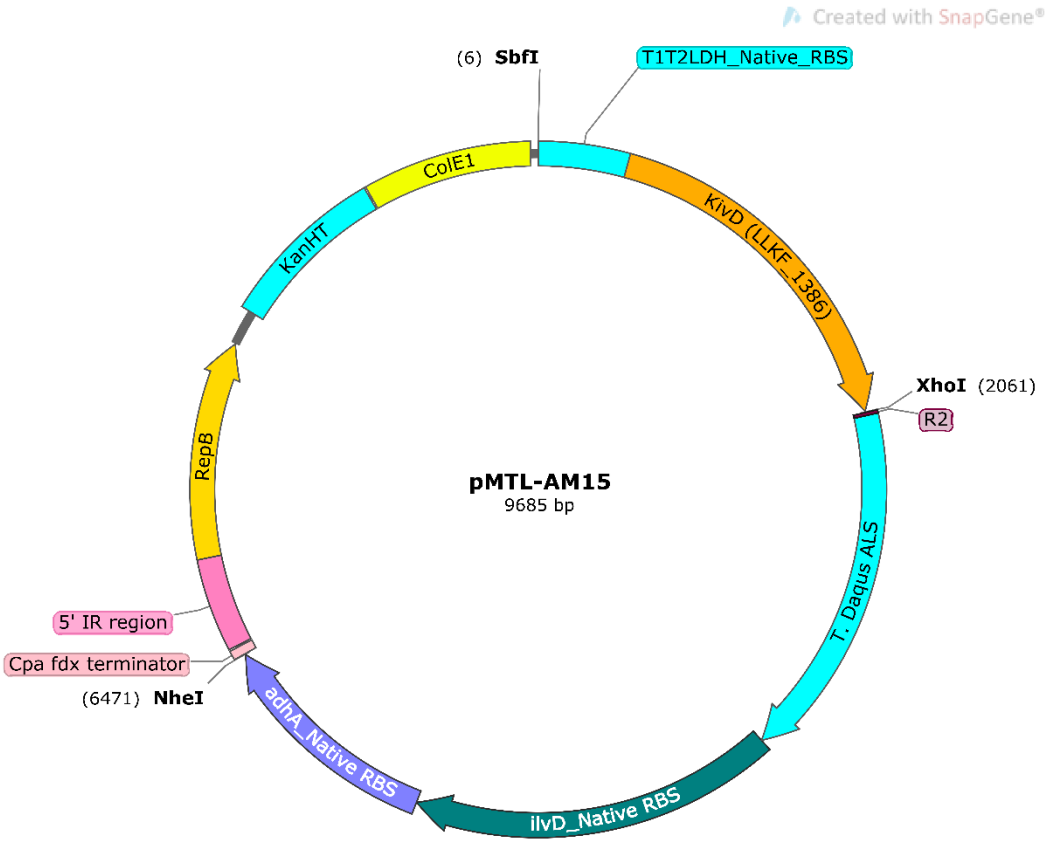

Created with SnapGene®
(6) SbfI
T1T2LDH_Native_RBS
ColE1
KivD (LLKF_1386)
KanHT
XhoI (2061)
R2
pMTL-AM11
8691 bp
RepB
B.coagulans-ALS
5' IR region
Cpa fdx terminator
(5477) NheI
ilvD_Native RBS


Created with SnapGene®
(6) SbfI
T1T2LDH_Native_RBS
ColE1
KivD (LLKF_1386)
KanHT
XhoI (2061)
R2
pMTL-AM12
8629 bp
RepB
T. Daqus ALS
5' IR region
Cpa fdx terminator
(5415) NheI
ilvD_Native RBS

Appendix X: Plasmid maps of pMTL-AM13, pMTL-AM14 and pMTL-AM15. The plasmids consist of a Gram-positive replicon, RepB, with the 5′ incompatibility region (IR); a selectable marker, *kanHT*, conferring resistance to kanamycin; a Gram-negative replicon, ColE1; a transcriptional terminator of the ribosomal RNA gene (*rrnB*) of *E. coli*, T1T2; lactate dehydrogenase promoter with native ribosome binding site from *Geobacillus stearothermophilus*, LDH_Native_RBS; ketoisovalerate decarboxylase from *L. lactis*, kivD (LLKF_1386); ribosome binding site, R2 (Lau *et al.*, 2021); acetolactate synthase (ALS) from *B. subtilis* (pMTL-AM13), *B. coagulans* (pMTL-AM14), and *T. daqus* (pMTL-AM15); dihydroxy acid dehydratase with native ribosome binding site from *P. thermoglucosidasius* NCIMB 11955, ilvD_Native RBS; alcohol dehydrogenase with native ribosome binding site from *P. thermoglucosidasius* NCIMB 11955, adhA_Native RBS; and a transcriptional terminator of the ferredoxin gene of *Clostridium pasteurianum*, Cpa fdx terminator.




Created with SnapGene®
(6) SbfI
T1T2LDH_Native_RBS
ColE1
KivD (LLKF_1386)
KanHT
XhoI (2061)
R2
pMTL-AM14
9747 bp
RepB
B.coagulans-ALS
5' IR region
Cpa fdx terminator
(6533) NheI
adhA_Native RBS
ilvD_Native RBS


Created with SnapGene®
(6) SbfI
T1T2LDH_Native_RBS
ColE1
KivD (LLKF_1386)
KanHT
XhoI (2061)
R2
pMTL-AM15
9685 bp
RepB
T. Daqus ALS
5' IR region
Cpa fdx terminator
(6471) NheI
adhA_Native RBS
ilvD_Native RBS