

https://dspace.org/

Parageobacillus thermoglucosidasius and the water-gas shift reaction: Investigating the influences of gas composition and H2- uptake hydrogenases

Mol, Michael

2024
University of the Witwatersrand, Johannesburg
https://hdl.handle.net/10539/41805
Mol, Michael. (2024). Parageobacillus thermoglucosidasius and the water-gas shift reaction: Investigating the influences of gas composition and H2- uptake hydrogenases [PhD thesis, University of the Witwatersrand, Johannesburg]. WireDSpace.

Downloaded from a DSpace repository. https://dspace.org/



Parageobacillus thermoglucosidasius and the water-gas shift reaction: Investigating the influences of gas composition and H_2-uptake hydrogenases

by

Michael Mol

(788559)

Thesis

Submitted in fulfilment of the requirements for the degree

Philosophiae Doctor

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Prof. P. De Maayer

May 2024

DECLARATION

I declare that this thesis is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



Michael Mol

May 2024

PREFACE

H_2 gas is an increasingly important commercial reagent in a range of industries, including as a potential fuel and energy carrier. Production of H_2 gas is carried out almost exclusively reliant on fossil fuel reformation through a relatively limited set of environmentally harmful processes. Pressing environmental concerns and the shift in climate policies towards greater incorporation of cleaner H_2 gas production strategies and the implementation of H_2 as a more sustainable and environmentally friendly alternative energy vector have necessitated development of renewable and cleaner H_2 production processes. **Chapter one** discusses available literature on the present applications, predominant production processes and emerging alternative production strategies of H_2 gas, with particular focus on the applicability and mechanism of biological H_2 gas production by the thermophile, *Parageobacillus thermoglucosidasius*, via the water-gas shift reaction.

The application of *P. thermoglucosidasius* to produce H_2 gas is presently in its formative laboratory-scale stage of development. Although progress has been made at a fundamental level, various aspects of *P. thermoglucosidasius* H_2 gas metabolism, remain uncharacterised or poorly understood. In **Chapter two**, the potential influences of two putative H_2-uptake hydrogenases, encoded on the genome of a *P. thermoglucosidasius* strain previously identified to conduct the hydrogenogenic biological water-gas shift reaction was explored through knock-out mutagenesis. Furthermore, in **Chapter three**, to establish the practical implementation of pertinent and inexpensive gas feedstocks for the water-gas shift reaction, we explored the influences of various industrial mimetic gas feedstock compositions on the H_2 gas evolution-, metabolic- and growth-profiles of *P. thermoglucosidasius*. Considering the highly variable compositions of industrial off-gases, which may contain oxygen and the highly sensitive nature of this system to oxygen, we further explored the impacts of varied volumetric and temporal oxygen perturbations on the system.

Aside from hydrogen production, *P. thermoglucosidasius* and closely related thermophilic taxa from the genera *Parageobacillus* and *Saccharococcus* represent appealing targets for various other biotechnological developments. In **Chapter**

four, we performed a comprehensive comparative genomic and phylogenomic analysis to further explore the biotechnological potential of *Parageobacillus* and *Saccharococcus* spp. for a range of whole cell applications as well as a source of industrially relevant thermostable enzymes.

This thesis is dedicated to my parents,

for their unwavering love and support

ACKNOWLEDGEMENTS

I would like to acknowledge and thank the following:

- My wife for her love and support.
- My family for their love and support.
- My supervisor, Prof. Pieter De Maayer at the University of the Witwatersrand, for his guidance, mentorship, and expertise.
- The members of the KIT TeBi lab, including Magda Ardila, Dr. Anke Neuman, Dr. Habibu Aliyu for their guidance, support and collaboration.
- Dr. Angela Botes, at the University of the Witwatersrand for her support.
- Michael Tobin at the University of the Witwatersrand for his support.
- The DSI-CSIR Interbursary support programme for funding my studies.

Publications arising from the research

Mol, M., Ardila, M.S., Mol, B.A., Aliyu, H., Neumann, A. and De Maayer, P., 2024. The effects of synthesis gas feedstocks and oxygen perturbation on hydrogen production by *Parageobacillus thermoglucosidasius*. *Microbial Cell Factories*, *23*(1), pp.125.

Mol, M. and De Maayer, P., 2024. Elucidating the biotechnological potential of the genera *Parageobacillus* and *Saccharococcus* through comparative genomic and pan-genome analysis. Manuscript under review at BMC Genomics.

Mol, M., Mol, B.A. and De Maayer, P. The bacterial hydrogen economy. Intended for submission to Microbiology & Molecular Biology Reviews.

TABLE OF CONTENTS

DECLARATION 2

PREFACE 3

ACKNOWLEDGEMENTS 6

TABLE OF CONTENTS 8

LIST OF FIGURES 12

LIST OF TABLES 15

LIST OF SUPPLEMENTARY TABLES 16

LIST OF ABBREVIATIONS 17

CHAPTER ONE: INTRODUCTION 19

1.1 Hydrogen 20
1.2 Hydrogen and the energy mix 20
1.3 Hydrogen and industry 21
1.4 Hydrogen production 22
1.4.1 Non-renewable H_2 gas production 24
1.4.1.1 Steam reformation 26
1.4.1.2 Partial oxidation 27
1.4.1.3 Autothermal oxidation 27
1.4.2 Renewable H_2 gas production 28
1.4.2.1 Electrolysis 28
1.4.2.2 Thermochemical approaches 30
1.4.2.3 Biological approaches 31
1.5 *Parageobacillus* 41
1.6 *Parageobacillus* and the water-gas shift reaction 42
1.5 Conclusions 45

CHAPTER TWO: EXPLORING THE INFLUENCES OF THE TWO *P. THERMOGLUCOSIDASIUS* GROUP 1D AND GROUP 2A NIFE-UPTAKE HYDROGENASES ON H_2 GAS PRODUCTION THROUGH KNOCKOUT MUTAGENESIS 46

Abstract 47

2.1 Background 48
2.2 Methods and materials 49
2.2.1 Genetic engineering protocol adapted from Cripps *et al*. (2009)..... 50
2.2.1.1 Media, strains and plasmids 53
2.2.1.2 Preparation of electrocompetent *E. coli* DH5α 55
2.2.1.3 Preparation of electrocompetent *P. thermoglucosidasius* 55
2.2.1.4 Generation of pUB190 transformation vector..... 55
2.2.1.5 Transformation of *E. coli* DH5α 60
2.2.1.6 pPlasA primary transformation vector construction 61
2.2.1.7 pPlasB *P. thermoglucosidasius* transformation vector construction 69
2.2.1.8 Electroporation of *P. thermoglucosidasius* with pPlasBΔ*phaB* and pPlasBΔ*phbE* 73
2.2.1.9 Generation of *P. thermoglucosidasius* primary integrants 74
2.2.1.10 Generation of *P. thermoglucosidasius* secondary integrants..... 77
2.2.2 Genetic engineering protocol adapted from Bacon *et al*. (2017)..... 80
2.2.2.1 Media strains and plasmids 83
2.2.2.2 Transformation vector construction using NEBuilder HiFi DNA assembly workflow 84
2.2.2.3 Transformation of *E. coli* 10β 88
2.2.2.4 Validation of pG2K-derived plasmid constructs 89
2.2.2.5 Transformation of *P. thermoglucosidasius*..... 90
2.3 Results..... 90
2.3.1 Genetic engineering protocol adapted from Cripps *et al*. (2009)..... 90
2.3.1.2 Transformation of *P. thermoglucosidasius*..... 92
2.3.1.3 Generation of primary and secondary *P. thermoglucosidasius* integrants..... 93
2.3.2 Genetic engineering protocol adapted from Bacon *et al*. (2017)..... 98
2.3.2.1 Transformation of *E. coli* 10β with pG2K-derived vectors 98
2.3.2.2 Transformation of *P. thermoglucosidasius* with pG2KΔ*phaB* and pG2KΔ*phbE*..... 98
2.4 Discussion 99
CHAPTER THREE: THE EFFECTS OF SYNTHETIC SYNGAS FEEDSTOCKS AND OXYGEN PERTURBATION ON HYDROGEN PRODUCTION BY *P. THERMOGLUCOSIDASIUS*..... 104

Abstract 105
3.1 Background 106
3.2 Methods and materials 108
3.2.1 Strains and media 108
3.2.2 Experimental setup 108
3.2.3 Sampling and analytics 111
3.2.4 Statistical analysis 112
3.3 Results 113
3.3.1 Anaerobic synthesis gas mimetic feedstocks promote earlier onset of hydrogenogenesis at the expense of biomass accumulation 113
3.3.2 Pre-hydrogenogenic phase O_2 perturbations extend the lag phase prior to H_2 production 118
3.3.3 Minor peak-hydrogenogenic phase O_2 perturbation minimally disrupts H_2 production in *P. thermoglucosidasius* DSM 6285 124
3.4 Discussion 130
3.5 Conclusions 141
CHAPTER FOUR: ELUCIDATING THE BIOTECHNOLOGICAL POTENTIAL OF THE GENERA *PARAGEOBACILLUS* AND *SACCHAROCOCCUS* THROUGH COMPARATIVE GENOMIC AND PAN-GENOMIC ANALYSIS 142
Abstract 143
4.1 Background 144
4.2 Methods 146
4.2.1 Genome assembly and annotation 146
4.2.2 Phylogenomic analyses 147
4.2.3 Pan-genome analyses 148
4.3 Results and discussion 148
4.3.1 Phylogenomic analysis delineates *Parageobacillus* and *Saccharococcus* as two distinct genera 148
4.3.2 *Parageobacillus* and *Saccharococcus* have open pan-genomes with *P. toebii* as a key driver of novel gene accrual 153
4.3.3 Plasmids, bacteriophages and transposable elements are key drivers of *Parageobacillus* and *Saccharococcus* diversification 160
4.3.4 Mining the *Parageobacillus* and *Saccharococcus* pan-genome for biotechnology 161

4.3.4.1 *Parageobacillus* as a source of novel antimicrobials 161
4.3.4.2 *Parageobacillus* and *Saccharococcus* as sources of bioindustrially relevant enzymes 166
4.3.4.3 Enzymes for the molecular laboratory 172
4.3.4.4 Whole-cell biotechnological applications for *Parageobacillus* and *Saccharococcus* 173
4.3.4.5 *Parageobacillus* as a producer of green energy 178
2.4 Conclusions 180
SUMMARY 181
APPENDIX A 183
APPENDIX B 184
APPENDIX C 220
REFERENCES 247
PLAGIARISM REPORT 289

LIST OF FIGURES

Figure 1.1: Hydrogen production strategies. 23
Figure 1.2: Schematic diagram of the membrane bound WGS system present in *P. thermoglucosidasius* DSM 6285. 43
Figure 2.1: Schematic diagram of knockout mutagenesis key steps. 51
Figure 2.2: Schematic diagram of primary (a) and secondary (b) plasmid integrants following successive double crossover homologous recombination events 52
Figure 2.3: Schematic diagrams of pUC19 (a) and pUB110 (b) plasmids and shuttle vector pUB190 (c). 57
Figure 2.4: 0.8% Agarose gel showing XbaI-HF/EcoRI-HF digestion of pUB110 60
Figure 2.5: Schematic diagram showing construction of mutant target gene constructs. 62
Figure 2.6: Schematic diagram showing pPlasA construct framework. 62
Figure 2.7: Schematic diagram showing *P. thermoglucosidasius* primer binding sites used to generate and validate *phaB* and *phbE* mutant gene fragments. 66
Figure 2.8: Schematic diagram showing primer binding sites targeted for verification of pPlasA constructs. 69
Figure 2.9: Schematic diagram of pPlasB orientation 1 (a) and orientation 2 (b) architectures. 72
Figure 2.10: Schematic diagram showing primer pairs and binding sites used to detect presence of pPlasB retention (yellow) and potential insertion of the *kan* gene (light blue) into the *P. thermoglucosidasius* chromosome for pPlasB construct orientation 1 (a) and 2 (b). 79
Figure 2.11: Schematic diagram of plasmid pG2K (a) and plasmid constructs pG2KΔ*phaB* (b) and pG2KΔ*phbE* (c). 81
Figure 2.12: Schematic diagram of *P. thermoglucosidasius* primary and secondary plasmid integrants following successive double crossover homologous recombination events. 82
Figure 2.13: 0.8% Agarose gels showing mutant gene insertion orientations 1 (a) and 2 (b). 92

Figure 2.14: 3% Agarose gels showing wild type and mutant *phaB* (a) and *phbE* (b) gene amplicons from kanamycin resistant, presumed primary integrant *P. thermoglucosidasius* colonies. 95
Figure 2.15: 3% Agarose gel showing wild type and mutant *phaB* (a) and *phbE* (b) gene amplicons from kanamycin sensitive, presumed secondary integrant *P. thermoglucosidasius* colonies. 96
Figure 2.16: 3% Agarose gel showing wild type and mutant *phaB* (a) and *phbE* (b) gene amplicons from kanamycin resistant *P. thermoglucosidasius* colonies..... 97
Figure 3.1: Gas composition (1), metabolite (2) and growth and pH (3) profiles for *P. thermoglucosidasius* DSM 6285 cultured under 50% CO/50% Air- (a) 50% CO/50% N_2- (b) 50% synthetic syngas- (c) and 100% synthetic syngas- (d) -gas headspaces. 117
Figure 3.2: Gas composition (1) metabolite (2) and growth and pH (3) profiles for *P. thermoglucosidasius* DSM 6285 cultured under a 50% CO/50% Air gas headspace with no O_2 addition (a), 14 h O_2 addition (b) and 24 h O_2 addition (c). 122
Figure 3.3: Gas composition (1), metabolite (2) and growth and pH (3) profiles for *P. thermoglucosidasius* DSM 6285 cultured under a 50% CO/50% Air gas headspace with no O_2 addition (a), 38 h O_2 addition (b) and 48 h O_2 addition (c). 123
Figure 3.4: Gas composition (1) metabolite (2) and growth and pH (3) profiles for *P. thermoglucosidasius* DSM 6285 cultured under a 50% CO/50% Air gas headspace with no O_2 addition (a), 1%- (b), 3%- (c), 5%- (d) and 10%- (e) -O_2 addition at 38 h. 129
Figure 3.5: Predicted glucose metabolism pathway for *P. thermoglucosidasius*. 140
Figure 4.1: Core genome ML phylogeny of the genera *Parageobacillus* and *Saccharococcus*, with *G. thermodenitrificans* DSM 465^T as an outgroup. 150
Figure 4.2: COG supra-functional categorisation of proteins encoded on the genomes of *Parageobacillus*, *Saccharococcus* and *G. thermodenitrificans* DSM 465^T. 152
Figure 4.3: Protein conservation and function in the genera *Parageobacillus* and *Saccharococcus*. 154

Figure 4.4: Pan- and core-genome graphs. 157
Figure 4.5: Protein conservation and function in the species *P. thermoglucosidasius*, *P. toebii* and *S. caldoxylosilyticus*. 158
Figure 4.6: *Parageobacillus* and *Saccharococcus* antimicrobial biosynthetic loci. 165
Figure 4.7: *Parageobacillus* and *Saccharococcus* Hemicellulose Utilisation (HUS) loci. 169
Figure 4.8: Schematic diagram showing potential whole-cell applications of the genera *Parageobacillus* and *Saccharococcus*. 175

LIST OF TABLES

Table 1.1: Efficiency of hydrogen production processes. 24
Table 1.2: H_2 gas production processes associated costs. 25
Table 1.3: Hydrogenogenic CO-oxidising bacteria. 39
Table 2.1: Strains and plasmids used 53
Table 2.2: Restriction digestion reaction components 59
Table 2.3: Primers used 64
Table 2.4: Q5® High-Fidelity DNA polymerase reaction components 66
Table 2.5: Q5® High-Fidelity DNA polymerase thermal cycling conditions used for PCR reactions 67
Table 2.6: Primers used in the validation and screening of mutant gene constructs 75
Table 2.7: OneTaq DNA polymerase reaction components 76
Table 2.8: OneTaq PCR amplification thermal cycling conditions 77
Table 2.9: Strains and plasmids used 83
Table 2.10: Primers used 86
Table 2.11: NEBuilder HiFi DNA assembly reaction components 88
Table 2.12: Transformation efficiencies achieved for plasmid constructs used to transform *E. coli* DH5α 91
Table 2.13: Transformation efficiencies achieved and electroporation parameters used for *P. thermoglucosidasius* DSM 2542^T transformation with pPlasB constructs. 93
Table 2.14: Transformation efficiencies achieved for the various plasmid constructs used to transform *E. coli* 10β 98
Table 3.1: Composition of synthetic syngas stock and initial gas headspace mixtures 109

LIST OF SUPPLEMENTARY TABLES

Supplementary Table S3.1: Significant differences observed between variable initial gas headspace compositions. 184
Supplementary Table S3.2: Significant differences observed between H_2/CO yields across all experiments. 191
Supplementary Table S3.3: Significant differences observed between pre-hydrogenogenic phase oxygen addition conditions. 193
Supplementary Table S3.4: Significant differences observed between hydrogenogenic phase oxygen addition conditions. 199
Supplementary Table S3.5: Significant differences in H_2 levels observed between timepoints. 205
Supplementary Table S3.6: Significant differences observed between variable oxygen additions during peak hydrogenogenesis. 207
Supplementary Table S4.1: Genome metadata of the taxa analysed. 220
Supplementary Table S4.2:ANI (upper quadrant) and dDDH (lower quadrant) values between the analysed taxa. 222
Supplementary Table S4.3: The number of glycosyl hydrolases, polysaccharide lyases, carboxylesterases/lipases and proteins encoded on the genomes of the analysed taxa. 223
Supplementary Table S4.4: Presence of DNA polymerase genes on the genomes of the analysed taxa. 225
Supplementary Table S4.7: Bioremediation potential related genes on of the genomes of the analysed taxa. 233

LIST OF ABBREVIATIONS

CCS	Carbon capture strategies
SR	Steam reformation
Syngas	Synthesis gas
WGS	Water-gas shift
PSA	Pressure swing adsorption
PO	Partial oxidation
AO	Autothermal oxidation
Fd	Ferredoxin
ATP	Adenosine triphosphate
VFA	Volatile fatty acid
MEC	Microbial electrolysis cell
CODH	Carbon monoxide dehydrogenase
DNA	Deoxyribose nucleic acid
gDNA	Genomic DNA
LB	Luria-Bertani
mLB	modified Luria-Bertani
GC	Gas chromatography
HPLC	High performance liquid chromatography
HSD	Honest significant difference
ANOVA	Analysis of variance
TCA	Tricarboxylic acid
NADH	Nicotinamide adenine dinucleotide
RNA	Ribose nucleic acid

rRNA	Ribosomal RNA
SCO	Single copy orthologous
ML	Maximum likelihood
ANI	Average nucleotide identity
dDDH	Digital DNA-DNA hybridisation
GGDC	Genome-to-Genome Distance Calculator
LAP	Linear azole containing peptide
AHL	*N*-acyl-homoserine lactones
CAZymes	Carbohydrate-active enzymes
GH	Glycoside hydrolases
GT	Glycosyltransferases
CE	Carbohydrate esterases
PL	Polysaccharide lyases
HUS	Hemicellulose utilization system
RM	Restriction modification
CRISPR/Cas	Clustered regularly interspaced short palindromic repeats/CRISPR associated
2,3-BDO	2,3-butanediol
ALS	Acetolactate synthase
ALD	Acetolactate dehydrogenase

CHAPTER ONE: INTRODUCTION

1.1 Hydrogen

Hydrogen is the most abundant element in the universe and on Earth is present mainly in the form of water and organic compounds (Züttel, 2004; Argun and Kargi, 2011). Consequently, molecular hydrogen or hydrogen (H_2) gas is not readily available, encompassing less than 1% of earthbound hydrogen (Abdalla *et al.*, 2018). H_2 gas is colourless, odourless and non-toxic, with the lowest density of all gases yet it has one of the highest energy densities of all fuels (119.9-141.9 kJ/g) and upon combustion produces only water and heat (Das *et al*., 2001; Nikolaidis and Poullikkas, 2017; Abe *et al.*, 2019). Furthermore, H_2 gas is amphoteric, capable of acting as both an oxidising agent when reacting with metals, and as a reducing agent when reacting with non-metals (Abe *et al.*, 2019). The abundance of hydrogen and its reactivity, combustion and physical qualities, have rendered it an attractive commercial compound with various applications in industry and as an energy carrier (Ramachandran and Menon, 1998; Rathore *et al.*, 2019).

1.2 Hydrogen and the energy mix

The shift towards renewable and environmentally friendly energy sources has seen increased momentum to align with sustainable development goals. This is in an effort to limit global warming to below 2˚C in compliance with the 2015 Paris Agreement (UNFCCC, 2015; IRENA, 2019). However, non-renewable energy sources, including finite fossil fuels (coal, natural gas, and oil) still accounted for ~81% of global primary energy generation in 2022 (BP, 2022). Global energy demand is estimated to increase by 50% by 2050; presently however, fossil fuel combustion is the principle contributor to global primary energy generation (IRENA, 2019). The combustion of fossil fuels releases greenhouse gases (CO_2, CH_4, and N_2O) among various other pollutants, further exacerbating environmental impacts (Olivier *et al*., 2017; IRENA, 2019). Consequently, the integration of more sustainable and less harmful renewable energy technologies into the energy mix is occurring with increasing frequency and is estimated to account for ~60% of global energy demands by 2050 (IRENA, 2019). Among these, systems utilising H_2 gas as an energy carrier have been developed and implemented (REN21, 2019).

H_2 gas, similar to electricity, is considered an energy carrier and a secondary form of energy, produced through the conversion of a primary energy source, predominantly fossil fuels at present. However, H_2 gas can also be produced from various renewable energy sources (Balat, 2008; Abe *et al.*, 2019). Subsequently, H_2 gas can be converted into various forms of energy, including steam and heat through catalytic combustion, a heat source/sink through different chemical reactions, heat through flame combustion and electricity through electrochemical processes (Veziroğlu and Şahin, 2008). The ease of interconverting H_2 gas to and from other energy forms, its high energy density and appealing combustion characteristics make it an enticing option as an energy vector and fuel (Kapdan and Kargi, 2006; Elbeshbishy *et al.*, 2017; Nikolaidis and Poullikkas, 2017; Rathore *et al.*, 2019). Accordingly, H_2 gas is anticipated to partially assume the role of fossil fuels in energy generation and transport (Dincer and Acar, 2014).

1.3 Hydrogen and industry

H_2 gas is utilised extensively (~115 Megatons in 2019) in numerous spheres of industry in both pure and mixed gas form, where its physical and chemical properties are exploited (IEA, 2019b). The most prominent application of H_2 gas is as a reactant in hydrogenation processes. These processes are typically used to produce lower molecular weight- or saturated-compounds, crack hydrocarbons or remove sulphur and nitrogen from compounds (Ramachandran and Menon, 1998). The largest portion of industrial H_2 gas consumption (~33%) is directed towards petroleum oil refinement, wherein hydrocracking and hydrotreatment is used to upgrade oils through removal of heavy metals, sulphur, nitrogen and carbon impurities (Sahu *et al.*, 2015; IEA, 2019b). The chemical production industry, predominantly, involves ammonia production (27%) for nitrogen-based fertilizers and methanol production (11%) as a base chemical or as an alternative fuel. These activities account for the second and third major streams of H_2 gas consumption of pure H_2 gas (IEA, 2019b). Globally, ammonia production is achieved almost exclusively via the Haber–Bosch process, wherein H_2 and N_2 gas are passed over an iron catalyst at elevated temperatures and pressures (Ausfelder and Bazzanella, 2016). Methanol production (~11%), is achieved through various processes,

however, principally relies on the reformation of H_2-containing synthetic gas mixtures at elevated temperatures and pressures (Dalena *et al.*, 2018; IEA, 2019b). Aside from these dominant applications, H_2 gas is further employed in a range of alternative processes. It is used as a reducing agent in conjunction with CO in steel production (Liu *et al.*, 2021), an oxygen scavenger in food packaging and preservation processes (Dey and Neogi, 2019), as anticorrosive for spent nuclear fuel (Broczkowski *et al*., 2010), as coolant (Balitskii *et al*., 2000) and as a process reagent in the production of numerous base chemicals and reactions (Rasul *et al.*, 2022).

1.4 Hydrogen production

As hydrogen is available predominantly in the form of compounds, H_2 gas production requires the liberation of H^+ ions from these compounds, followed by their subsequent conversion to H_2 gas (Züttel, 2004; Winter, 2009). This process is not spontaneous and requires an energy investment along with appropriate conditions to form H_2 gas (Winter, 2009). A wide range of both renewable and non-renewable H_2 gas generation processes have been developed (Figure 1.1). Currently, the overwhelming majority of industrial H_2 gas production processes make use of dwindling, non-renewable energy sources, primarily coal and natural gas, with the further drawback that these processes produce undesirable greenhouse gases and other pollutants (Elbeshbishy *et al.*, 2017).

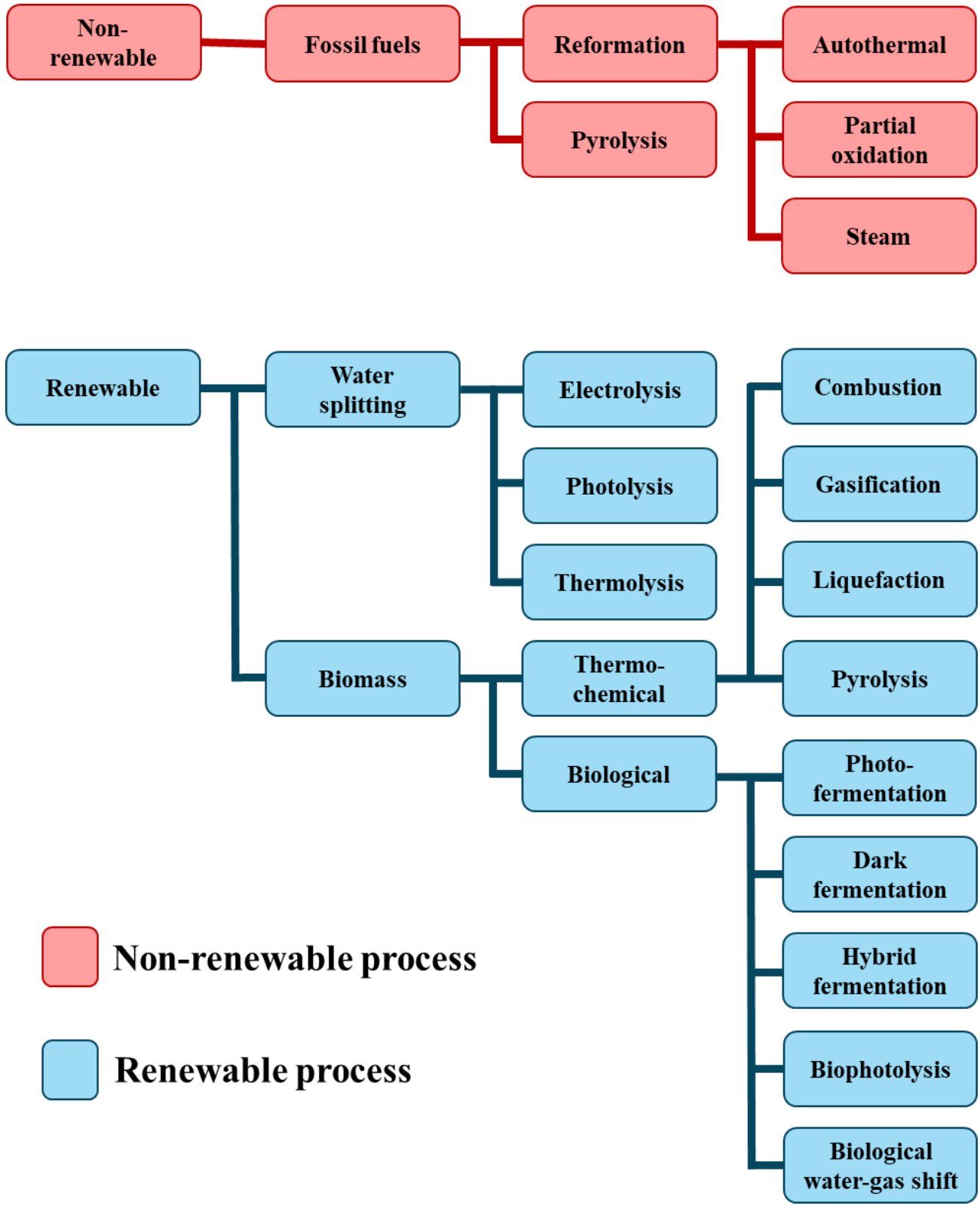


Figure 1.1: Hydrogen production strategies.

1.4.1 Non-renewable H_2 gas production

Currently ~95% of H_2 gas is produced through steam reformation, partial oxidation and autothermal oxidation (Dincer and Acar, 2014; Nikolaidis and Poullikkas, 2017; Singh and Dutta, 2018; IEA, 2019a). These processes utilise non-renewable fossil fuels (predominantly natural gas) as their primary feedstocks and result in the generation of greenhouse gas emissions (Dincer and Acar, 2014; Abdalla *et al.*, 2018). While not environmentally friendly, the predominant use of these technologies is the result of associated process efficiencies (Table 1.1), economic costs (Table 1.2) and availability of existing infrastructure, which is vastly more matured and established compared to emerging alternative renewable technologies (Nikolaidis and Poullikkas, 2017).

Table 1.1: Efficiency of hydrogen production processes.

Process	**CO/H_2 Conversion Efficiency (%)**	**Reference**
Steam reformation	74-85	Nikolaidis and Poullikkas, 2017
Partial oxidation	60-75	Nikolaidis and Poullikkas, 2017
Autothermal oxidation	60-75	Nikolaidis and Poullikkas, 2017
Biomass pyrolysis	35-50	Nikolaidis and Poullikkas, 2017
Biomass gasification	44-48	Staffell *et al.*, 2019
Bio-photolysis	12	Hallenbeck, 2014
Coal gasification	45-65	Staffell *et al.*, 2019
Dark fermentation	60-80	Nikolaidis and Poullikkas, 2017
Photofermentation	10	Hallenbeck, 2011
Electrolysis	40-60	Nikolaidis and Poullikkas, 2017
	51-67	Staffell *et al.*, 2019
Thermolysis	20-45	Nikolaidis and Poullikkas, 2017
Photo-electrolysis	0.06	Nikolaidis and Poullikkas, 2017

Table 1.2: H_2 gas production processes associated costs.

Process	Energy Source	Feedstock	Hydrogen Cost ($/kg)	Year referred to	Reference
SR with CCS	Fossil fuels	Natural gas	0.8	2024	Ayub et al, 2024
SR without CCS	Fossil fuels	Natural gas	0.66	2024	Ayub et al, 2024
Coal gasification with CCS	Fossil fuels	Coal	1.47	2024	Ayub et al, 2024
Coal gasification without CCS	Fossil fuels	Coal	1.17	2024	Ayub et al, 2024
Autothermal reforming with CCS	Fossil fuels	Natural gas	1.66	2022	Oni *et al*, 2022
Autothermal reforming without CCS	Fossil fuels	Natural gas	1.23	2022	Ayub et al, 2024
Methane pyrolysis	Internally generated steam	Natural gas	1.15	2024	Shokrollahi *et al*, 2024
Biomass pyrolysis	Internally generated steam	Woody biomass	1.25-2.20	2024	Zhang *et al*, 2024
Biomass gasification	Internally generated steam	Woody biomass	0.82-1.06	2024	Ayub et al, 2024
Direct biophotolysis	Solar	Water + Algae	2.13	2023	Qureshi *et al*, 2023
Indirect biophotolysis	Solar	Water + Algae	1.42	2023	Qureshi *et al*, 2023
Dark fermentation	Organic biomass	Organic biomass	2.3	2024	Bhatia *et al*, 2024

Process	Energy Source	Feedstock	Hydrogen Cost ($/kg)	Year referred to	Reference
Photofermentation	Solar	Organic biomass	3.5	2024	Bhatia *et al*, 2024
Solar PV electrolysis	Solar	Water	3.59	2023	Achour *et al*, 2023
Solar thermal electrolysis	Solar	Water	5.10	2019	Shen e*t al*, 2024
Wind electrolysis	Wind	Water	5.3	2018	Moshin *et al*, 2018
Nuclear electrolysis	Nuclear	Water	3.18-3.63	2018	Sorgulu and Dincer, 2018
Nuclear thermolysis	Nuclear	Water	1.63	2016	Hosseini and Wahid, 2016
Solar thermolysis	Solar	Water	7.98-8.40	2007	Nikolaidis and Poullikkas, 2017
Photo-electrolysis	Solar	Water	8	2024	Athia *et al*, 2024

1.4.1.1 Steam reformation

Steam Reformation (SR) is the primary contributor towards annual H_2 gas production, accounting for more than 50% of H_2 gas produced (Nikolaidis and Poullikkas, 2017; IEA, 2019b). SR encompasses the degradation of gaseous or liquid hydrocarbons in the presence of an oxidant or mix of oxidants, steam or CO_2, typically at temperatures above 700°C and under pressures between 3-25 bar (Eq. 1; Chaubey *et al.*, 2013; Zhang *et al.*, 2016). Reformation (Eq. 2) results in the production of synthesis gas (syngas), a mixture of CO_2, CO, H_2 and CH_4. Syngas is passaged to a water-gas shift (WGS) converter which utilizes steam and CO produced during reformation to further produce additional H_2 gas (Eq. 3) (Chaubey *et al.*, 2013). CO_2 produced during WGS conversion is captured through chemical absorption and the hydrogen rich gas is further purified through pressure swing

adsorption (PSA) or upgraded into value-added products through processes such as, the Fischer-Tropsch process (Hosseini and Wahid, 2016; Nanda *et al*, 2017). PSA is a technique generally used to purify gases, relying on selective high-pressure adsorption (retention) of a target gas, while non-target gas components are passaged out of the system (Muin *et al.*, 2020). Adsorbed target gas components are subsequently desorbed through pressure reduction, facilitating collection of the purified target gas.

$$C_nH_m + nH_2O \rightarrow nCO + (n + \frac{m}{2})H_2 \qquad \Delta H<0 \qquad (1)$$

$$CH_4 + H_2O \rightarrow CO + 3H_2 \qquad \Delta H = -206\text{kj/mol at } 15°C \qquad (2)$$

$$CO + H_2O \rightarrow CO_2 + H_2 \qquad \Delta H = +41.2\text{kj/mol at } 15°C \qquad (3)$$

1.4.1.2 Partial oxidation

Partial oxidation (PO) is the second most applied H_2 gas production process (Chaubey *et al.*, 2013). Unlike SR, PO requires the use of O_2, steam and may make use of a variety of hydrocarbon feedstocks (typically coal) to produce H_2 gas. This is done through gasification and combustion of the hydrocarbon with available oxygen, which may be conducted with (Eq. 4) or without the presence of a catalyst (Eq. 5) (Holladay *et al.*, 2009). Through combustion of the hydrocarbon feedstock at high temperatures (950-1500°C) with limited O_2, the hydrocarbon is only partially oxidized, forming syngas rich in H_2, and CO in the reformer. The syngas is then subjected to a desulfurization step prior to WGS conversion (Eq. 3) and H_2 gas purification by PSA (Nikolaidis and Poullikkas, 2017).

$$C_nH_m + \frac{n}{2}O_2 \rightarrow nCO + (\frac{m}{2})H2 \qquad \Delta H>0 \qquad (4)$$

$$C_nH_m + nH_2O \rightarrow nCO + (n + \frac{m}{2})H_2 \qquad \Delta H>0 \qquad (5)$$

1.4.1.3 Autothermal oxidation

Autothermal oxidation (AO) is the combined application of SR and PO, where the heat generated by exothermic PO is utilized by the application of SR to increase H_2 gas production (Abdalla *et al.*, 2018). In this technique, both steam and O_2 are injected into the reformer, allowing simultaneous oxidation and reformation to

occur (Eq. 6; Nikolaidis and Poullikkas, 2017). The hydrocarbon is initially oxidized to produce syngas, of which the CO component along with the injected steam further form CO_2 and H_2 gas via the WGS reaction (Eq. 3; Chaubey *et al.*, 2013). Residual CO_2 is then captured via carbon absorption and the H_2 gas purified by PSA (Nikolaidis and Poullikkas, 2017).

$$C_nH_m + \frac{n}{2}H_2O + \frac{n}{4}O_2 \rightarrow nCO + (\frac{n}{2} + \frac{m}{2})H_2 \quad (6)$$

1.4.2 Renewable H_2 gas production

Renewable approaches account for ~ 5% of H_2 gas production. However, there is a drive towards expanding renewable H_2 gas generation processes to assume the central roles that fossil fuel-based technologies currently occupy (Abdalla *et al.*, 2018; Singh and Dutta, 2018). As such, numerous renewable H_2 gas producing technologies have been developed (Figure 1.1). However, these tend to be financially uncompetitive with non-renewable means of H_2 gas production (Table 1.2; Nikolaidis and Poullikkas, 2017). Mainstream applications of these processes have gravitated towards the use of electrolysis and bioenergy systems to produce H_2 gas (Bartels *et al*., 2010; Hosseini and Wahid, 2016).

1.4.2.1 Electrolysis

Electrolysis strategies account for ~90% of renewable H_2 gas production strategies and encompasses the application of electricity to split water (Bhandari *et al*., 2014; Singh and Dutta, 2018). Electrolysis is carried out through the use of electrolyzers, supplied with water and an electric current, which may be derived from any form of renewable or non-renewable energy (Ji *et al*., 2012; Hosseini and Wahid, 2020). Electrolyzers facilitate the conversion of electrical to chemical energy in the form of H_2 and O_2 gas (Eq.7; Holladay *et al.*, 2009). Three forms of electrolyzers are commonly utilised at commercial and industrial scales, namely alkaline, polymer electrolyte membrane and high-temperature solid oxide electrolyzers (Bhandari *et al*., 2014).

$$H_2O \xrightarrow{Electrical\ energy} H_2 + \frac{1}{2}O \quad (7)$$

Despite the environmental appeal of electrolysis technologies, associated operational and construction costs are substantially higher than those presented by dominant non-renewable technologies (Table 1.2; Balat, 2008; Chaubey *et al.*, 2013; Nikolaidis and Poullikkas, 2017). However, more cost effective alternative H_2 gas production strategies, such as bioenergy strategies are under development which may further contribute to H_2 gas production efforts.

While electrolysis strategies are expected to facilitate the transition from the use of fossil fuels to clean H_2 gas generation in the near-term, bioenergy technologies, are predicted to evolve as the principle means for future H_2 gas generation (Hosseini and Wahid, 2016; Staffell *et al.*, 2019). Bioenergy technologies are highly versatile, capable of utilizing a variety of feedstocks including dedicated crops, agricultural and industrial wastes and wastewater sources in a range of thermochemical and biological processes (Holladay *et al.*, 2009; Hosseini and Wahid, 2016; Staffell *et al.*, 2019).

1.4.2.2 Thermochemical approaches

Thermochemical approaches essentially encompass the use of traditional fossil-fuel reformation processes, substituting fossil fuels with sustainably sourced organic feedstocks to produce H_2 rich gas (Zhang *et al.*, 2016). This typically involves the initial combustion, liquefaction, pyrolysis or gasification of the feedstock, followed by further reformation of the products to high purity H_2 gas. Due to pollutant emissions and strict operational requirements, combustion and liquefaction approaches are not widely employed. Instead, pyrolysis and gasification are the most utilised approaches, producing H_2 rich gas and other biofuels, which undergo further processing (Ni *et al.*, 2006; Nikolaidis and Poullikkas, 2017; Liu *et al.*, 2019).

Pyrolysis

Pyrolysis involves the heating of biomass to between 650-800 K at 0.1-0.5 MP in the absence of air, which converts the biomass into liquid oils, solid charcoal and producer gas (Eq. 8; Ni *et al.*, 2006; Nikolaidis and Poullikkas, 2017). Slow pyrolysis results in the production of predominantly charcoal and as such is not normally considered for H_2 gas production. Fast pyrolysis, however, where the biomass is heated rapidly to high temperatures to form vapour and then condensed to bioliquid, may be used in the production of H_2 gas. However, it is more targeted towards non-H_2 gas containing biofuels (Ni *et al.*, 2006). The pyrolysis process is typically conducted without oxygen, except for when partial oxidation is utilized to generate the heat which facilitates the highly energy intensive process (Parthasarathy and Narayanan, 2014). Steam reformation may be conducted along with the WGS reaction following pyrolysis to maximise H_2 gas production from the producer gas (Nikolaidis and Poullikkas, 2017).

$$\text{Biomass pyrolysis} \rightarrow H_2 + CO + CO_2 + \text{hydrocarbon gases} + \text{tar} + \text{char} \quad (8)$$

Biomass gasification

Biomass gasification is the process of converting biomass into gaseous fuel/syngas in a gaseous medium such as air, oxygen and/or steam, typically between 500-

1400°C and pressures up to 33 bar (Eq. 9 and 10; Nikolaidis and Poullikkas, 2017). Biomass gasification is a variation of pyrolysis and is based on the partial oxidation of biomass into producer gas, char and tar. As with pyrolysis, the H_2 rich producer gas formed during gasification may then undergo reformation to maximize H_2 gas yield (Ni *et al.*, 2006). Both processes display similar CO/H_2 conversion efficiencies, peaking around 50% (Table 1.1), are mature and are considered to present easy system management and relatively low cost (Table 1.2; Nikolaidis and Poullikkas, 2017; Liu *et al.*, 2019).

$$\text{Biomass} + \text{Air} \xrightarrow{\square} H_2\ CO_2 + CO + N_2 + CH_4 + \text{other CHs} + \text{tar} + H_2O + \text{char}$$

(transformation to syngas using air) (9)

$$\text{Biomass} + \text{steam} \xrightarrow{\square} H_2 + CO_2 + CO + CH_4 + \text{other CHs} + \text{tar} + \text{char}$$

(transformation to syngas using steam) (10)

Although thermochemical H_2 gas production technologies are highly developed and relatively low cost to employ, they harbour numerous disadvantages, rendering them uncompetitive with traditional H_2 gas production technologies (Acar and Dincer, 2014). These technologies suffer from low thermal efficiency due to the energy required to vaporize the moisture held by biomass feedstocks (Holladay *et al.*, 2009). Furthermore tar, a by-product of these processes, has negative effects on infrastructure through slugging and fouling of equipment (Dincer, 2012). Moreover, yields from these technologies are low compared to dominant non-renewable approaches (Demirbas, 2016; Nikolaidis and Poullikkas, 2017). Thus, these technologies are presently unable to economically produce H_2 gas at a large scale, rendering them unviable despite their more environmentally friendly aspects (Acar and Dincer, 2014).

1.4.2.3 Biological approaches

Biological H_2 gas production strategies exploit the intrinsic capacities of microorganisms to produce H_2 gas. These approaches present numerous advantages which identify them as ideal H_2 gas production processes (Hosseini and Wahid, 2016). Biological H_2 gas production strategies utilise a wide variety of feedstocks,

including both dedicated agricultural resources as well as organic waste materials from a broad range of sources (Das *et al.*, 2001; Hosseini and Wahid, 2016). These feedstocks can be viewed as inexhaustible, and as they further contribute to waste recycling, are considered substantially more environmentally friendly and sustainable than strategies reliant on fossil fuels (Das *et al.*, 2001). Furthermore, as the biological processes utilised are (predominantly) undertaken at ambient temperature and pressure, these technologies are minimally energy demanding, especially compared to dominant H_2 gas production processes (Nikolaidis and Poullikkas, 2017). Biological H_2 gas production has been observed in numerous microorganisms, including archaea, bacteria and green algae, which utilise either light-dependent or light-independent processes to produce H_2 gas (Das *et al.*, 2001; Kapdan and Kargi, 2006; Holladay *et al.*, 2009; Rittmann *et al.*, 2015; Van Niel, 2016; Liu *et al.*, 2019). Light-dependent H_2 gas production processes include biophotolysis and photofermentation, whilst light-independent processes include dark fermentation, microbial electrolysis and the water-gas shift reaction (WGS) (Dinesh *et al.*, 2018).

Biophotolysis

Biophotolysis is a light-dependent process conducted by algae and cyanobacteria, whereby solar energy is converted and stored as H_2 gas, using hydrogen derived from water lysis (Eq. 11; Arimi *et al.*, 2015). Biophotolysis proceeds as either a direct process, where H_2 and O_2 gas are produced simultaneously, or as an indirect process, where H_2 and O_2 gas are produced under spatially or temporally separated conditions, respectively (Manish and Banerjee, 2008). In direct biophotolysis, carried out by green algae, H_2 gas is produced via two steps. Initial photochemical oxidation of water is conducted by photosystem II mediated photon absorption in the thylakoid membrane, accompanied by obligate O_2 generation (Ghiasian, 2019). The liberated electrons are then passaged to an electron carrier ferredoxin (Fd) through the plastoquinones, cytochromes b_6 and f, plastocyanin and photosystem I using further light derived energy, simultaneously generating ATP. Electrons are finally transferred from the Fd to a hydrogenase or nitrogenase enzyme, which evolves H_2 gas (Das and Veziroğlu, 2008; Show *et al.*, 2018).

$$2H_2O + \text{Solar Energy} \rightarrow 2H_2 + O_2 \quad (11)$$

H_2 gas production through indirect biophotolysis is carried out by cyanobacteria and comprises two distinct stages (Nikolaidis and Poullikkas, 2017). The first stage is similar to direct biophotolysis. In the second stage, however, electrons transported via the photosystem I and the electron carrier Fd are instead targeted towards synthesis of cellular biomass and storage compounds (Eq. 12). Storage compounds produced then undergo fermentation, liberating electrons which are transported to hydrogenase or nitrogenase enzymes and used to reduce available H^+ ions to produce H_2 gas (Eq. 13; Ni *et al.*, 2006).

$$12H_2O + 6CO_2 \xrightarrow{light} C_6H_{12}O_6 + 6O_2 \quad (12)$$

$$C_6H_{12}O_6 + 12H_2O \xrightarrow{light} 12H_2 + 6CO_2 \quad (13)$$

Application of biophotolysis technologies at industrial scale is inherently problematic as the hydrogenases involved are strongly suppressed by the presence of oxygen, a concomitant product of the photosynthetic H_2 gas production process (Show *et al.*, 2018). Light conversion efficiency further dictates potential process efficiency. Although maximum potential light conversion is estimated to peak at ~12.2%, achieved conversion efficiencies have not exceeded 3% (Hallenbeck, 2011). Further efficiency loss through light absorption through reactor units is another major limitation to the scalability of this technology, as large volume reactor units suffer from increasingly lower light conversion efficiency (Show *et al.*, 2018; Ghiasian, 2019). In practice photolysis typically produces low H_2 gas production rates (0.07 mmol $H_2.h^{-1}.L^{-1}$) that is neither practical nor competitive for industrial scale application (Rathore *et al.*, 2019; Pareek *et al.*, 2020).

Photofermentation

Photofermentation is another light dependent hydrogenogenic process undertaken by non-sulphur purple and green bacteria, which utilise carbon sources such as simple sugars, organic and volatile fatty acids (VFAs) and solar energy to produce H_2 and CO_2 gas (Eq. 14; Dinesh *et al.*, 2018). As with biophotolytic approaches, photofermentation presents several key hinderances to industrial scale application. Theoretical yields for photofermentation stand as high as 12 mol H_2/ mol hexose and has been achievable (Koku *et al.*, 2002). However, H_2 gas productivity through photofermentation is, like biophotolysis, limited by numerous factors including low light conversion efficiency (max ~10%; Table 1.1), substrate pre-treatment requirements, low H_2 gas productivities and enzyme sensitivity to both oxygen and nitrogen (Van Niel, 2016; Sharma *et al.*, 2020; Abdelsalam *et al.*, 2023). These limitations continue to restrict large scale H_2 gas production application (Argun and Kargi, 2011; Hay *et al.*, 2013).

$$CH3COOH + 2H_2O \xrightarrow{light} 4H_2 + 2CO_2 \qquad (14)$$

Dark fermentation

Dark fermentation is a light-independent hydrogenogenic process used by a range of archaea, bacteria and microalgae under anaerobic conditions. Particular research focus has been placed on acidogenic and acetogenic bacteria and methanogenic archaea (Hay *et al.*, 2013; Kumar and Mohan, 2018; Yasin *et al.*, 2019). Dark fermentation approaches make use of a range of organic substrates that are metabolized by various fermentative pathways (Hay *et al.*, 2013). These pathways generally produce VFA intermediates and by-products, such as acetic acid, in conjunction with H_2 gas production (Eq. 15; Liu *et al.*, 2019).

$$C_6H_{12}O_6 + 2H_2O \rightarrow 2CH_3COOH + 4H_2 + 2CO_2 \qquad (15)$$

When applied in isolation, photo- and dark-fermentation technologies show numerous challenges, especially regarding poor yields. While a theoretical maximum yield of 12 mol H_2/ mol glucose is thought to be achievable, in practice, maximum achievable yields thus far for mesophilic fermentations have been 4 mol

H_2/ mol glucose (Sarangi and Nanda, 2020). This can in part be attributed to a build-up of organic acid by-products (typically acetic-, butyric- or propionic-acid) of the respective fermentation processes (Osman *et al.*, 2020; Sharma *et al.*, 2020). Build-up of organic acids can result in a decrease in medium pH, stimulating a shift from hydrogenogenic acidogenesis towards solventogenesis or methanogenesis, reducing H_2 gas yields (Sarkar *et al.*, 2017). Sequential photo- and dark fermentation approaches have attempted to mitigate these shortfalls (discussed below).

Microbial electrolysis cells

Microbial electrolysis cells (MECs) are essentially modified microbial fuel cells, where a biodegradable substrate is non-spontaneously oxidized at an anode under anaerobic conditions by an exoelectrogen supplied with an external voltage (Eq. 16; Van Niel, 2016). Exoelectrogens are taxa capable of transferring electrons extracellularly, typically through direct contact with an oxidant through outer membrane c-type cytochromes or nanowires (Kumar *et al.*, 2015). In principle, MECs facilitate organic substrate oxidation by exoelectrogenic taxa under a supply of low voltage at an anode, which serves as the oxidant (Van Niel, 2016). Electrons liberated during substrate oxidation are transferred to the anode and shuttled to a cathode. Protons (H^+ ions) produced during substrate oxidation travel from the anode chamber across a semipermeable membrane to a cathode where they are reduced, using electrons supplied by the cathode, to produce H_2 gas (Eq. 17; Holladay *et al.*, 2009; Van Niel, 2016). Studies concerning the application of electrogenic taxa in MECs have predominantly utilised mixed cultures predominated by Proteobacteria or pure cultures of *Geobacter sulfurreducens* (Van Niel, 2016; Liu *et al.*, 2019).

$$CH_3CO_2^- + 4H_2O \rightarrow 2HCO_3^- + 9H^+ + 8e^- \quad (16)$$

$$2H^+ + 2e^- \rightarrow 2H_2 \quad (17)$$

MECs present a rapidly developing alternative biological hydrogenogenic approach which present several advantages over other approaches, principally, their wide substrate ranges and low energy investment requirement compared to conventional

electrolysis (Zhang and Angelidaki, 2014; Van Niel, 2016). However, two key restrictions, low volumetric hydrogen yield upon upscaling and electrode material costs, have emerged as hurdles limiting upscale application of this technology (Van Niel, 2016).

Integrated processes

Both photo- and dark- fermentation processes may be undertaken simultaneously within the same reactor unit or in a sequential manner (Singh and Wahid, 2015). Although, single stage hybrid dark-/photo-fermentations using co-cultures present higher potential H_2 gas yields (12 mol H_2/ mol hexose) than photo- or dark-fermentations conducted in isolation, their H_2 gas evolution times are generally substantially slower than those of dark fermentations (Argun and Kargi, 2011). Furthermore, obtaining maximal H_2 gas production is challenging due to the varied conditions required for the respective fermentation processes being undertaken and the nutritional and environmental conditions required by the different microbial taxa employed (Pachapur *et al.*, 2015). Consequently, limited research has been conducted on hydrogenogenic mixed fermentation systems (Singh and Wahid, 2015). Despite this, yields as high as 6.6 mol H_2/ mol glucose have been achieved in fed-batch cultures, using a co-culture of *Clostridium butyricum* (dark fermentative) and *Rhodobacter* sp. (photo-fermentative) (Yokoi *et al.*, 1998). Although, typically co-culture applications have produced substantially lower (<2 mol H_2/ mol hexose) yields (Zagrodnik and Łaniecki, 2017).

Sequential dark-/photo-fermentations, have received substantially more attention than hybrid single stage strategies and are typically characterised by a two-stage fermentation process (Pachapur *et al.*, 2015). Generally, a primary stage dark fermentation is performed followed by a secondary stage photofermentation (Argun and Kargi, 2011). In the primary stage anaerobic bacteria are used to convert the primary feedstock to a low yield of H_2 gas and a bulk component of short chain organic acids and other simple fermentation metabolites. These products are then fed into a second stage where they are metabolized through photofermentation pathways, to produce H_2 gas (Sharma *et al.*, 2020). Despite high theoretical yields of 12 mol H_2/ mol hexose, achieved yields, although higher than standalone

processes, still remain low (Zagrodnik and Duber, 2024). Cai *et al*, (2019) have exemplified this by improving mixed-culture dark fermentation yields of 1.14 mol H_2/ mol glucose to 4.25 mol H_2/ mol glucose through second stage mixed-culture photofermentation. These hybrid processes produce substantially higher H_2 gas yields than single fermentation processes and present a promising mitigation strategy for the challenges presented by standalone technologies (Show *et al.*, 2018).

The (biological) water-gas shift reaction

While the aforementioned approaches have been key focus areas for biological H_2 gas production studies, there are additional but less explored biological approaches that show promise for renewable industrial scale H_2 gas production (Mohr *et al.*, 2018a; Mohr *et al.*, 2018b). One potential route with growing appeal is the biological WGS reaction. The WGS mediates processing of industrial waste gases and syngases containing variable CO levels to produce value added products and biofuels, such as H_2 (Rittmann *et al.*, 2015; Akhlaghi and Najafpour-Darzi, 2020). CO-rich gases such as natural gas, syngas (produced during oil/shale/natural gas refining and coal gasification) and various industrial waste gases (e.g. from steel milling) currently present appealing and abundant feedstock streams, with up to 10-60% v/v CO (Dürre and Eikmanns, 2015; Alfano and Cavazza, 2018). Furthermore, syngas may also be produced from the gasification of organic biomass (e.g. lignocellulosic material) and solid wastes, which is gaining traction as a renewable and cleaner alternative to fossil fuel derived syngas (Latif *et al.*, 2014).

Several bacterial and archaeal species across a broad taxonomic spectrum have been identified to metabolize CO anaerobically, including methanogenic archaea and carboxydotrophic, acetogenic, sulphate- and ferric iron-reducing bacteria (Ragsdale, 2004; Rittmann *et al.*, 2015). These taxa utilise several strategies to proliferate utilising CO. One strategy in particular, the biologically mediated WGS reaction, catalysed by a restricted subset of bacteria and archaea is an enticing route for H_2 production (Rittmann *et al.*, 2015; Mohr *et al.*, 2018a). The WGS reaction is an equilibrium dependant reaction, whereby CO and water are reversibly converted into CO_2 and H_2 (Eq. 3), by either inorganic or organic catalysts (Schoelmerich and

Müller, 2019; Chen and Chen, 2020). The inorganic WGS reaction is achieved at elevated temperatures (180-450˚C) using metal-based catalysts and is typically conducted as a means to reduce CO content in process gasses following hydrogenogenic reformation processes (Chen and Chen, 2020). However, application of the inorganic WGS reaction is relatively limited due to the cost-prohibitive nature of the catalysts employed, poor conversion efficiencies and low potential energy yields of the gas feedstocks employed, resulting in a deficit in efficient processing of small-scale syngas and waste gas streams (Haddad *et al.*, 2014; Chen and Chen, 2020). The biological WGS reaction provides a more cost-effective alternative to traditional inorganic WGS reaction applications, capable of mediating low temperature and pressure production of H_2 from highly variable grade syngases (Haddad *et al.*, 2014). The biological WGS reaction is a light-independent reaction that utilises CO as a source of energy and water as an electron acceptor to produce H_2 and CO_2 gas at ambient temperatures (Eq. 3; Ni *et al.*, 2006; Schoelmerich and Müller, 2019). Despite research concerning the biologically mediated WGS being in its infancy compared to mainstream biological H_2 gas production processes, there is growing interest in developing this technology as a means for cost-effective, clean and high yield H_2 gas production process (Levin *et al*., 2004; Alfano and Cavazza, 2018).

A limited taxonomic set of hydrogenic CO-oxidising bacteria have been identified and their capacity to undertake the WGS reaction variably characterised (Alfano and Cavazza, 2018). These include predominantly mesophilic members of the phylum Proteobacteria and thermophilic members of the phylum Firmicutes (Table 1.3; Fukuyama *et al.*, 2020). Of these taxa two model organisms, namely the CO-oxidizing mesophile *Rhodospirillum rubrum* and the carboxydotrophic thermophile *Carboxydothermus hydrogenoformans*, have been extensively studied and characterized (Younesi *et al.*, 2008; Haddad *et al.*, 2014; Alfano and Cavazza, 2018).

Table 1.3: Hydrogenogenic CO-oxidising bacteria. Adapted from Fukuyama *et al*, (2020).

Growth temperature	Phylum	Class	Order	Family	Organism
Mesophilic	Proteobacteria	Gammaproteobacteria	Enterobacterales	*Enterobacteriaceae*	*Citrobacter amalonaticus*
		Alphaproteobacteria	Rhizobiales	*Methylocystaceae*	*Pleomorphomonas carboxyditropha*
				Bradyrhizobiaceae	*Rhodopseudomonas palustris*
			Rhodospirillales	Rhodospirillaceae	*Rhodospirillum rubrum*
		Betaproteobacteria	Burkholderiales	unclassified Burkholderiales	*Rubrivivax gelatinosus*
Thermophilic	Firmicutes	Clostridia	Thermoanaerobacterales	*Thermoanaerobacteraceae*	*Caldanaerobacter subterraneus*
					Calderihabitans maritimus
					Carboxydothermus hydrogenoformans
					Moorella sp.
					Thermoanaerobacter sp.
			Clostridiales	*Peptococcaceae*	*Desulfotomaculum nigrificans*
					Therminicola sp.
				Family XVI	*Carboxydocella* sp
		Bacilli	Bacillales	*Bacillaceae*	*Parageobacillus thermoglucosidasius*
		Thermolithobacteria	Thermolithobacterales	*Thermolithobacteraceae*	*Thermolithobacter carboxydivorans*
		Negativicutes	Selenomonadales	*Sporomusaceae*	*Thermosinus carboxydivorans*
Hyperthermophilic	Dictyoglomi	Dictyoglomia	Dictyoglomales	*Dictyoglomaceae*	*Dictyoglomus carboxydivorans*

R. rubrum is an extensively studied anaerobic, mesophilic, purple non-sulphur bacterium capable of photoautotrophic, photoheterotrophic and chemoheterotrophic growth depending on growth conditions (Alfano and Cavazza, 2018). Complementary to the wide metabolic capability of this taxon, it is capable of utilising CO as a sole energy source under dark anaerobic growth conditions via the WGS reaction (Kerby *et al.*, 1995). While not capable of utilising CO as a sole carbon source, it is capable of conducting hydrogenogenesis when grown on a variety of supplemented carbon sources including acetate, malate and formate (Najafpour *et al.*, 2006). CO/H_2 conversion yields by *R. rubrum* of up to 98% have been achieved using acetate as a carbon source and a 55% CO, 10% CO_2, 20% H_2, 15% Ar synthetic syngas mix, and further demonstrated H_2 evolution efficiencies of up to 16 mmol $H_2.g_{substrate}^{-1}.cell^{-1}.h^{-1}$ (Najafpour and Younesi, 2007; Younesi *et al.*, 2008). *C. hydrogenoformans* is an anaerobic, thermophilic bacterium that can produce H_2 via the WGS reaction and is capable of growth using CO as a sole carbon source (Svetlichny *et al.*, 1991; Haddad *et al.*, 2014). Despite its potential for use with synthetic syngas feedstocks, it has only been used in pure CO-fed hydrogenogenic studies (Alfano and Cavazza, 2018). Using pure CO as a feedstock a CO/H_2 conversion yield of up to 95% has been achieved (Haddad *et al.*, 2014). Furthermore, CO conversion rates achieved by *C. hydrogenoformans* in a hollow fibre immobilized membrane reactor fed with pure CO have reached 3.5 mol per CO per g(volatile suspended solid).d^{-1} at 0.55 mM CO (Zhao *et al.*, 2013).

The application of carboxydotrophic thermophiles in particular, such as *C. hydrogenoformans* has emerged as a promising route for H_2 gas production. Their natural occurrence in extreme high temperature environments and the limited availability of potential energy sources, necessitate efficient metabolic strategies for proliferation (Zeigler, 2014; Hosseini *et al.*, 2015). Additionally, standard metabolism under high temperature conditions confers several advantageous qualities, including increased gas diffusion, growth, reaction and mass transfer rates and decreased gas solubility. Taken together these factors render these thermophiles as ideal mediators of processes like the biological WGS reaction (Hosseini *et al.*, 2015; Alfano and Cavazza, 2018). However, the vast majority of carboxydotrophic thermophiles that conduct the WGS reaction are obligate anaerobes (Fukuyama *et*

al., 2020). Cultivation and reactor operation under strictly anaerobic conditions necessitates more complex and costly reactor configurations and practices (Morya *et al*., 2023). Furthermore, biomass accumulation rates of strictly anaerobic bacterial taxa are much lower than aerobes, delaying the build-up of catalytically active biomass (Russel and Cook, 1995). However, the recent characterisation of the WGS reaction carried out by the facultatively anaerobic thermophile *Parageobacillus thermoglucosidasius* has presented an appealing avenue for the exploration of high temperature, O_2 tolerant hydrogenogenic carboxydotrophic biotechnology strategies (Mohr *et al.*, 2018a; Inoue *et al.*, 2019; Aliyu *et al.*, 2021; Imaura *et al.*, 2023).

1.5 *Parageobacillus*

The genus *Parageobacillus* was delineated from the genus *Geobacillus* on the basis of whole genome phylogenomic approaches (Aliyu *et al.*, 2016; Aliyu *et al.*, 2018). Presently, the genera *Geobacillus* and *Parageobacillus* comprise twelve and six validly described species (Parte *et al.*, 2020). Members of both genera are broadly characterized as thermophilic Gram-positive, aerobic or facultatively anaerobic, rod-shaped, spore forming bacteria, capable of growth over temperatures ranging from 37 to 80˚C (Najar and Thakur, 2020). These taxa have been readily isolated across disparate, globally dispersed environments, including high temperature environments, such as hot springs (Brumm *et al.*, 2015), oil wells (Nazina *et al.*, 2005), hot composts (Poli *et al.*, 2006) and geothermal sites (Omae *et al.*, 2019; Imaura *et al.*, 2023), as well as moderate and low temperature environments, such as temperate soils (Marchant *et al.*, 2002), plant matter (Sung *et al.*, 2002; Valladares Juárez *et al.*, 2009) and marine sediments (Inoue *et al.*, 2019). Consequently, representative taxa have demonstrated highly versatile metabolic capabilities (Hussein *et al.*, 2015; Najar and Thakur, 2020; Khaswal *et al.*, 2022). Due to their thermophilic nature and the concomitant thermostability of their enzymes, extensive research efforts have been directed towards characterizing the biotechnological potential of the genus as well as its sister genus *Geobacillus*. Consequently, a number of thermostable enzymes including proteases, xylanases, amylases, and lipases have been characterised from members of the two genera and

highlighted for their commercial value (Hussein *et al.*, 2015; Najar and Thakur, 2020; Khaswal *et al.*, 2022). In terms of whole cell applications of these taxa, various studies have emphasized their potential role in the production of novel antimicrobials (Zebrowska *et al.*, 2022), the bioremediation of petroleum compounds (Banat and Marchant, 2011; Peng *et al.*, 2023), as well as their application in the production of biofuels such as ethanol and butanol (Cripps *et al.*, 2009; Lin *et al.*, 2014). One novel area of biotechnological potential harboured by a subset of *Parageobacillus* is the capacity to produce H_2 gas via the WGS reaction (Sung *et al.*, 2002; Mohr *et al.*, 2018a; Najar and Thakur, 2020).

1.6 *Parageobacillus* and the water-gas shift reaction

The capacity for H_2 production by *P. thermoglucosidasius* DSM 2542^T was first demonstrated by growing this taxon in an aerobic atmosphere containing CO (Mohr *et al.*, 2018a). This hydrogenogenic activity was shown to be mediated by a carbon monoxide dehydrogenase (CODH) and a group 4a H_2-evolving hydrogenase encoded on the genome of the strain (Mohr *et al.*, 2018a; Adachi *et al.*, 2020). These enzymes are predicted to form a complex, facilitating CO oxidation and H_2 evolution (Figure 1.2), as observed for the model WGS conducting taxa *R. rubrum* and *C. hydrogenoformans* (Fox *et al.*, 1996; Soboh *et al.*, 2002; Mohr *et al.*, 2018a). The WGS machinery are encoded within a 29 kb chromosomally located region, by three co-localised operons, comprising three, ten and two genes, respectively (Mohr *et al.*, 2018a; Mohr *et al.*, 2018b). The first operon (*cooCSF*), codes for a carbon monoxide dehydrogenase, with *cooC*, *cooS* and *cooF* encoding a CO dehydrogenase maturation factor, the CO dehydrogenase catalytic subunit and a CO dehydrogenase Fe-S protein that transfers electrons between the CODH and the associated hydrogenase, respectively (Mohr *et al.*, 2018a). The second operon (*phcABCDEFGHIJ*) codes for major (PhcG) and minor (PhcE) catalytic subunits of a group 4a H_2-evolving [NiFe]-hydrogenase, as well as proteins involved in the formation and maturation of the mature hydrogenase (Mohr *et al.*, 2018a; Adachi *et al.*, 2020). The third operon (*phcKL*), encodes HypA- (PhcK) and HypB-like (PhcL) proteins, predicted to participate in incorporation of nickel into the catalytic core of the mature hydrogenase enzyme (Chan *et al.*, 2012). These enzymes are

predicted to form a complex, facilitating CO oxidation and H_2 evolution, as observed for the model WGS conducting taxa *R. rubrum* and *C. hydrogenoformans* (Fox *et al.*, 1996; Soboh *et al.*, 2002; Mohr *et al.*, 2018a).

Through this complex, CO is oxidised by the CODH to CO_2, liberating electrons which are passaged via a ferredoxin or ferredoxin-like protein to the hydrogenase enzyme, which catalyses the reduction of H^+ ions, using the liberated electrons to produce H_2 gas (Figure 1.2; Alfano and Cavazza, 2018).

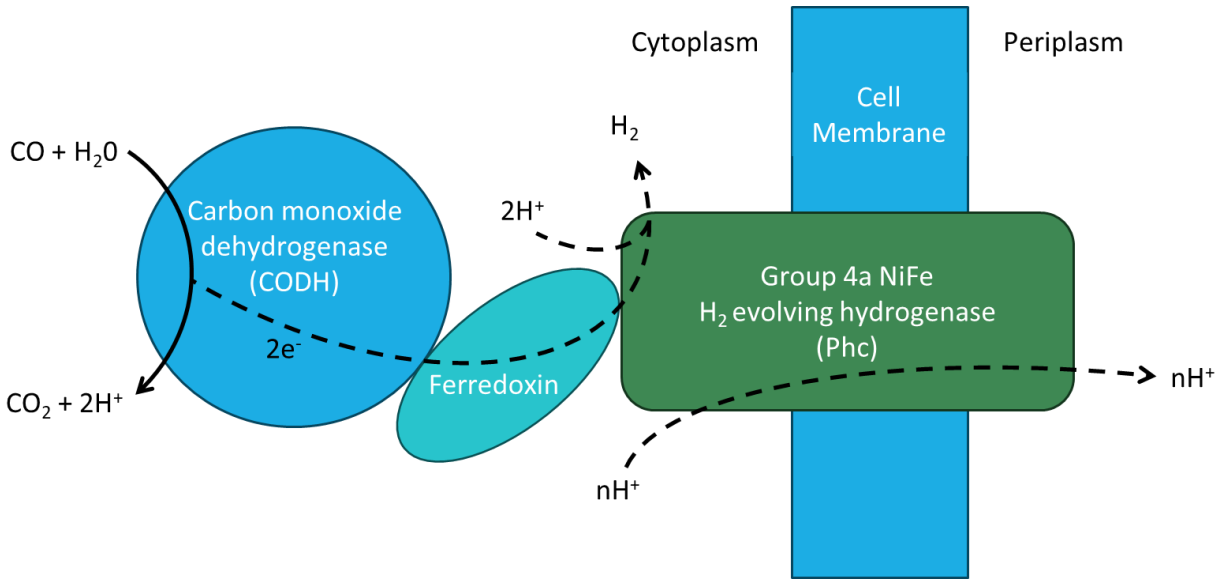


Figure 1.2: Schematic diagram of the membrane bound WGS system present in *P. thermoglucosidasius* DSM 6285. Adapted from Alfano and Cavazza, (2018).

In a following study by (Mohr *et al.*, 2019), several additional *P. thermoglucosidasius* strains were shown to conduct the hydrogenogenic reaction with varying H_2 gas yields and production rates. The study illustrated that strain *P. thermoglucosidasius* DSM 6285 had superior hydrogenogenic capacity compared to other strains, having demonstrated a shorter lag phase between aerobic inoculation and anaerobic H_2 gas production, as well as elevated H_2 gas yields. In contrast, strain *P. thermoglucosidasius* DSM 21625 did not exhibit H_2 gas evolution, despite harbouring the implicated CODH/[NiFe]-hydrogenase loci. Extensive differences between the proteomes of the strains analysed, alongside various polymorphisms and deletions in the CODH/[NiFe]-hydrogenase loci, were proposed to account for this discrepancy in hydrogenogenic activity (Mohr *et al.*, 2018a). An additional factor which may have contributed to the variable hydrogen

yields in these strains may be the influence of the two [NiFe]-uptake hydrogenases encoded on the genomes of *P. thermoglucosidasius*. Two loci, *phaABCDEFGHIJK* and *phbABCDEFGHIJKLMNOP*, encode putative group 1d and group 2a [NiFe]-uptake hydrogenases, Pha and Phb respectively (Mohr *et al.*, 2018a). While the function of these hydrogenases in *P. thermoglucosidasius* have not previously been characterised, other members of group 1d and 2a [NiFe]-uptake hydrogenases have been implicated in H_2 consumption (Vignais and Billoud, 2007; Greening *et al.*, 2016). Consequently, it is possible that these hydrogenases may negatively influence terminal H_2 yield in *P. thermoglucosidasius* WGS reaction fermentations.

Subsequent optimisation of fermentation parameters for *P. thermoglucosidasius* DSM 6285 was conducted by the same group, yielding further enhancements to H_2 production rates (Mohr *et al.*, 2019). Further investigation of the dynamics of the *P. thermoglucosidasius* DSM 6285 WGS reaction was conducted through a time-course transcriptome analysis of the WGS reaction fermentation profile, which highlighted the stark changes in cell metabolism over the course of the fermentation (Aliyu *et al.*, 2020). This study further noted the potential regulation of the CODH and [NiFe]-hydrogenase by O_2 in conjunction with the reversal of metabolic deterioration linked with transition from aerobic respiration to anaerobic growth and sporulation upon commencement of the WGS reaction. This suggested that the transition to CO metabolism further supports the maintenance of growth and proliferative state (Aliyu *et al.*, 2020). Subsequently, elucidation of the influence of CO on the physiology of *P. thermoglucosidasius* was undertaken through analysis of several key metabolic end-products produced under cultivation with varying CO comprising atmospheres (Aliyu *et al.*, 2021). This highlighted the flux of key metabolic pathways in response to CO, in particular, the focused production of lactate and acetate to balance reducing equivalents under a CO-compromised respiratory pathway. The study also further reinforced the notion that WGS reaction dependant metabolism derived energy through CO oxidation, facilitating anaerobic growth (Aliyu *et al.*, 2021).

1.5 Conclusions

H_2 gas is a versatile and increasingly important commercial and industrial reagent as well as an increasingly prominent alternative energy carrier. The global shift towards utilisation of renewable energy sources, such as H_2 gas, has highlighted the demand for environmentally friendly H_2 gas production. Presently, H_2 gas production is conducted almost exclusively through non-renewable processes that gravitate around fossil fuel reformation. However, numerous renewable H_2 gas production processes have been developed to accommodate the transition of H_2 gas production towards more sustainable means. One promising avenue for renewable H_2 gas production is the utilisation of the biologically mediated water-gas shift reaction conducted by various bacterial taxa, such as *Parageobacillus thermoglucosidasius*. These bacteria can utilise CO rich gases to produce H_2 gas, making them a potential candidate for sustainable hydrogen production. Ongoing research is focused on fully characterising and optimising this process to enhance H_2 gas yields and efficiency.

CHAPTER TWO: EXPLORING THE INFLUENCES OF THE TWO *P. THERMOGLUCOSIDASIUS* GROUP 1D AND GROUP 2A NIFE-UPTAKE HYDROGENASES ON H_2 GAS PRODUCTION THROUGH KNOCKOUT MUTAGENESIS

Abstract

Parageobacillus thermoglucosidasius produces H_2 gas through the biological water-gas shift reaction. This reaction is mediated by a complex comprised of a carbon monoxide dehydrogenase and H_2-evolving [NiFe]-hydrogenase, universally encoded by a single operon on the genome of the species. In addition, two operons identified to encode a predicted group 1d and group 2a [NiFe]-uptake hydrogenase, Pha and Phb respectively, are universal to the genomes of all taxa within the species. To date, the function or influence of these hydrogenases on *P. thermoglucosidasius* H_2 metabolism remain uncharacterised. Here, genetic inactivation of genes coding for catalytic subunits (*phaB* and *phbE*, respectively) of the group 1d (Pha) and group 2a (Phb) [NiFe]-uptake hydrogenases was undertaken using two distinct approaches. Neither approach was capable of generating stable mutants for either target gene. Notably, following introduction of either mutant-gene-bearing transformation vectors into *P. thermoglucosidasius*, substantial reduction in host growth was observed. Reductions in growth did not appear to be influenced by either the transformation vectors used, nor the fitness costs associated with the resistance genes they carried. Wild type-like growth of transformants only resumed coincidental with failure to detect the presence of the introduced mutant genes, suggesting that the function of these genes may influence cell viability. Further investigation of the essential roles of these hydrogenases using alternative genetic engineering techniques may assist in resolving their functional relevance.

2.1 Background

Hydrogen (H_2) has a highly negative redox potential (-0.42 V), high diffusion coefficient (4×10^{-9} $m^2.s^{-1}$) and reactive promiscuity, readily serving as either an oxidising or reducing agent (Schwartz *et al.*, 2013). As such, it serves as cornerstone to the metabolism of many prokaryotes, as a readily accessible energy source. Furthermore, it serves as a reductant or reductant sink to facilitate various mechanisms for energy conservation, carbon assimilation, mixotrophic growth and cell survival and maintenance strategies during substrate limitation (Berney and Cook, 2010; Benoit *et al.*, 2020; Greening and Grinter, 2022). Consequently, H_2 metabolism supports several metabolic pathways that facilitate growth in highly adverse or dynamic conditions (Schwartz *et al.*, 2013; Berney *et al.*, 2014). One such pathway, the biological water-gas shift (WGS) reaction, couples CO-oxidation with H_2 reduction to supplement energy generation (Robb and Techtmann, 2018). A broad range of carboxydotrophic prokaryotes are capable of conducting the WGS reaction, including mesophilic, thermophilic and hyperthermophilic bacteria and hyperthermophilic archaea (Fukuyama *et al.*, 2020). These constitute predominantly anaerobic taxa, although a number of strains of the facultative anaerobic, heterotrophic species *P. thermoglucosidasius* have been demonstrated to conduct the hydrogenogenic WGS reaction (Mohr *et al.*, 2018a; Mohr *et al.*, 2019; Aliyu *et al.*, 2021).

This activity has been ascribed to a 29 kb region comprising three distinct operons and fifteen genes. The first operon (*cooCSF*) encodes a carbon monoxide dehydrogenase (CODH) responsible for oxidation of CO to CO_2 (Mohr *et al.*, 2018a; Adachi *et al.*, 2020). The second (*phcA-J*) operon encodes proteins that produce a mature group 4a H_2-evolving [NiFe]-hydrogenase, with small (PhcE) and large (PhcG) catalytic subunits, responsible for H_2 evolution using electrons derived from CO oxidation by the associated CODH (Mohr *et al.*, 2018a; Adachi *et al.*, 2020). The third operon (*phcKL*), encodes HypA- and HypB-like proteins, predicted to participate in nickel incorporation into the enzyme (Chan *et al.*, 2012; Mohr *et al.*, 2018a). In addition to the CODH-group 4a H_2-evolving [NiFe]-hydrogenase locus, two loci, (*phaA-K* and *phbA-P*) encoding putative group 1d and

group 2a [NiFe]-uptake hydrogenases, Pha and Phb respectively, have been identified on the genomes of *P. thermoglucosidasius* (Mohr *et al.*, 2018a). Unidirectional, O_2-tolerant membrane-bound group 1d [NiFe]-uptake hydrogenases are associated with a range of H_2 oxidising aerobic, facultatively anaerobic and anaerobic taxa, including *Wautersia eutropha*, *Escherichia coli* and *Wolinella succinogenes* (Vignais and Billoud, 2007; Greening *et al.*, 2016). H_2 consumed by these hydrogenases is typically understood to be coupled with aerobic respiration or anaerobic respiratory reduction of various organic and inorganic oxidants (Brugna-Guiral *et al.*, 2003; Fritsch *et al.*, 2011; Volbeda *et al.*, 2012). Group 2a [NiFe]-uptake hydrogenases have likewise been implicated in aerobic respiration and recycling H_2 produced through H_2-producing nitrogenase activity or fermentative pathways in aerobic soil and Cyanobacteria (Greening *et al.*, 2016). In these instances, electrons derived from H_2 oxidation are typically believed to be targeted towards the eventual reduction of O_2 (Houchins and Burris, 1981; Greening *et al.*, 2014a; Koch *et al.*, 2014). Presently, the roles of the predicted group 1d (Pha) and group 2a (Phb) hydrogenases in *P. thermoglucosidasius* are not understood. However, it is plausible that these uptake hydrogenases could couple uptake of H_2 produced via the H_2-evolving hydrogenase/carbon monoxide dehydrogenase complex to other metabolic processes, potentially reducing H_2 yield. Here we attempted to determine the potential roles of the Pha and Phb hydrogenases in *P. thermoglucosidasius* through knock-out mutagenesis.

2.2 Methods and materials

Two separate genetic engineering approaches were attempted to inactivate two target genes in *P. thermoglucosidasius*, namely the *phaB* (1,722 bp) and *phbE* (1,635 bp) genes, encoding the large catalytic subunits of the group 1d [NiFe]-uptake hydrogenase or group 2a [NiFe]-uptake hydrogenases, respectively. The first protocol was modelled on the knock-out mutagenesis approach developed by Cripps *et al.* (2009), while the second was adapted from the protocol employed by Bacon *et al.* (2017).

2.2.1 Genetic engineering protocol adapted from Cripps *et al*. (2009)

This protocol is based on the generation of markerless knock-out mutants by introducing a plasmid construct encoding a truncated presumed non-functional mutant of a target gene, which replaces the native gene (Figure 2.1). Cells which successfully take up the plasmid construct exhibit kanamycin resistance, conferred by the plasmid, and are selected for through kanamycin selective culturing (Cripps *et al.*, 2009). The cells are then cultured under kanamycin selective conditions at a temperature (>65˚C) preventing plasmid replication, maintaining dependence of cell proliferation on retention of the kanamycin resistance gene harboured by the plasmid (Liao *et al.*, 1986). This facilitates selection of cells which have undergone a primary integration event through double crossover homologous recombination, integrating the plasmid at the site of the target gene and maintaining kanamycin resistance (Figure 2.2; Cripps *et al.*, 2009). The resultant primary integrants are then cultured under non-selective conditions to facilitate a second recombination event (Figure 2.2), wherein the mutant gene is maintained in lieu of the wild type gene and all elements of the integrated plasmid are removed, resulting in markerless inactivation of the target gene (Cripps *et al.*, 2009).

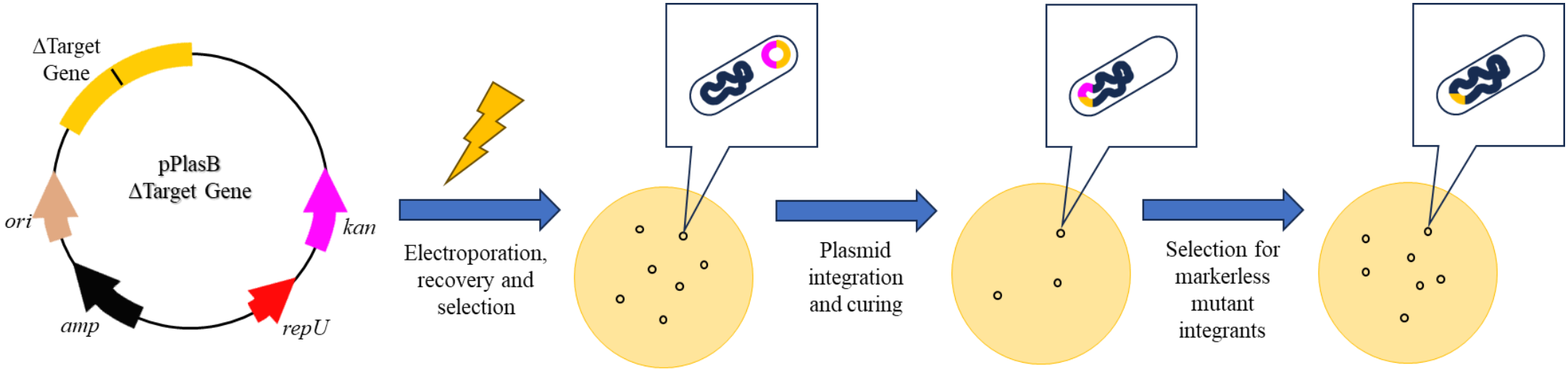


Figure 2.1: Schematic diagram of knockout mutagenesis key steps. (*left to right*) A plasmid construct, pPlasB harbouring a thermosensitive origin of replication, kanamycin resistance marker and a presumed non-functional deleterious target gene sequence homologous to that of the native target gene is electroporated into the host strain. Following plasmid uptake, the plasmid is integrated into the host chromosome through double crossover homologous recombination. These primary integrants are selected for by temperature curing of the plasmid with continued kanamycin selection. A second recombination event is facilitated and the portion of cells containing the markerless deletion of the target gene are identified through replica plating and PCR screening. Genomic DNA (gDNA) is indicated in black, the mutant gene in yellow and remaining plasmid DNA in pink. Adapted from Cripps *et al.* (2009).

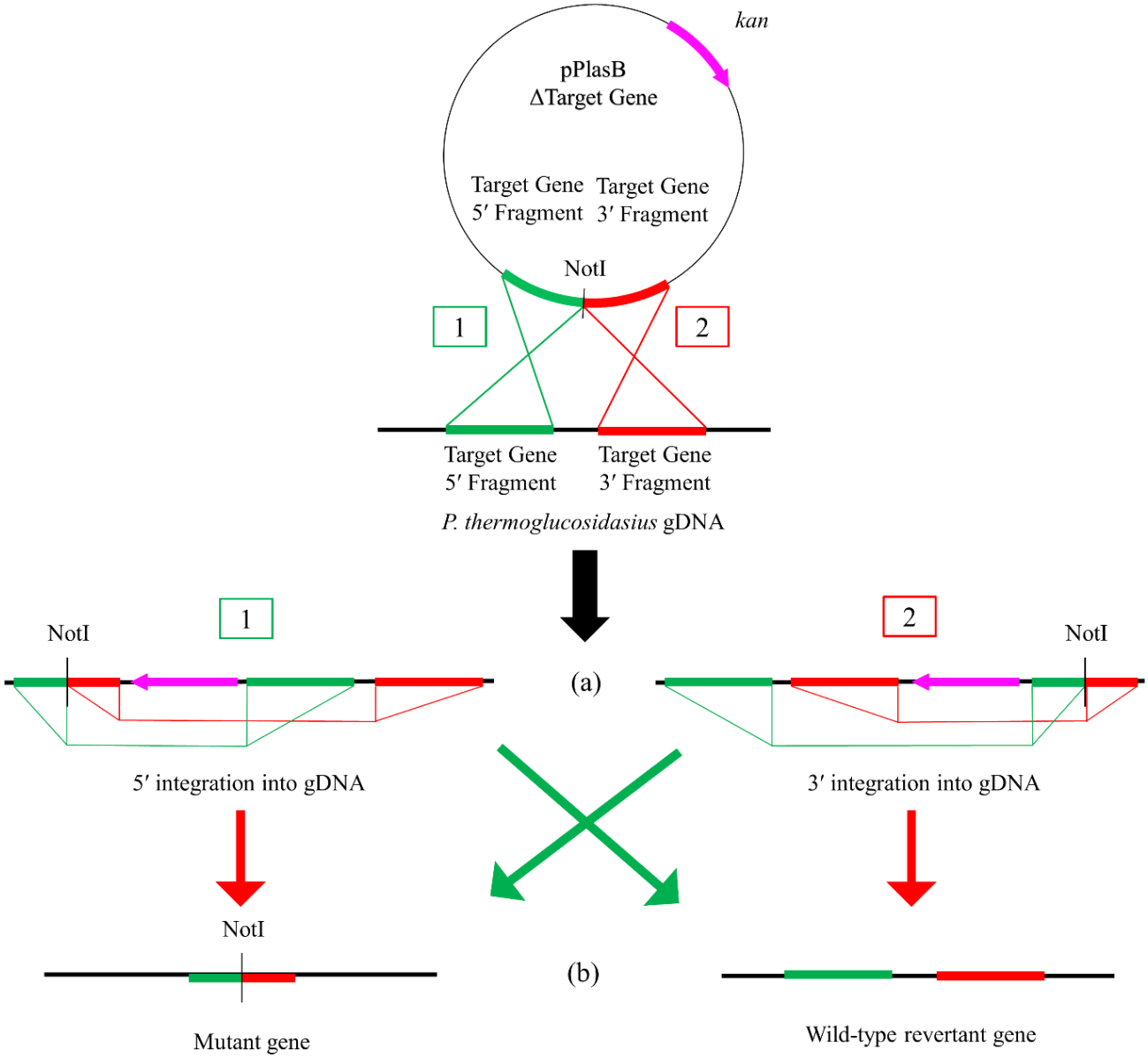


Figure 2.2: Schematic diagram of primary (a) and secondary (b) plasmid integrants following successive double crossover homologous recombination events. Green and red lines indicate homologous double strand recombination events between target and mutant gene 5′ and 3′ fragments respectively. Pathway 1 (left) follows an initial recombination event between the homologous target gene 5′ fragments. Pathway 2 (right) follows an initial recombination event between the target gene 3′ fragments. Adapted from Huang and Wilks, (2017).

2.2.1.1 Media, strains and plasmids

All strains and plasmids used are listed in Table 2.1. *Escherichia coli* DH5α was donated by Dr. Angela Botes at the School of Molecular & Cell Biology, University of the Witwatersrand.

E. coli strains were routinely cultured using Luria–Bertani (LB) medium supplemented (Appendix A) with antibiotics as required on solid medium, incubated at 37˚C for 24 hours, or in liquid LB broth and incubated at 37˚C for 16 hours and shaking at 200 rpm. LB media was supplemented with either 100 µg/mL ampicillin, 50 µg/mL kanamycin or both antibiotics as required. *Bacillus subtilis* DSM 4514, harbouring plasmid pUB110, was routinely cultured using LB medium supplemented with 50 µg/mL kanamycin on solid media and incubated at 25˚C for 24 hours or in liquid media at 25˚C for 16 hours at 200 rpm. *P. thermoglucosidasius* was routinely cultured using modified Luria–Bertani (mLB) medium (Appendix A) and incubated on solid medium at 56˚C for 24 hours or in liquid medium at 56˚C for 16 hours at 120 rpm. While undergoing genetic engineering, *P. thermoglucosidasius* strains were cultured using TGP or 2TY media (Appendix A; Cripps *et al.*, 2009) supplemented with 12.5 µg/mL kanamycin as required. Glycerol stocks (50% glycerol: 50% liquid culture) were used for long term storage of all bacterial strains in a microcentrifuge tube and stored at -80 ºC.

Table 2.1: Strains and plasmids used

Strain	Relevant Characteristics	Source
P. thermoglucosidasius DSM 2542^{T}	Wild type isolate	DSMZ, Braunschweig, Germany
B. subtilis DSM 4514 pUB110	pUB110 plasmid source	DSMZ, Braunschweig, Germany
E. coli BL23 DE3 pUC19	pUC19 plasmid source	DSMZ, Braunschweig, Germany

Strain	**Relevant Characteristics**	**Source**
E. coli DH5α	*fhuA2Δ(argF-lacZ) U169 phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	NEB, Ipswich, MA USA
E. coli DH5α pPlasA*ΔphaB*	pPlasA*ΔphaB* containing *E. coli* DH5α	This study
E. coli DH5α pPlasA*ΔphbE*	pPlasA*ΔphbE* containing *E. coli* DH5α	This study
E. coli DH5α pPlasB*ΔphaB*	pPlasB*ΔphaB* containing *E. coli* DH5α	This study
E. coli DH5α pPlasB*ΔphbE*	pPlasB*ΔphbE* containing *E. coli* DH5α	This study
P. thermoglucosidasius DSM 2542^{T} pPlasB*ΔphaB*	pPlasB*ΔphaB* containing *P. thermoglucosidasius* DSM 2542^{T}	This study
P. thermoglucosidasius DSM 2542^{T} pPlasB*ΔphbE*	pPlasB*ΔphbE* containing *P. thermoglucosidasius* DSM 2542^{T}	This study
pPlasB*ΔphbE*	*ori, amp, ΔphaB*	This study
Plasmid	**Relevant Characteristics**	**Reference**
pUB110	*repU, kan*	Mckenzie *et al.*, (1986)
pUC19	*ori, amp*	Yanisch-Perron *et al.*, (1985)
pUB190	*repU, kan, ori, amp*	Cripps *et al.*, (2009)
pPlasA*ΔphaB*	*ori, amp, ΔphaB*	This study
pPlasA*ΔphbE*	*ori, amp, ΔphbE*	This study
pPlasB*ΔphaB*	*repU, kan, ori, amp, ΔphaB*	This study
pPlasB*ΔphbE*	*ori, amp, ΔphaB*	This study

2.2.1.2 Preparation of electrocompetent *E. coli* DH5α

Electrocompetent *E. coli* DH5α cell stocks were prepared by inoculating a single colony in 50 mL LB broth, followed by incubation for 16 hours at 37˚C and 200 rpm. An aliquot of 5 mL of this culture was used to inoculate 500 mL of LB broth, which was incubated for 6-8 hours at 37˚C and 300 rpm to an absorbance (OD_{600}) of 0.5-0.7. The culture was then chilled on ice for 20 minutes prior to pelleting in pre-chilled 50 mL centrifuge tubes through centrifugation using a Beckman Coulter Allegra X-12R (Beckman Coulter, Brea, USA) at 3000 x g for 15 minutes at 4˚C. Cell pellets were washed four times with 50 mL, 25 mL, 10 mL and 10 mL of ice cold 10% (v/v) glycerol consecutively. Subsequently the pellet was resuspended in 1 mL ice-cold 10% (v/v) glycerol, aliquoted in 100 µL aliquots and stored at -80˚C.

2.2.1.3 Preparation of electrocompetent *P. thermoglucosidasius*

Electrocompetent *P. thermoglucosidasius* stocks were prepared by inoculating 50 mL TGP broth with a single colony prior to incubation at 52˚C and 200 rpm to an OD_{600} of 1.8. Cultures were then chilled on ice for 10 minutes and pelleted through centrifugation at 3,000 x g for 15 minutes at 4˚C in pre-chilled 50 mL centrifuge tubes. Cell pellets were washed four times with 50 mL, 25 mL, 10 mL and 10 mL ice cold electroporation buffer (0.5 M mannitol, 0.5 M sorbitol, 10% (v/v) glycerol). Following the final wash, cells were resuspended in 1 mL ice cold electroporation buffer before preparation of 100 µL aliquots which were stored at -80˚C.

2.2.1.4 Generation of pUB190 transformation vector

A universal shuttle vector, pUB190 was generated from the pUC19 and pUB110 plasmids (Figure 2.3). pUC19 is a general *E. coli* cloning vector, with an origin of replication (*ori*) capable of replication at 37˚C and confers host ampicillin resistance through the encoded *amp* gene at 37˚C (Yanisch-Perron *et al.*, 1985). pUB110 is a Gram-positive cloning vector, originally isolated from *Staphylococcus aureus* and commonly used as a cloning vector in *B. subtilis* (Shivakumar and Dubnau, 1978). This plasmid possesses an origin of replication capable of replicating at both mesophilic (37˚C) and thermophilic temperatures (<60˚C; Shivakumar and Dubnau, 1978; McKenzie *et al.*, 1986; Cripps *et al.*, 2009), and

confers kanamycin and phleomycin resistance through the presence of *kan* and *fhl* genes (Matsumura *et al.*, 1984). The pUB190 construct confers both ampicillin and kanamycin resistance at 37˚C and kanamycin resistance above 37˚C. Limited replicative ability is observed for the plasmid above 60˚C, allowing it to serve as a temperature sensitive "pseudo-suicide" vector and served as the backbone for the *P. thermoglucosidasius* transformation vectors (Shivakumar and Dubnau, 1978; Cripps *et al.*, 2009; Sheng *et al.*, 2017).

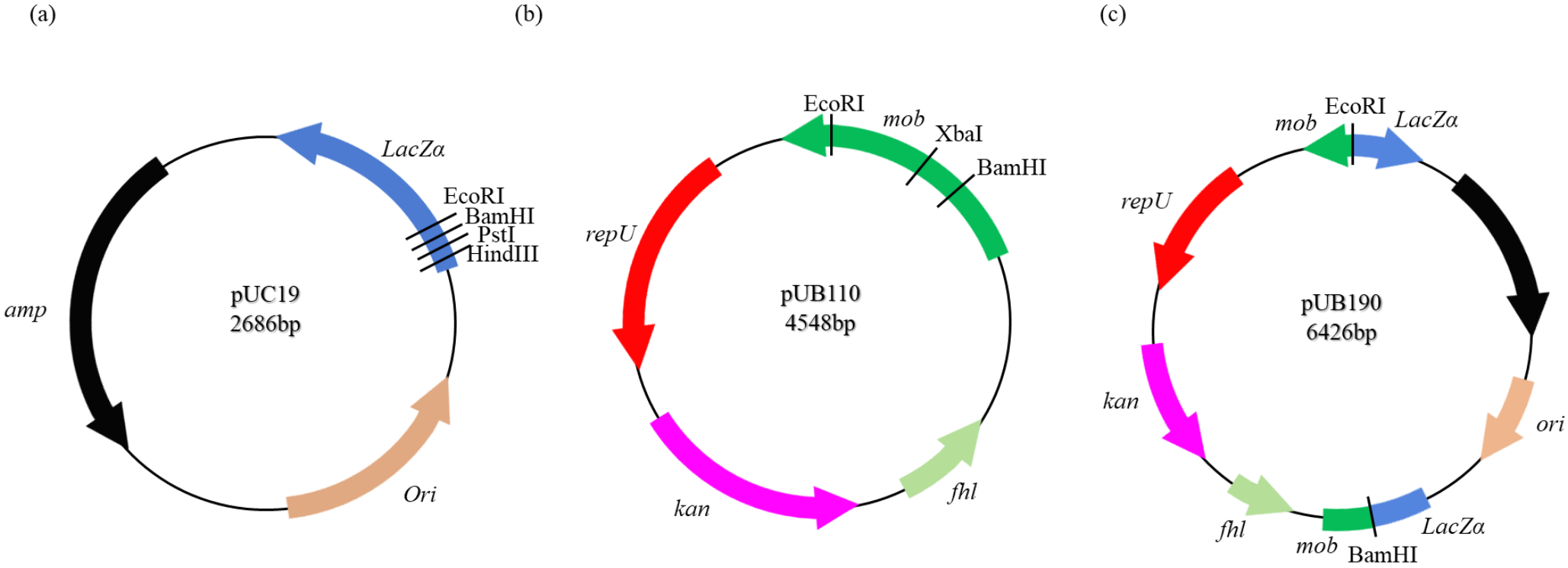


Figure 2.3: Schematic diagrams of pUC19 (a) and pUB110 (b) plasmids and shuttle vector pUB190 (c).

Plasmid extraction

pUB110 was extracted from *B. subtilis* DSM 4514 using a ZymoPURE Plasmid Miniprep kit (Zymo Research, Irvine, CA, USA), according to the Gram-positive bacteria protocol. Following extraction, plasmid concentration and extract purity were evaluated using a Nanodrop^{TM} 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, USA). DNA samples which exhibited A260/280 and A260/230 ratios of 1.8 ± 0.2 and 2.1 ± 0.1 were deemed of sufficient quality (Lucena-Aguilar *et al.*, 2016) for further downstream application and stored at -20˚C. Plasmid extract purity was further confirmed through agarose gel electrophoresis (results not shown), using a 0.8% TAE agarose gel run at 100 V for 1 h. Samples were loaded and stained with Novel Juice in a 1:6 DNA:Novel Juice (Sigma-Aldrich, St. Louis, MO, USA) ratio and visualised using a Gel Doc XR (BioRad, Hercules, CA, USA).

Restriction digestion, purification and ligation

The extracted pUB110 plasmid and pUC19 were digested with EcoRI-HF (NEB, Ipswich, MA, USA) and XbaI (NEB, Ipswich, MA, USA) as per the manufacturer's recommendations (Table 2.2). Digest specificity was evaluated by 0.8% TAE agarose gel electrophoresis, loaded and stained with Novel Juice. Non-specific cleavage of the pUB110 plasmid by XbaI was observed using recommended parameters (20 U XbaI, incubated at 37˚C for 15 minutes; Figure 2.4). Optimization of the digestion was attempted by halving enzyme concentration and incubation time; however, non-specific cleavage was still observed. Consequently, XbaI-HF digestion was substituted with BamHI-HF (NEB, Ipswich, MA, USA). Following EcoRI-HF/BamHI-HF digestion, pUC19 digests were column purified using a ZYMO Research DNA Clean and Concentrator-25 kit (Zymo Research, Irvine, CA, USA) as per the manufacturer's recommendations. pUB110 digests were separated by 0.8% TAE agarose gel electrophoresis run at 100 V for 1 h. Samples were loaded with 1:5 amplicon: 6x Purple loading dye (NEB, Ipswich, MA, USA) ratio. The gel was post-stained with GelRed (Sigma-Aldrich, St. Louis, USA), using 15 µL GelRed in 50 mL distilled water for 30 minutes at 50 rpm. The 3,761 bp fragment containing the *repU*, *kan* and *fhl* genes, was excised and purified using a Zymoclean Gel DNA Recovery Kit as per the manufacturer's recommendations. The DNA

fragment was quantified using a Nanodrop^{TM} 1000 Spectrophotometer. The 3,761 bp pUB110 DNA fragment and the 2,665 bp pUC19 DNA fragment, containing cleaved *lacZα*, *amp* and *ori* genes, were ligated at a 1:3 fmol DNA ends ratio using 30 ng and 64 ng of each plasmid fraction, respectively. Ligation was performed with T4 DNA ligase (ThermoFisher Scientific, Waltham, MA, USA) using the rapid ligation of cohesive ends protocol (Weiss *et al.*, 1968).

Table 2.2: Restriction digestion reaction components

Component	Volume
Single digest reaction	
Restriction enzyme	1 µL
DNA	Volume to 1 µg
10x NEBuffer	5 µL
Nuclease-free water	Volume to 50 µL
Double digest reaction	
Restriction enzyme 1	1 µL
Restriction enzyme 2	1 µL
DNA	Volume to 1 µg
10x NEBuffer	5 µL
Nuclease-free water	Volume to 50 µL

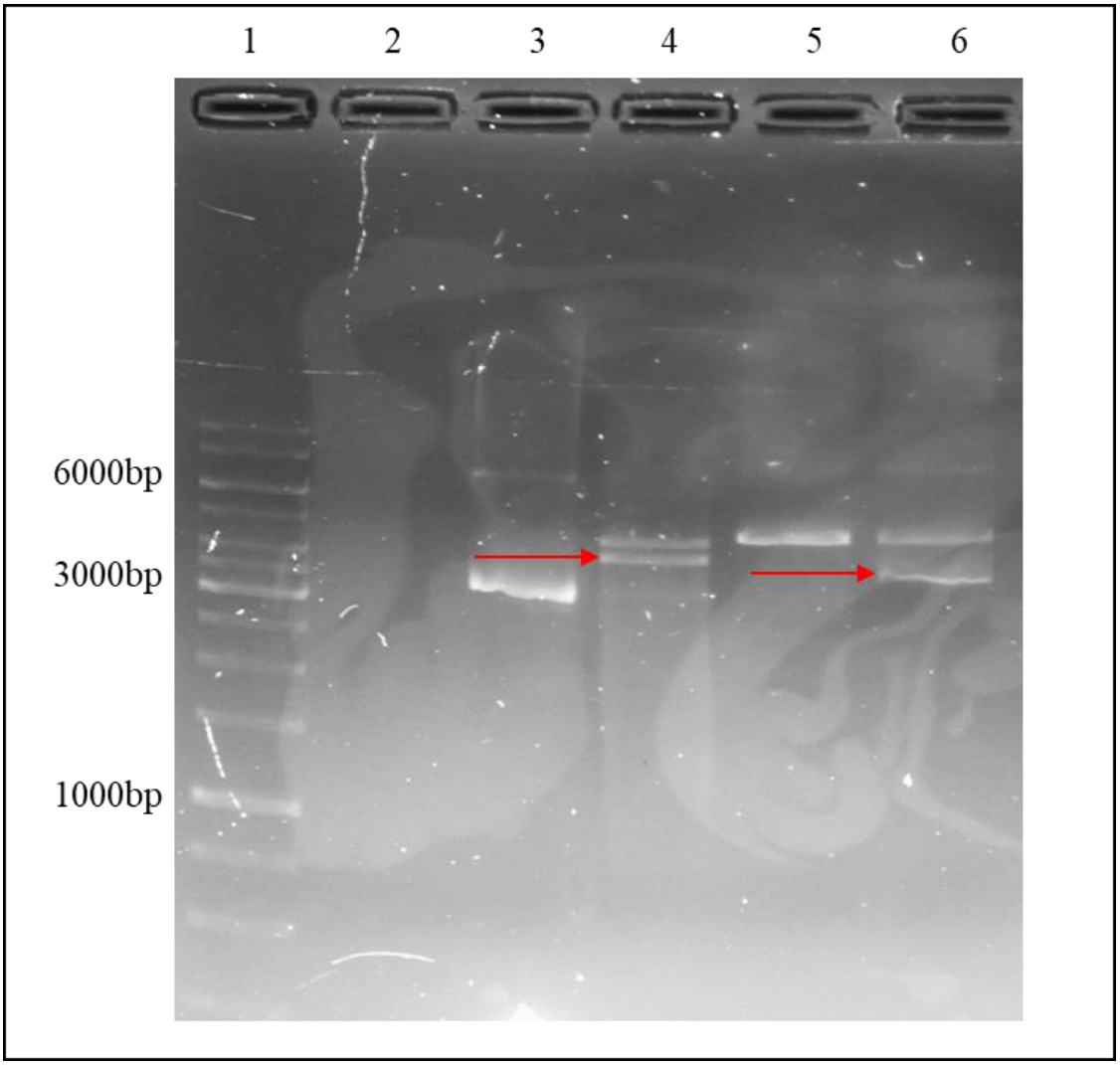


Figure 2.4: 0.8% Agarose gel showing XbaI-HF/EcoRI-HF digestion of pUB110 (1) 1 Kbp DNA ladder (2) blank (3) pUB110 plasmid (4) pUB110 XbaI-HF restriction digest (5) pUB110 EcoRI-HF restriction digest (6) pUB110 XbaI-HF / EcoRI-HF restriction digest. Red arrows indicate non-specific cleavage.

2.2.1.5 Transformation of *E. coli* DH5α

Following ligation, 5 µL of the reaction was mixed with 60 µL of thawed (10 min) electrocompetent *E. coli* DH5α cells and incubated on ice for 1 minute. The suspension was then transferred to a 0.1 cm electroporation cuvette (BioRad, Hercules, CA, USA) and electroporated at 1.8 kV using a MicroPulser Electroporator (BioRad, Hercules, CA, USA). pUC19 plasmid was used as a positive control (~1 ng in 5 µl nuclease free water), to confirm that the prepared

electrocompetent *E. coli* DH5α cells were suitable for electroporation. Cells were recovered in 1 mL *E. coli* SOC medium (Appendix A; Sharma and Schimke, 1996) for 1 hour at 37˚C and shaking at 200 rpm in a thin-walled test tube. A phosphate buffered saline (Appendix A) dilution series (undiluted to 1 x 10^{-7}) of the recovered cells were then plated on LB agar supplemented with 100 µg/mL ampicillin (Sigma-Aldrich, St. Louis, USA), 50 µg/mL kanamycin (Glentham Life Sciences) and 200 µg/mL X-Gal (Invitrogen). The latter was added to facilitate blue/white screening (Padmanabhan *et al.*, 2011). The reaction mix was incubated at 37˚C for 24 hours. Successful transformants were identified through the development of white colonies, indicative of retention of the correctly assembled plasmid with disrupted LacZα gene (Padmanabhan *et al.*, 2011). Transformation efficiency was calculated using the following equation: $\frac{number\ of\ white\ colonies\ on\ plate}{\mu g\ DNA\ \ transformed \times dilution}$. Successful transformants were sub-cultured in 50 mL LB broth supplemented with 100 µg/mL ampicillin and 5 µg/mL kanamycin and incubated for 16 hours at 37˚C and 120 rpm prior to plasmid extraction using a ZymoPURE Plasmid Miniprep kit (Zymo Research, Irvine, CA, USA).

2.2.1.6 pPlasA primary transformation vector construction

Non-functional mutant gene copies of *phaB* and *phbE* were constructed through PCR amplification of a 5′ and 3′ fragment of each target gene which when ligated omitted central regions of 50 and 47 bp, respectively of each target gene (Figure 2.5). The central deletion regions were substituted with a rare 8 bp NotI restriction site sequence, introduced by the primers used. The use of only a relatively minor deletion of the target genes was used to increase the likelihood for subsequent recombination attempts between the native and mutant gene constructs. While the wild type *phaB* and *phbE* genes are 1,722 bp and 1,635 bp in size, respectively, successful mutants should present genes 1,680 bp and 1,596 bp in size, respectively. The mutant gene copies were then integrated into a pUC19 backbone to produce the primary transformation vectors pPlasAΔ*phaB* and pPlasAΔ*phbE* (Figure 2.6).

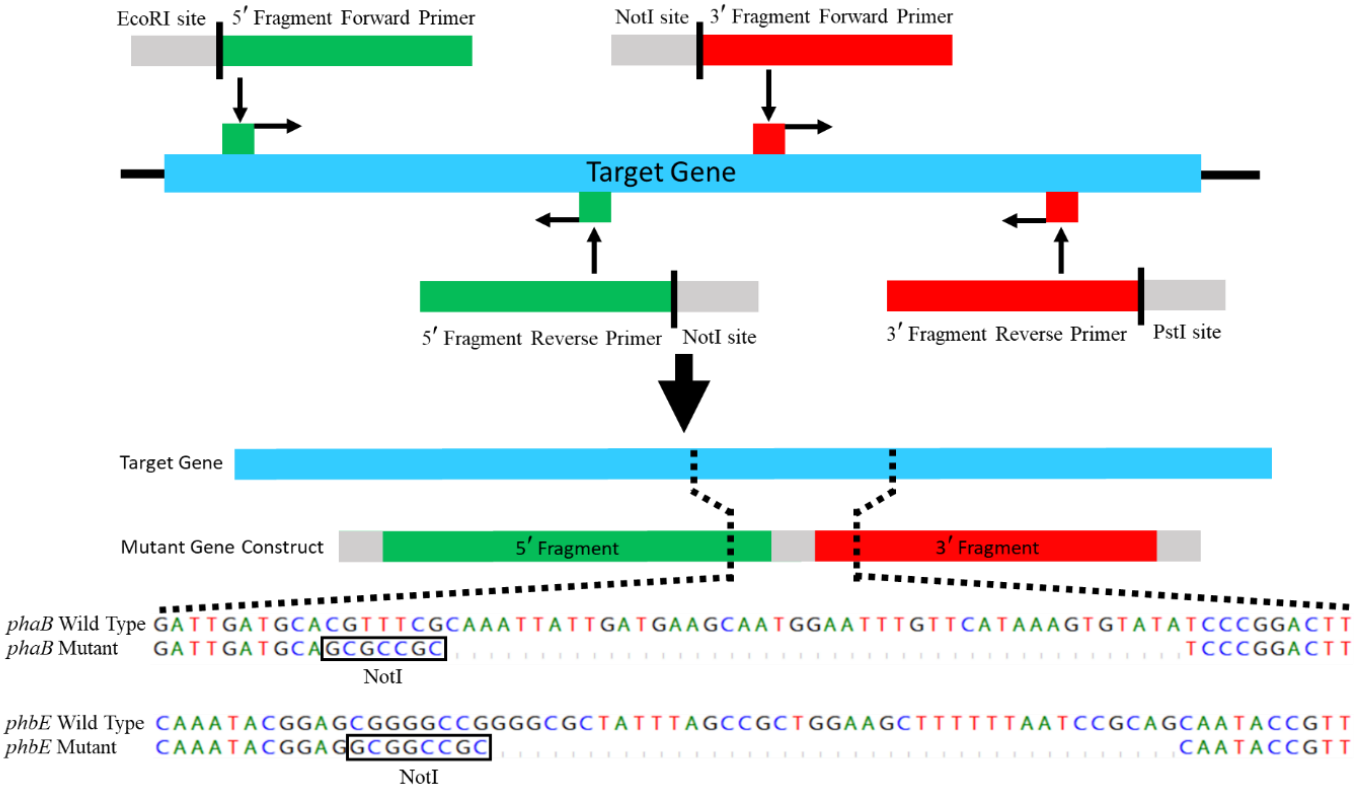


Figure 2.5: Schematic diagram showing construction of mutant target gene constructs.

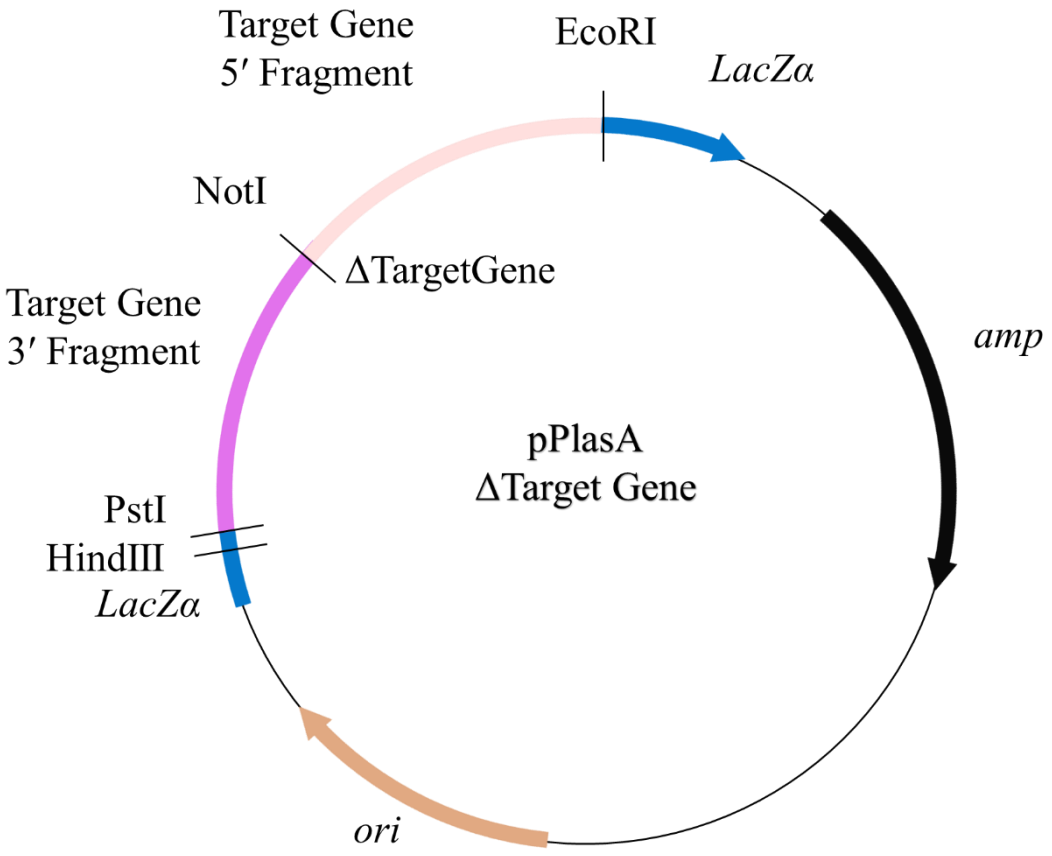


Figure 2.6: Schematic diagram showing pPlasA construct framework.

Genomic DNA extraction

Genomic DNA was extracted from *P. thermoglucosidasius* cultures grown in mLB for 16 h at 56˚C at 120 rpm, using the Monarch® Genomic DNA Purification Kit (NEB, Ipswich, MA, USA), per the manufacturer's protocol for gDNA purification from Gram-positive bacteria and archaea. Purified DNA was subsequently quantified using a Nanodrop™ 1000 Spectrophotometer.

PCR amplification and purification

All primers were manually designed and specificity assessed through local BlastN analysis against *P. thermoglucosidasius* and *E. coli* host genomes and plasmids using Bioedit v. 7.7.1 (Hall, 1999), and are listed in Table 2.3. Two fragments (a 5′ and 3′ fragment) of each target gene were amplified (Figure 2.5; Figure 2.7). Primers *phab*87F-1 and *phab*725R (T_m=67˚C) and *phab*775F and *phab*1541R (T_m =68˚C) were used to incorporate 5′-EcoRI/3′-NotI and 5′-NotI/3′-PstI restriction sites into the 5′ (638 bp) and 3′ (766 bp) fragment amplification products of the *phaB* gene. Primers *phbE*327F-1 and *phbE*781R (T_m =66˚C) and *phbE*828F and *phbE*1454R (T_m =66˚C) were used to amplify the 5′ (454 bp) and 3′ (626 bp) fragments of the *phbE* gene. PCR amplifications of target gene fragments were conducted using Q5® High-Fidelity DNA Polymerase (NEB, Ipswich, MA, USA) as per the manufacturer's recommendations (Table 2.4; Table 2.5). PCR reactions for each target amplicon were column purified using a ZYMO Research DNA Clean and Concentrator-25 kit as per the manufacturer's recommendations. Purified PCR amplicons were quantified using a Nanodrop™ 1000 Spectrophotometer.

Table 2.3: Primers used

Primer	Sequence 5′ – 3′	Target
*phaB*87F-1	AAGATTGAATTCCAAGATGCATTCAGCTC TGGTAC Underlined bases indicate EcoRI-HF restriction sequence	87 bp 3′ of *phaB* start
*phaB*725R	AATATAGCGGCCGCTGCATCAATCGTTCG GCATG Underlined bases indicate NotI-HF restriction sequence	725 bp 3′ of *phaB* start
*phaB*775F	ATTATTGCGGCCGCTCCCGGACTTGCTTG CTATC Underlined bases indicate NotI-HF restriction sequence	775 bp 3′ of *phaB* start
*phaB*1541R-1	ATATTACTGCAGGGTGAAGCGTTCCACGT TG Underlined bases indicate PstI-HF restriction sequence	1541 bp 3′ of *phaB* start
*phbE*327F-1	ATAGTAGAATTCACGATTCAAAGCATTCC TCGTTG Underlined bases indicate EcoRI-HF restriction sequence	327 bp 3′ of *phbE* start
*phbE*781R	TAATATGCGGCCGCCTCCGTATTTGTCCA ATCCGATTTC Underlined bases indicate NotI-HF restriction sequence	781 bp 3′ of *phbE* start
*phbE*828F	AATACTGCGGCCGCCAATACCGTTATCCG ACGATTGAAG Underlined bases indicate NotI-HF restriction sequence	828 bp 3′ of *phbE* start

Primer	Sequence 5′ – 3′	Target
*phbE*1454R-1	ATGATACTGCAGTGACGATCTCTTGGCCCAATG Underlined bases indicate PstI-HF restriction sequence	1454 bp 3′ of *phbE* start
pUC19 *LacZα*5R	CATGATTACGCCAAGCTTGCATG	2 bp 5′ of pPlasA/pPlasB mutant gene insertion site
*phaB*87F-2	GCGAATTGTTGTTGATCCAGTACC	87 bp 3′ of *phaB* start
*phbE*327F-2	AAGCATGGGTCAAACGAGTG	327 bp 3′ of *phbE* start

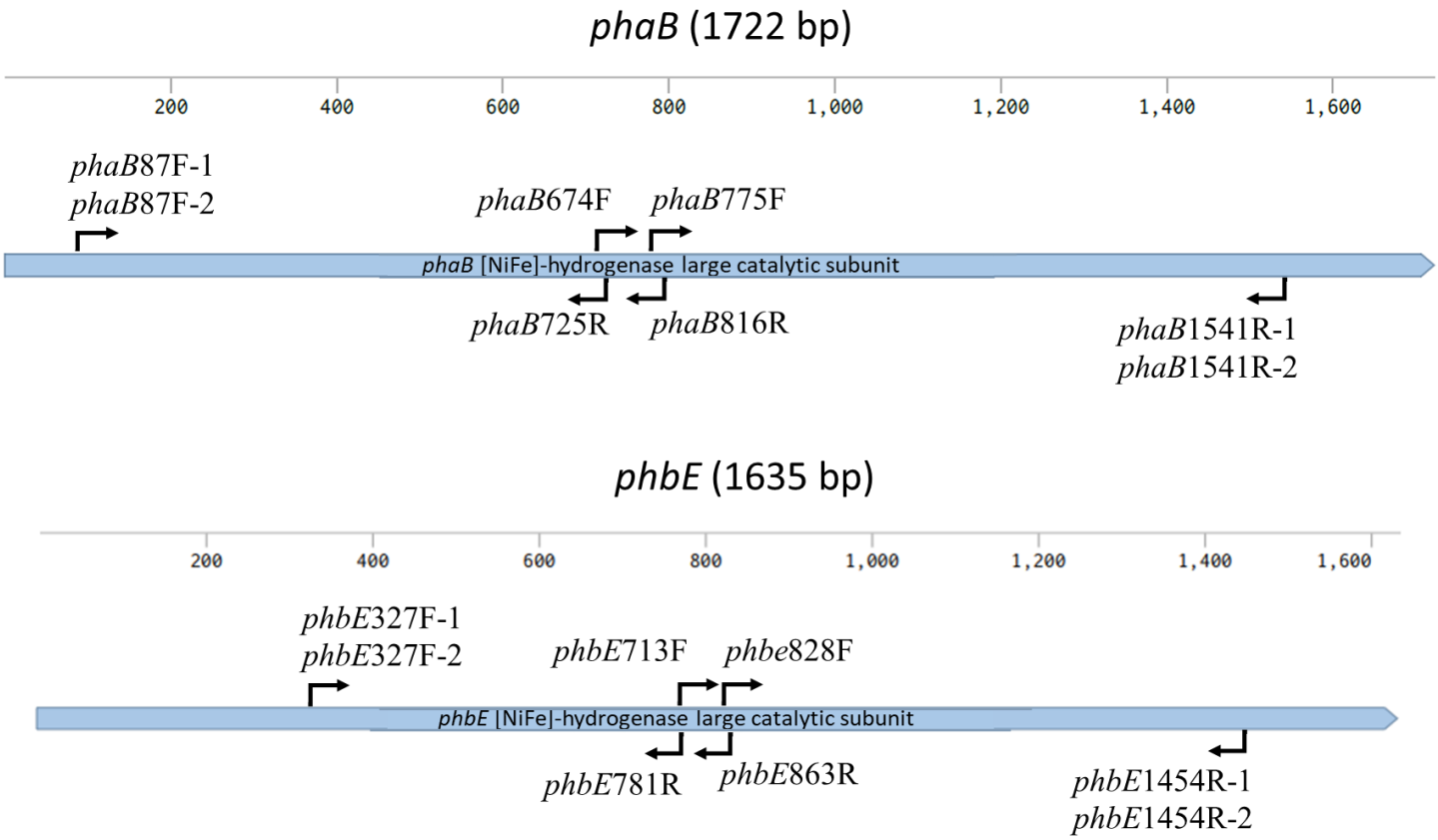


Figure 2.7: Schematic diagram showing *P. thermoglucosidasius* primer binding sites used to generate and validate *phaB* and *phbE* mutant gene fragments. Arrows indicate direction of primer extension.

Table 2.4: Q5® High-Fidelity DNA polymerase reaction components

Component	Volume (µL)	Final Concentration
5x Q5 Reaction buffer	10	1x
10 mM dNTPs	1	200 µM
10 µM Forward primer	1	0.5 µM
10 µM Reverse primer	1	0.5 µM
Template DNA	Volume to 200 ng	4 ng/µL
Q5® High-Fidelity DNA polymerase	0.5	0.02 U/µL
Nuclease-free water	Volume to 50 µL	1x

Table 2.5: Q5® High-Fidelity DNA polymerase thermal cycling conditions used for PCR reactions

Step	Temperature (˚C)	Time (s)	Cycles
Initial Denaturation	98	30	1
Denaturation	98	10	35
Annealing	Variable = (T_m)	10	
Extension	72	20 s/kb	
Final extension	72	120	1

Restriction digestion, purification and electroporation

The generated *phaB* and *phbE* amplicons were digested with EcoRI-HF/NotI-HF as specified by the manufacturer's recommendations (Table 2.2). Digests were column purified using a ZYMO Research DNA Clean and Concentrator-25 kit (Zymo Research, Irvine, CA, USA) and the DNA quantified using a Nanodrop™ 1000 Spectrophotometer. *phaB* and *phbE* 3′ fragments were similarly digested with NotI-HF/PstI-HF, as per the manufacturer's recommendations. Purified DNA fragments were subsequently quantified using a Nanodrop™ 1000 Spectrophotometer. Similarly, pUC19 was digested with EcoRI-HF/PstI-HF, as specified by the manufacturer's recommendations, column purified and the digest DNA quantified. The respective purified 5′ and 3′ fragments of each target gene and pUC19 fragments were then ligated in three-point ligation reactions using T4 DNA Ligase, using the rapid ligation of cohesive ends protocol at a 1:1:1 ratio, using 30 ng of each digest. This yielded the mutant gene-containing plasmids pPlasAΔ*phaB* and pPlasAΔ*phbE*. A volume of 5 µL of each ligation reaction was used to transform electrocompetent *E. coli* DH5α cells as described previously (Section 2.2.1.5).

Validation of pPlasA constructs

Correct construct orientation and mutant gene sequences were assessed by PCR amplification of mutant genes from plasmid extracts. To confirm correct directional insertion of the *phaB* and *phbE* mutant gene constructs into the pUC19 backbone, the primer pUC19*LacZα*5R was used in conjunction with primers *phaB*1541R-2 and *phbE*1454R-2 (T_m = 68˚C), respectively. pUC19*LacZα*5R targeted a plasmid sequence 2 bp upstream of intended mutant gene insertion site, located near the PstI restriction site in the *LacZα* gene of the pUC19 backbone (Figure 2.8). PCR amplification of the 1,442 and 1,118 bp amplicons targeted by primers pUC19*LacZα*5R and *phaB*1541R-2, and pUC19*LacZα*5R and *phbE*1454R-2 respectively, was conducted using Q5® High-Fidelity DNA Polymerase as per the manufacturer's recommendations as conducted previously (Section 2.2.1.6). Amplicons were then separated through agarose gel electrophoresis, using a 0.8% TAE agarose gel run at 100 V for 1 h. Samples were loaded and stained with Novel Juice in a 1:6 DNA:Novel Juice ratio and visualised using a Gel Doc XR. Samples were excised from the gel and purified using a Zymoclean Gel DNA Recovery Kit as per the manufacturer's recommendations. Purified amplicons were then sequenced through Sanger sequencing (ABI 3500XL Genetic Analyzer, POP7™, BrilliantDye™ Terminator v3.1) at Inqaba Biotech (South Africa). Base correction of the produced chromatograms was conducted using FinchTV v1.4 and the curated nucleotide sequences aligned against the respective wild type target gene nucleotide sequences through clustalW alignment using Bioedit v. 7.7.1 (Thompson *et al.*, 1994; Hall, 1999).

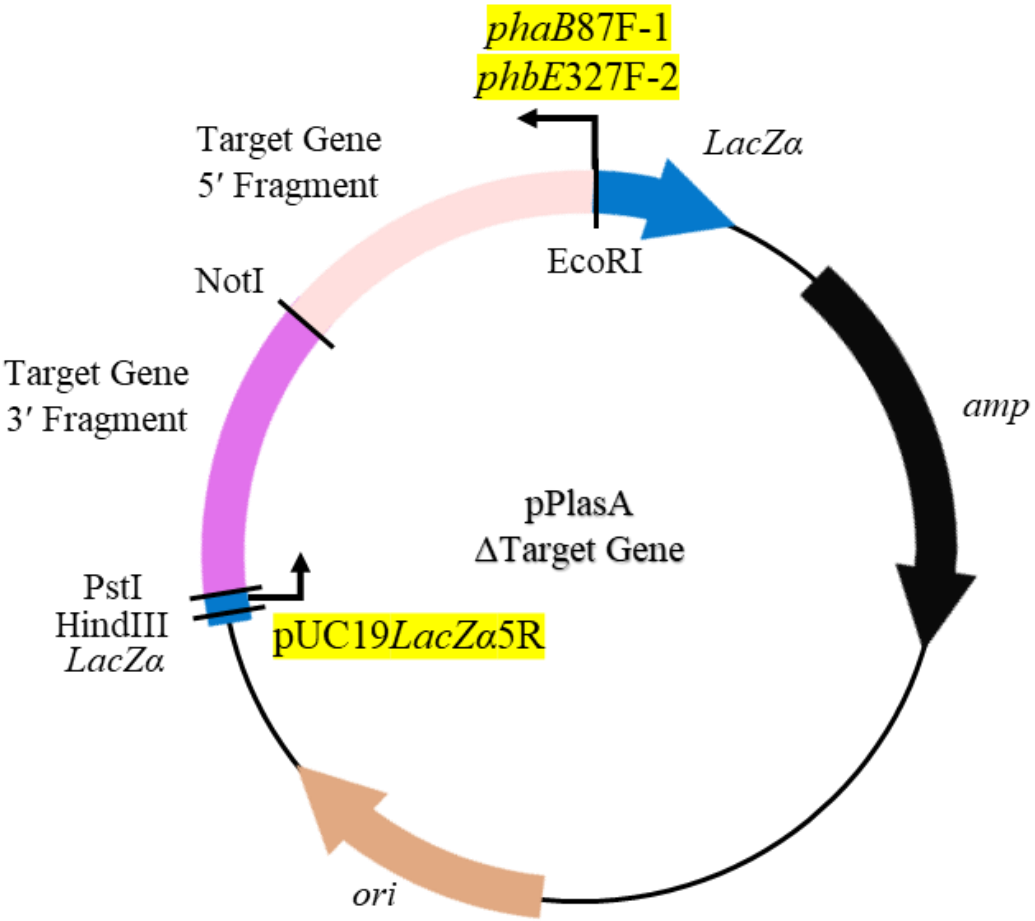


Figure 2.8: Schematic diagram showing primer binding sites targeted for verification of pPlasA constructs. Primers are highlighted in yellow. Primer binding sites and primer amplification direction are indicated by the black arrows.

2.2.1.7 pPlasB *P. thermoglucosidasius* transformation vector construction

Plasmid extraction, restriction digestion, purification, ligation and electroporation into *E. coli*

pPlasB plasmids comprised the constructed mutant target genes derived from the corresponding pPlasA plasmids, inserted into the pUB190 shuttle vector backbone. The pPlasA plasmids generated in section 2.2.1.6 were extracted from their respective *E. coli* DH5α host as previously described (Section 2.2.1.4). pPlasA plasmids were then digested with EcoRI-HF/HindIII-HF as per the manufacturer's recommendations (Table 2.2). Five digest replicates were conducted and pooled

together prior to being column purified using a ZYMO Research DNA Clean and Concentrator-25 kit. This was done to mitigate low yield during the purification steps. Purified digests were separated through 0.8% TAE agarose gel run at 100 V for 1 h to isolate the 1,429 bp *ΔphaB* and 1,106 bp *ΔphbE* mutant gene containing fragments. Samples were loaded and stained with Novel Juice in a 1:6 DNA:Novel Juice ratio and visualised using a Gel Doc XR. The target mutant gene fragments were then excised and purified using a Zymoclean Gel DNA Recovery Kit as per the manufacturer's recommendations. The DNA fragments were subsequently quantified using a NanodropTM 1000 Spectrophotometer. pUB190 plasmid was digested with BamHI-HF as per the manufacturer's recommendations, column purified, and quantified. The purified 6,426 bp fragment of the purified pUB190 digest was then Klenow treated to produce blunt ends, using DNA Polymerase I (NEB, Ipswich, MA, USA) as per the manufacturer's recommendations. Subsequently, the 5′ ends were dephosphorylated to prevent excessive downstream self-ligation using QuickCIP (NEB, Ipswich, MA, USA) in accordance with the manufacturer's instructions. The dephosphorylated pUB190 fragment was subsequently column purified using a ZYMO Research DNA Clean and Concentrator-25 kit (Zymo Research, Irvine, CA, USA) as per the manufacturer's recommendations. The purified fragment was subsequently quantified using a Nanodrop^{TM} 1000 Spectrophotometer. The blunt-end mutant genes and 6,426bp pUB190-fragments were ligated with 5x T4 DNA ligase (ThermoFisher Scientific, Waltham, MA, USA) as per the manufacturer's recommendations using the rapid ligation of blunt ends protocol in 3:1 ratios. A volume of 5 µL of each ligation reaction was used to transform electrocompetent *E. coli* DH5α cells as previously described (Section 2.2.1.5), to facilitate storage and replication of the plasmid constructs. This plasmid set was capable of replication in both *E. coli* and *P. thermoglucosidasius* and conferred both ampicillin and kanamycin resistance at 37˚C and kanamycin resistance above 37˚C. Successful transformants were cultured in 50 mL LB broth supplemented with 100 µL ampicillin and 50 µL kanamycin, for 16 hours at 37˚C and 120 rpm prior to plasmid extraction and quantification.

Validation of pPlasB constructs

Due to the blunt-end ligation protocol used, pPlasB constructs could exhibit two predominant construct orientations (Figure 2.9). Orientations differed such that in the first orientation (Figure 2.9a), the mutant gene insert was flanked by the shuttle vector *ori* origin of replication (upstream of the target gene 3′ fragment) and *mob* (downstream of the target gene 5′ fragment). The second orientation (Figure 2.9b) was structured such that the mutant gene insert was flanked by the shuttle vector *ori* region (upstream of the target gene 5′ fragment) and *mob* (downstream of the target gene 3′ fragment). *E. coli* pPlasB transformants were screened for construct orientation. pPlasB plasmids were extracted from isolated *E. coli* transformants as described in section 2.2.1.4, and digested with PstI-HF (NEB, Ipswich, MA, USA) as per the manufacturer's recommendations (Table 2.2). Digest banding patterns were subsequently observed through 0.8% TAE agarose gel electrophoresis run at 100 V for 1 h. Samples were loaded with 1:5 6x Purple loading dye and post-stained with GelRed, using 15 µL GelRed in 50 mL distilled water for 30 minutes at 50 rpm. In order to maintain consistency, constructs with the second construct orientation (Figure 2.9b) were used in subsequent transformation attempts.

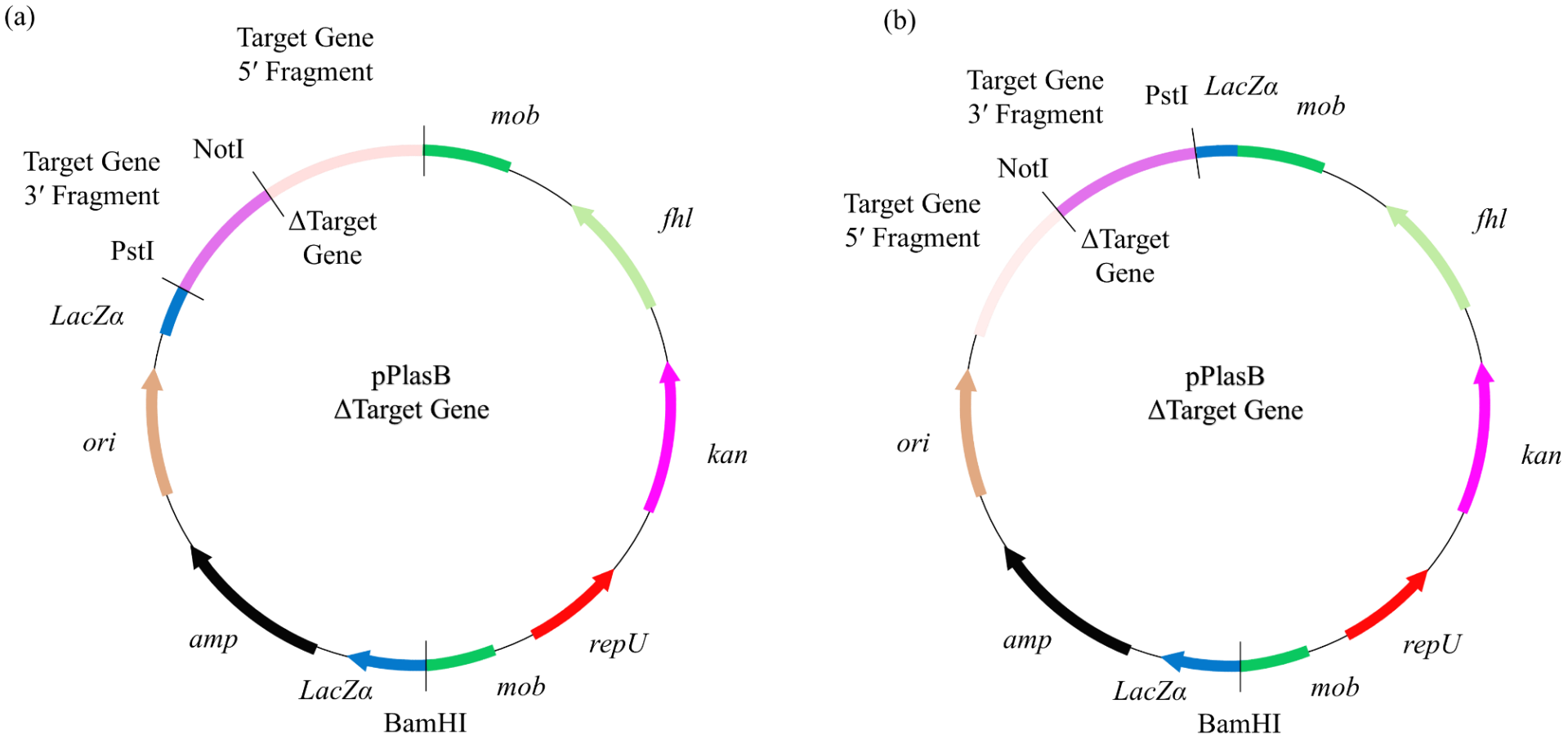


Figure 2.9: Schematic diagram of pPlasB orientation 1 (a) and orientation 2 (b) architectures.

2.2.1.8 Electroporation of *P. thermoglucosidasius* with pPlasBΔ*phaB* and pPlasBΔ*phbE*

Transformation of *P. thermoglucosidasius* DSM 6285 with pPlasBΔ*phaB* and pPlasBΔ*phbE* was attempted as per the protocol employed by Cripps *et al.* (2009). Frozen electrocompetent *P. thermoglucosidasius* DSM 6285 stocks were thawed for 10 minutes on ice prior to use. Once thawed, 60 µL of electrocompetent cells were mixed with 200-500 ng of pPlasB transformation vectors, transferred to a 0.1 cm electroporation cuvette and electroporated at 2.5 kV using a MicroPulser Electroporator. Cells were recovered in 1 mL TGP medium for 1-4 hours at 52˚C and 200 rpm in a thin-walled test tube. Cells were then plated in 100 µL aliquots on TGP agar supplemented with 12.5 µg/mL kanamycin and incubated at 52˚C for 24 hours. No successful transformation attempts were achieved following this protocol and numerous mitigations were evaluated. Varied combinations of plasmid volume (100, 200, 500, 1,000 ng), electroporation voltage (1.8, 2.5 kV), recovery period (1, 2, 3, 4 h), recovery medium (TGP, 2TY, mLB), recovery agitation (120, 160, 200 rpm) outgrowth period (24, 48, 72 h) and recovery/outgrowth temperature (52, 56, 60˚C) were used. As no outgrowth was observed following optimization, transformation using the type strain of *P. thermoglucosidasius*, namely strain DSM 2542^T, was undertaken. This strain has previously been successfully transformed through the electroporation protocol applied here (Cripps *et al.*, 2009). Although superior hydrogen production has been observed for *P. thermoglucosidasius* DSM 6285, the DSM 2542^T strain has also exhibited H_2 production activity attributed to the WGS reaction (Mohr *et al.*, 2018a; Mohr *et al.*, 2018b). The *phaB* and *phbE* genes of *P. thermoglucosidasius* DSM 2542^T, share 99.71% and 99.76% identity, respectively, with those of *P. thermoglucosidasius* DSM 6285. This amounted to five (*phaB*) and four (*phbE*) single-nucleotide polymorphisms between the respective genes, none of which overlapped with any of the utilized primer binding regions. Transformation of *P. thermoglucosidasius* DSM 2542^T was conducted as per the protocol employed by Cripps *et al.* (2009). However, where Cripps *et al.* (2009) recovered transformed colonies after 24 h, no colony development was observed for either transformed plasmid in the suggested time-frame. As such, outgrowth was extended to between 48 and 72 h.

2.2.1.9 Generation of *P. thermoglucosidasius* primary integrants

Genomic integration and plasmid curing of pPlasB

Following outgrowth of transformants *P. thermoglucosidasius* DSM 2542^T, colonies were assumed to be transformants harbouring either pPlasB*ΔphaB* or pPlasBΔ*phbE* plasmids. To facilitate a double crossover homologous recombination event between homologous regions of the native target gene and the plasmid-borne mutant gene (Figure 2.2), transformants were cultured in 50 ml TGP broth supplemented with 12.5 µg/ml kanamycin at 52°C for 16 h. Following incubation, 100 µL aliquots of a PBS dilution series ranging from undiluted to 1 x 10^{-7} were plated in triplicate on 2TY agar supplemented with 12.5 µg/ml kanamycin and cultured at 68°C for up to 72 h. Cultivation above >65˚C restricts plasmid replication facilitating plasmid curing (Liao *et al.*, 1986). Use of kanamycin selection at this step necessitates retention of the *kan* resistance gene encoded by the plasmid, selecting for recombinants cured of free plasmids, but possessing genome integrated plasmids, incorporated at the site of the target genes through double crossover homologous recombination (Cripps *et al.*, 2009). No growth was observed from the dilution series, for either pPlasBΔ*phaB* or pPlasBΔ*phbE* transformants, necessitating plating of greater cell concentrations. Aliquots of the same liquid culture, cultivated at 68˚C for 16 h (0.1, 0.2, 0.5, 1, 2, 3 and 5 mL), were pelleted by centrifugation at 3000 x g for 5 minutes, resuspended in 100 µL of PBS and plated in triplicate on 2TY agar supplemented with 12.5 µg/ml kanamycin and cultured at 68°C for 48 h. Colonies that developed following 48 h incubation were presumed primary integrants which had undergone plasmid integration and had been cured of autonomously replicating plasmids. Given that colonies which developed following 48 h incubation at 68˚C were minute, the wild type strain was cultured similarly, sparing use of kanamycin, to confirm if poor growth was the result of the elevated temperatures applied or plasmid transformation.

Screening for primary integrants

DNA was extracted from scrapings of the presumptive primary integrants by resuspending in 20 µL of 2 mM NaOH and heating to 100°C for 8 minutes. The reaction was then vortexed and centrifuged at 16,000 x g for 30 seconds. Supernatants were then transferred to clean 0.2 ml microcentrifuge tubes and used to conduct PCR amplifications. PCR amplification of a region spanning the targeted deletion regions of each target gene was conducted to assess the presence of wild type and mutant gene sequences. Primers *phaB*674F and *phaB*816R (T_m = 48˚C) targeting the *phaB* gene region and primers *phbE*713F and *phbE*863R (T_m = 47˚C) targeting the *phbE* gene region (Table 2.6; Figure 2.7). PCR amplification was conducted using OneTaq DNA polymerase (Table 2.7; Table 2.8; NEB, Ipswich, MA USA). Target gene PCR products were separated through 3% TAE agarose gel electrophoresis (100 V for 1 h 20 min, loaded and stained with Novel Juice). Wild type and mutant *phaB* and *phbE* amplicons were distinguished based on length, with mutant amplicons being 58 and 39 bp shorter than those of the wildtypes, respectively. Colonies that showed successful amplification of both the wild type and mutant gene were presumed to have undergone the primary recombination event and cured of the plasmid.

Table 2.6: Primers used in the validation and screening of mutant gene constructs

Primer	Sequence 5′ – 3′	Target
*phaB*674F	GCCGATTGATATTAACAGCG	674 bp 3′ of *phaB* start
*phaB*816R	CCAATCTTTGTAGTAAGAAGCG	816 bp 3′ of *phaB* start
*phbE*713F	GCACCGCAATTCTGATCTTG	713 bp 3′ of *phbE* start
*phbE*863R	GCGTTGCGTCCTTCAATC	863 bp 3′ of *phbE* start
pUC19*LacZα*76F	TGGCCGCGTTTTACAACGTC	76 bp 3′ of pUC19 *LacZα* start

Primer	Sequence 5′ – 3′	Target
pUC19*amp*401R	TTGGCCGCAGTGTTATCACTC	401 bp 3′ of pUC19 *amp* start
pUB110*kan*81F	GGGGATGATGTTAAGGCTATTGG	81 bp 3′ of pUB110 *kan* start
*phaB*87F-2	CAAGATGCATTCAGCTCTGGTAC	87 bp 3′ of *phaB* start
*phbE*327F-2	TCCCGGACTTGCTTGCTATC	327 bp 3′ of *phbE* start
*phaB*1451R-2	GGTGAAGCGTTCCACGTTG	1451 bp 3′ of *phaB* start
*phbE*1454R-2	TGACGATCTCTTGGCCCAATG	1454 bp 3′ of *phbE* start

Table 2.7: OneTaq DNA polymerase reaction components

Component	Volume (µL)	Final Concentration
5x OneTaq Standard Reaction Buffer	10	1x
10 mM dNTPs	1	200 µM
10 µM Forward primer	1	0.2 µM
10 µM Reverse primer	1	0.2 µM
Template DNA	Volume to 200 ng	4 ng/µL
OneTaq DNA Polymerase	0.25	1.25 U/µL
Nuclease-free water	Volume to 50 µL	1x

Table 2.8: OneTaq PCR amplification thermal cycling conditions

Step	Temperature (˚C)	Time (s)	Cycles
Initial Denaturation	94	30	1
Denaturation	94	15	35
Annealing	Variable = (T_a)	15	
Extension	72	60 s/kb	
Final extension	72	300	1

2.2.1.10 Generation of *P. thermoglucosidasius* secondary integrants

Three isolated colonies for each target gene identified as primary integrants were serially subcultured five times in 10 mL of TGP without kanamycin in 50 ml Erlenmeyer flasks. Incubation was performed at 60°C for 8 h and shaking at 120 rpm, using 100 µL inoculum. This was done to facilitate a secondary crossover event generating stable secondary integrants that possessed only the mutant target gene, were kanamycin sensitive and harboured no remnants of the previously integrated plasmid (Figure 2.1; Figure 2.2). Following the fifth subculture, a dilution series ranging from undiluted to 1 x 10^{-7} aliquots were generated and plated in triplicate on TGP agar. For each of the three cultures per target gene, 500 individual colonies were replica plated on TGP agar with 12.5 µg/mL kanamycin to identify kanamycin sensitive colonies. Kanamycin sensitive colonies were screened for the presence of the mutant genes. DNA was extracted from portions of the developed colonies as previously described (Section 2.2.1.9) and target genes amplified by PCR amplification. PCR amplification was performed using OneTaq DNA polymerase (Table 2.7; Table 2.8), primers *phaB*674F and *phaB*816R (T_m = 48˚C) for the *phaB* gene target and primers *phbE*713F and *phbE*863R (T_m = 47˚C) for the *phbE* gene target (Table 2.6). PCR products were separated though 3% agarose gel electrophoresis (loaded and stained with Novel Juice) as previously described (Section 2.2.1.9). Failing detection of any secondary integrant mutants through this, following replica plating, a subset of 100 kanamycin resistant colonies for each target gene transformation attempt were similarly screened. To determine

if elements of the pPlasB plasmids were present primers pUC19*LacZα*76F and pUC19*amp*401R (T_m = 54˚C), targeting a 994 bp region of the incorporated pUC19 plasmid (Table 2.6; Figure 2.10). To establish if the *kan* resistance gene had incorporated near the target gene, the primer pUB110*kan*81F was used in combination with primers *phaB*87F-2 (T_m = 52˚C), *phbE*327F-2 (T_m = 51˚C), *phaB*1451R-2 (T_m = 52˚C) and *phbE*1454R-2 (T_m = 52˚C; Table 2.6; Figure 2.10). pUB110*kan*81F targeted the 5′ region of the potentially inserted pUB110 *kan* gene that was harboured on pPlasB transformation vectors. Potential pairing of this primer with the latter primers, which considered both potential mutant gene orientations in pPlasB constructs (Figure 2.9) would theoretically produce ~3,100 bp amplicons if the *kan* gene had integrated into the host chromosome as intended.

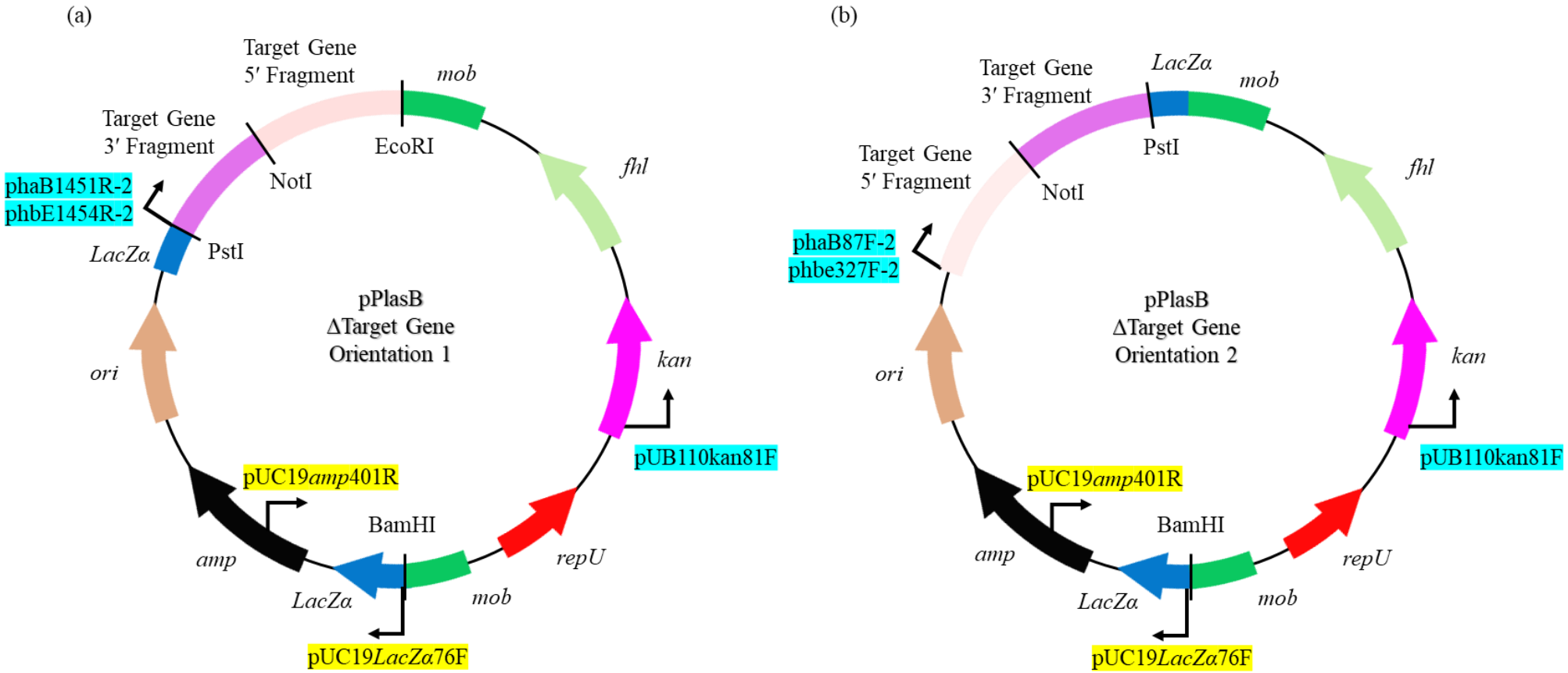


Figure 2.10: Schematic diagram showing primer pairs and binding sites used to detect presence of pPlasB retention (yellow) and potential insertion of the *kan* gene (light blue) into the *P. thermoglucosidasius* chromosome for pPlasB construct orientation 1 (a) and 2 (b).

2.2.2 Genetic engineering protocol adapted from Bacon *et al.* (2017)

Following failure of the first protocol to generate stable knockout mutants for either target gene, a second knockout mutagenesis protocol was attempted. The second protocol relied on a transformation principle loosely adapted from the approach employed by Bacon *et al.*, 2017. Unlike the first protocol, which utilised elements of pUC19 and pUB110 plasmids to construct the transformation vector backbone, this approach made use of elements of the pG2K plasmid (Figure 2.11a; Reeve *et al.*, 2016; Bacon *et al.*, 2017). Transformation constructs (Figure 2.11b and c) were created with modified architectures to eliminate the need for plasmid curing and facilitate alternative integration patterns via homologous recombination (Figure 2.12). Unlike the Cripps *et al.* (2009) protocol, which replaced the wild type target gene with a deleterious copy, this protocol relied on the removal of the targeted sequences in their entirety (Bacon *et al.*, 2017). Like the approach applied by Cripps *et al.* (2009; Section 2.2.1), integration of the plasmid-borne kanamycin resistance gene (*kan*) into the host chromosome is required to permit proliferation under kanamycin selective conditions. Dissimilarly though, generation of primary integrants through plasmid electroporation into host cells, integration of the plasmid into the host chromosome and plasmid curing were attempted as a single step instead of three successive steps (Figure 2.12). The primary integrants generated through this approach are then cultured under non-selective conditions to facilitate a second recombination event (Figure 2.12), wherein the mutant gene is maintained in lieu of the wild type gene and all elements of the integrated plasmid are removed, resulting in markerless inactivation of the target gene sequence (Bacon *et al.*, 2017).

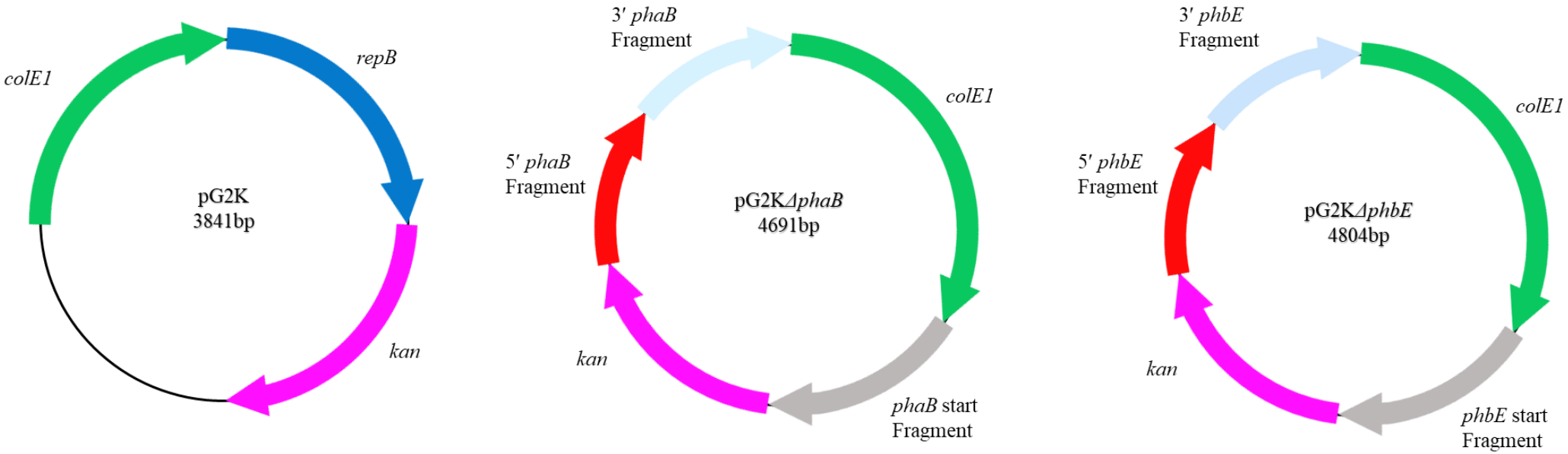


Figure 2.11: Schematic diagram of plasmid pG2K (a) and plasmid constructs pG2KΔ*phaB* (b) and pG2KΔ*phbE* (c). Grey fragments represent sequences homologous to the 5′ start regions of the target genes. Red fragments represent 5′ homology arms, homologous to regions upstream of the target genes. Light blue fragments represent 3′ homology arms, homologous to regions downstream of the target genes. Adapted from Reeve *et al.* 2016.

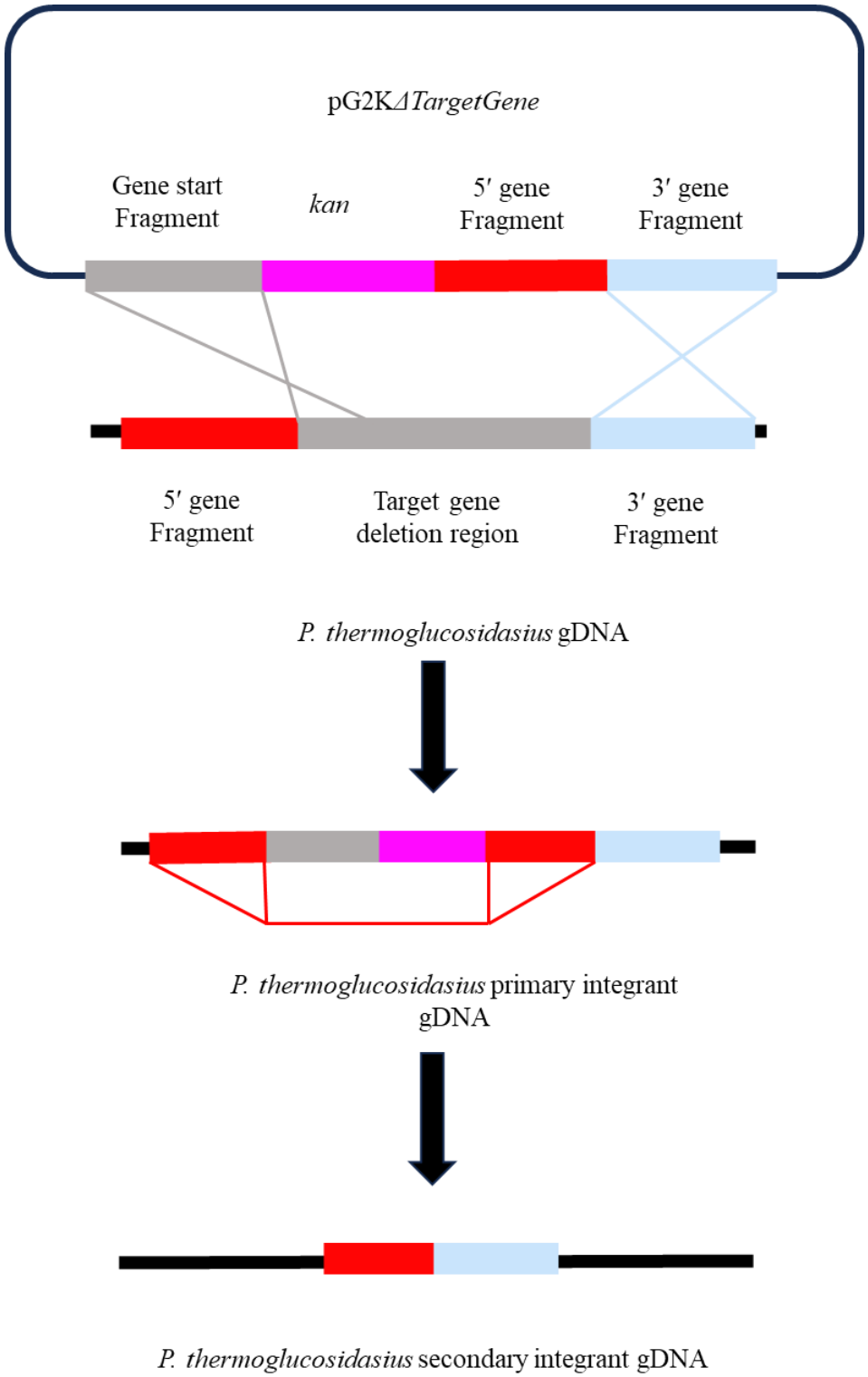


Figure 2.12: Schematic diagram of *P. thermoglucosidasius* primary and secondary plasmid integrants following successive double crossover homologous recombination events. Grey fragments represent homology arms homologous to the 5′ start regions of the target genes. Red fragments represent homology arms homologous to the regions upstream of the target genes. Light blue fragments represent homology arms homologous to the regions downstream of the target genes. Adapted from Bacon *et al.* 2017.

2.2.2.1 Media strains and plasmids

All strains, plasmids and relevant strain phenotypes used are listed in Table 2.9. *E. coli* strains were routinely cultured using Luria–Bertani (LB) medium (Appendix A) on solid medium with 50 µg/mL kanamycin as required. Cultures were incubated at 37˚C for 24 hours, or in liquid LB broth and incubated at 37˚C for 16 hours with shaking at 200 rpm. *P. thermoglucosidasius* was routinely cultured using modified Luria–Bertani (mLB) medium (Appendix A) and incubated at 56˚C on solid media for 24 hours or in liquid media for 16 hours with shaking at 120 rpm. While undergoing genetic engineering, *P. thermoglucosidasius* was cultured using TGP or SPYNG media supplemented (Appendix A) with 12.5 µg/mL kanamycin as required.

Table 2.9: Strains and plasmids used

Strain	**Relevant Characteristics**	**Source**
P. thermoglucosidasius DSM 6285	Wild type isolate	DSMZ, Braunschweig, Germany
E. coli DH5α	*fhuA2Δ(argF-lacZ) U169 phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	NEB, Ipswich, MA, USA
E. coli DH5α pG2K	pG2K containing *E. coli* DH5α	Addgene, Watertown, NY, USA
E. coli 10β	Δ(ara-leu) 7697 araD139 *fhuA* ΔlacX74 galK16 galE15 e14- ϕ80d*lacZ*ΔM15 *recA1 relA1 endA1 nupG rpsL* (StrR) *rph spoT1* Δ(mrr-hsdRMS-mcrBC)	NEB, Ipswich, MA, USA
E. coli 10β pG2K*ΔphaB*	pG2K*ΔphaB* containing *E. coli* 10β	This study
E. coli 10β pG2K*ΔphbE*	pG2K*ΔphbE* containing *E. coli* 10β	This study

Plasmid	Relevant Characteristics	Reference
pG2k	*colE1, kan, repB*	Reeve *et al.*, (2016)
pG2K*ΔphaB*	*colE1, kan, phaB* knockout cassette	This study
pG2K*ΔphbE*	*colE1, kan, phbE* knockout cassette	This study

2.2.2.2 Transformation vector construction using NEBuilder HiFi DNA assembly workflow

Plasmid and gDNA extraction

The *Geobacillus* cloning vector pG2K (Reeve *et al.*, 2016), carrying the mesophilic *colE1* origin of replication and thermostable *kan* resistance gene, was extracted from *E. coli* DH5α harbouring the plasmid, using a Monarch Plasmid Miniprep Kit (NEB, Ipswich, MA, USA) as per the manufacturers' recommendations. Following extraction, plasmid concentration and extract purity were evaluated using a Nanodrop One Spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). Plasmid extract purity was further confirmed through agarose gel electrophoresis, using a 0.8% TAE agarose gel run at 100 V for 1 h. Samples were loaded in a 1:6 ratio with ROTI Load DNA (Carl Roth, Karlsruhe, Germany) and stained with ROTI Gel stain (Carl Roth, Karlsruhe, Germany), using 2.5 µL ROTI Gel Stain per 50 mL agarose gel solution. Gels were visualised using a FAS-Digi Compact Gel Imaging System (Nippon Genetics Europe, Düren, Germany). gDNA was extracted from *P. thermoglucosidasius* using a Zymo Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research, Irvine, CA USA), according to the manufacturer's recommendations. Following extraction, gDNA concentration and extract purity were evaluated using a Nanodrop One Spectrophotometer.

PCR amplification and purification of plasmid construct fragments

Primers (Table 2.10) were designed using the NEBuilder assembly tool v 2.8.1 (www.*nebuilder.neb.com*) and assessed for specificity through local BlastN analysis against the *P. thermoglucosidasius* DSM 6285 genome using Bioedit v. 7.7.1 (Hall, 1999). Primers generated using this tool were designed such that, following PCR

amplification of targeted sequences, the generated amplicons could be used to construct a circularized vector, similar to the principle applied by Zuo and Rabie, (2010). Distinct primer sets were generated for each intended plasmid construct to ensure optimal overhang complementarity and prevent unintended non-specific binding in the targeted sequences. Primers included 3′ sequences complementary to targeted genomic/plasmid target sequences. Primers also included 5′ overhangs designed to complement the overhang region of intended neighbouring primer amplicons. Inclusion of these overhangs on both sense and antisense primers facilitated subsequent circularisation of the assembled fragments.

Transformation vectors were designed to include two pG2K-derived genes and three *P. thermoglucosidasius* derived sequences. The mesophile *colE1* origin of replication from the pG2K plasmid (Reeve *et al.*, 2016) was PCR amplified as a 1,834 bp fragment, using primers pG2K*colE1*F1 and pG2K*colE1*R1 (T_m = 51.2˚C) for the pG2KΔ*phaB* construct and as a 1,831 bp fragment, using primers pG2K*colE1*F2 and pG2K*colE1*R2 (T_m = 51.2˚C) for the pG2KΔ*phbE* construct. This facilitated replication and maintenance of the derived plasmid constructs in host *E. coli* 10β. A thermophilic origin of replication capable of facilitating plasmid replication above 42˚C in *P. thermoglucosidasius* was not included in the plasmid constructs to prevent potential undesirable retention of the plasmids and eliminate the need for plasmid curing. The kanamycin resistance gene (*kan*) was used as a selective marker for the plasmid constructs and was PCR amplified from the pG2K plasmid as a 1,009 bp fragment, using primers pG2K*kan*F1 and pG2K*kan*R1 (T_m = 56.8˚C) for the pG2KΔ*phaB* construct and primers pG2K*kan*F2 and pG2K*kan*R2 (T_m = 50.1˚C) for the pG2KΔ*phbE* construct.

A 673 bp and 675 bp fragment, respectively, of the *phaB* and *phbE* genes targeted for deletion were PCR amplified from *P. thermoglucosidasius* gDNA using primers *phaB*startF and *phaB*startR (T_m = 51.4˚C) for the pG2KΔ*phaB* construct and primers *phbE*startF and *phbE*startR (T_m = 51.4˚C) for the pG2KΔ*phbE* construct. DNA sequences both upstream and downstream of the *P. thermoglucosidasius* genes targeted for deletion were PCR amplified, to serve as homology arms, facilitating primary integration of the plasmid constructs into the host chromosome (Figure 2.11; Figure 2.12). The 5′ homology arms (626 bp and 733 bp, respectively)

of the target genes were PCR amplified from *P. thermoglucosidasius* gDNA using primers *phaB*5′F and *phaB*5′R (T_m = 53.3˚C) for the pG2KΔ*phaB* construct and primers *phbE*5′F and *phbE*5′R (T_m = 49.2˚C) for the pG2KΔ*phbE* construct. The 3′ homology arms (674 bp and 681 bp, respectively) of the target genes were PCR amplified from *P. thermoglucosidasius* gDNA using primers *phaB*3′F and *phaB*3′R (T_m = 52.6˚C) for the pG2KΔ*phaB* construct and primers *phbE*3′F and *phbE*3′R (T_m = 47.6˚C) for the pG2KΔ*phbE* construct. Primers were designed such that when assembled through NEBuilder DNA HiFi assembly, plasmid constructs would assemble as follows: *colE1* - 5′ homology arm – *kan* - target gene start fragment - 3′ homology arm.

All PCR amplifications in this section were conducted using OneTaq Hot Start Quick-Load 2x Master Mix with Standard Buffer (Table 2.7; Table 2.8; NEB, Ipswich, MA USA), as per the manufacturer's recommendations. PCR products were subsequently separated through 0.8% agarose gel electrophoresis, run at 100 V for 1 h. Samples were loaded in a 1:6 ratio with ROTI Load DNA and stained with ROTI Gel stain, using 2.5 µL ROTI Gel Stain per 50 mL agarose gel solution. Gels were visualised using a FAS-Digi Compact Gel Imaging System and target amplicons excised from the gels and purified using a Zymoclean Gel DNA Recovery Kit, as per the manufacturer's recommendations. Following purification, DNA concentration and purity were evaluated using a Nanodrop One Spectrophotometer.

Table 2.10: Primers used

Primer	Sequence 5′ – 3′	Target
pG2K*colE1* F1	TCATTTAGTCGCATTTCCATAGGCTAAG CGTACTAGTCCTAGGC	pG2K ColE 5′ start
pG2K*colE1* R1	ATCAACAACAATTCGCTCTGCCATTTTT AACAGATGGGGTTTTTCTGTG	pG2K ColE 3′ end
pG2K*colE1* F2	CTGACGTCAACGGTTTCCATAGGCTAA GCGTACTAGTCCTAGGCTAG	pG2K ColE1 5′ start
pG2K*colE1* R2	AGAATGTACTTTCTTGAGATCCTTTAAC AGATGGGGTTTTTCTGTGC	pG2K ColE1 3′ end

Primer	**Sequence 5′ – 3′**	**Target**
pG2K*kan*F1	TCACTACGTTGTATGAGAATAGTGAGG GCCAGTTTGTTGAAGATTAG	pG2K *kan* 5′ start
pG2K*kan*R1	ATACCTCCCCATTCAAAATGGTATGGC GCGCCTTATCAAAATG	pG2K *kan* 3′ end
pG2K*kan*F2	ACGTTCTTGCATATGAGAATAGTGAGG GCCAGTTTGTTGAAGATTAG	pG2K *kan* 5′ start
pG2K*kan*R2	GATGAGCCGGACTCAAAATGGTATGGC GCGCCTTATCAAAATG	pG2K *kan* 3′ end
*phaB*startF	AATGGCAGAGCGAATTGTTGTTGATCC AG	*phaB* start
*phaB*startR	TCACTATTCTCATACAACGTAGTGAGG GTGAGGG	659 bp 3′ of *phaB* start
*phbE*startF	AAAGGATCTCAAGAAAGTACATTCTCT AGGGCG	44 bp 3′ of *phbE* start
*phbE*startR	TCACTATTCTCATATGCAAGAACGTCGT TCC	676 bp 3′ of *phbE* start
*phaB*5′F	CATACCATTTTGAATGGGGAGGTATTC CGGC	600 bp 5′ of *phaB* start
*phaB*5′R	ATAATGCTCCCCCTTTCATTCGCCTCAC TCATGC	0 bp 5′ of *phaB* start
*phbE*5′F	CATACCATTTTGAGTCCGGCTCATCCTG ACTG	698 bp 5′ of *phbE* start
*phbE*5′R	TATATTCAAGCTACTTTTTCATTCTCTTT AAGCACCTC	9 bp 3′ of *phbE* start
*phaB*3′F	AGGCGAATGAAAGGGGGAGCATTATG CCGG	0 bp 3′ of *phaB* end
*phaB*3′R	AGCCTATGGAAATGCGACTAAATGATG CCATGATC	627 bp 3′ of *phaB* end
*phbE*3′F	AGAATGAAAAAGTAGCTTGAATATATG GTGTTTGC	0 bp 3′ of *phbE* end

Primer	Sequence 5′ – 3′	Target
*phbE*3′R	AGCCTATGGAAACCGTTGACGTCAGTT CAAG	638 bp 3′ of *phbE* end

NEBuilder HiFi DNA assembly of plasmid constructs

The transformation vectors pG2KΔ*phaB* and pG2KΔ*phbE* (Figure 2.11) were constructed using NEBuilder DNA HiFi assembly (www.*nebuilder.neb.com*). pG2KΔ*phaB* and pG2KΔ*phbE* were assembled using appropriate fragment combinations (Section 2.2.2.2) using the NEBuilder HiFi DNA assembly protocol for 4-6 fragment assembly as per the manufacturer's recommendations (Table 2.11). Briefly, equimolar amounts of the DNA fragments (five fragments per plasmid construct), to a cumulative total amount of 0.5 pmol of DNA fragments, were incubated together with the NEBuilder HiFi DNA assembly reaction mix (Table 2.11) at 50˚C for 1 h. Following incubation, reactions were stored at -20˚C.

Table 2.11: NEBuilder HiFi DNA assembly reaction components

	4–6 Fragment Assembly
DNA Molar ratio	Fragment ratios 1:1
Total amount of fragments	0.5 pmols
NEBuilder HiFi DNA Assembly Master Mix	10 µL
Nuclease free water	To 20 µL
Total volume	20 µL

2.2.2.3 Transformation of *E. coli* 10β

Following NEBuilder HiFi DNA assembly, 2 µL of the reaction was mixed with 60 µL of thawed (10 min) electrocompetent *E. coli* 10β cells. The suspension was then incubated on ice for 1 minute, transferred to a 0.1 cm electroporation cuvette and electroporated at 1.8 kV using a MicroPulser Electroporator. Cells were recovered in 1 mL NEB 10β Stable Outgrowth Medium (NEB, Ipswich, MA, USA) in a thin-walled test tube, with 1 h incubation at 37˚C and shaking at 200 rpm. Aliquots of

the recovered cells (50 µL) were then plated on LB agar supplemented with 100 µg/mL ampicillin (Sigma-Aldrich, St. Louis, USA), 5 µg/mL kanamycin and incubated at 37˚C for 24 h. Transformants were subcultured in 50 mL LB broth supplemented with 50 µg/mL kanamycin and incubated for 16 h at 37˚C with shaking (120 rpm).

2.2.2.4 Validation of pG2K-derived plasmid constructs

Following cultivation of transformed *E. coli* 10β (Section 2.2.2.3), pG2KΔ*phaB* and pG2KΔ*phbE* plasmid constructs were extracted from host *E. coli* 10β using a Monarch Plasmid Miniprep Kit (NEB, Ipswich, MA, USA) as previously described (Section 2.2.2.2). Following extraction, plasmid concentration and extract purity were evaluated using a Nanodrop One Spectrophotometer. Plasmid extract purity was further confirmed through agarose gel electrophoresis (results not shown), using a 0.8% TAE agarose gel, run at 100 V for 1 h. Samples were loaded in a 1:6 ratio with ROTI Load DNA and stained with ROTI Gel stain, using 2.5 µL ROTI Gel Stain per 50 mL agarose gel solution. Gels were visualised using a FAS-Digi Compact Gel Imaging System.

Plasmid extracts were then used as template DNA for PCR amplification of each target knockout cassette, spanning the target gene 5′ fragment, *kan*, target gene start fragment and target gene 3′ fragment (Figure 2.11). PCR amplification of the 2,907 bp pG2KΔ*phaB* knockout cassette was conducted using primers *phaB*5′F and *phaB*3′R (Table 2.10). PCR amplification of the 3,023 bp pG2KΔ*phbE* knockout cassette was conducted using primers *phbE*5′F and *phbE*3′R (Table 2.10). PCRs were conducted using OneTaq Hot Start Quick-Load 2x Master Mix with Standard Buffer (Table 2.7; Table 2.8; NEB, Ipswich, MA USA), as per the manufacturer's recommendations. PCR products were subsequently viewed through 0.8% agarose gel electrophoresis, run at 100 V for 1 h. Samples were loaded in a 1:6 ratio with ROTI Load DNA and stained with ROTI Gel stain, using 2.5 µL ROTI Gel Stain per 50 mL agarose gel solution. Gels were subsequently visualised using a FAS-Digi Compact Gel Imaging System. Following visualisation of correspondingly sized amplicons, plasmid constructs were assumed correctly assembled and subsequently used to conduct transformation of *P. thermoglucosidasius*.

2.2.2.5 Transformation of *P. thermoglucosidasius*

Transformation of *P. thermoglucosidasius* DSM 6285 was undertaken as per the protocol employed by Bacon *et al*, (2017). Frozen electrocompetent *P. thermoglucosidasius* DSM 6285 (Section 2.2.1.3) were thawed for 10 minutes on ice prior to use. Once thawed, 60 µL of electrocompetent cells were mixed with 200 ng of each transformation vector, respectively, and with the original pG2K backbone as a positive control. The suspensions were then transferred to 0.1 cm electroporation cuvettes and electroporated at 1.8 kV using a MicroPulser Electroporator. Cells were recovered in 1 mL SPYNG medium for 1.5 h at 52˚C and 200 rpm in a thin-walled test tube. Cells were then plated in 100 µL aliquots on SPYNG agar supplemented with 12.5 µg/mL kanamycin and incubated at 52˚C for 48 h. No successful transformations were achieved using the two plasmid constructs, despite the control showing colony development. Subsequently, varied combinations of plasmid volume (100, 200, 500, 1,000 ng), electroporation voltage (1.8, 2.5 kV), recovery period (1, 2, 3, 4 h), recovery medium (TGP, mLB, SPYNG), recovery agitation (120, 160, 200 rpm) outgrowth period (24, 48, 72 h) and recovery/outgrowth temperature (37, 42, 52, 60˚C) were used to facilitate transformation of the electroporated cells. Following no observed outgrowth following these optimization attempts, further transformation attempts were abandoned.

2.3 Results

2.3.1 Genetic engineering protocol adapted from Cripps *et al*. (2009)

2.3.1.1 Transformation of *E. coli* DH5α with plasmid constructs

Electroporation of *E. coli* DH5α with the pUB190, pPlasA and pPlasB vectors and the pUC19 positive control was successful. However, electroporations exhibited lower than ideal time constants ranging between 3.8-4.1 ms. pUC19 and its derived plasmids pPlasAΔ*phaB* and pPlasAΔ*phbE* showed much higher transformation efficiencies than the shuttle vector pUB190 and its derived plasmids pPlasBΔ*phaB* and pPlasBΔ*phbE* (Table 2.12). Validation of the PCR amplified sequences (generated using primers pUC19*LacZα*5R/*phaB*1541R-2 and

pUC19*LacZα*5R/*phbE*1454R-2) spanning 1,442 and 1,118 bp of the pPlasAΔ*phaB* and pPlasAΔ*phbE* plasmids, respectively, showed the correct insertion of the truncated mutant genes within the pUC19 backbones. Mutant gene homology arms showed 100% nucleotide identity with the wild type genes homology regions, not including the intended manipulations, which presented the intended sequence modifications (as illustrated in Figure 2.7). Given that construction of the pPlasB vector constructs relied on blunt end ligation of DNA fragments, unlike the sticky-end ligation of the shuttle vector and pPlasA plasmids, two mutant gene insertion orientations were observed following PstI-HF digestion of isolated colony plasmid extractions (Figure 2.9; Figure 2.13). The more prevalent orientation observed (orientation 2) was subsequently used in *P. thermoglucosidasius* transformations.

Table 2.12: Transformation efficiencies achieved for plasmid constructs used to transform *E. coli* DH5α

Plasmid	Transformation efficiency (cfu/µg DNA)
pUC19	1.05×10^9
pUB190	2.08×10^7
pPlasAΔ*phaB*	1.73×10^8
pPlasAΔ*phbE*	1.33×10^8
pPlasBΔ*phaB*	8.06×10^6
pPlasBΔ*phbE*	2.18×10^6

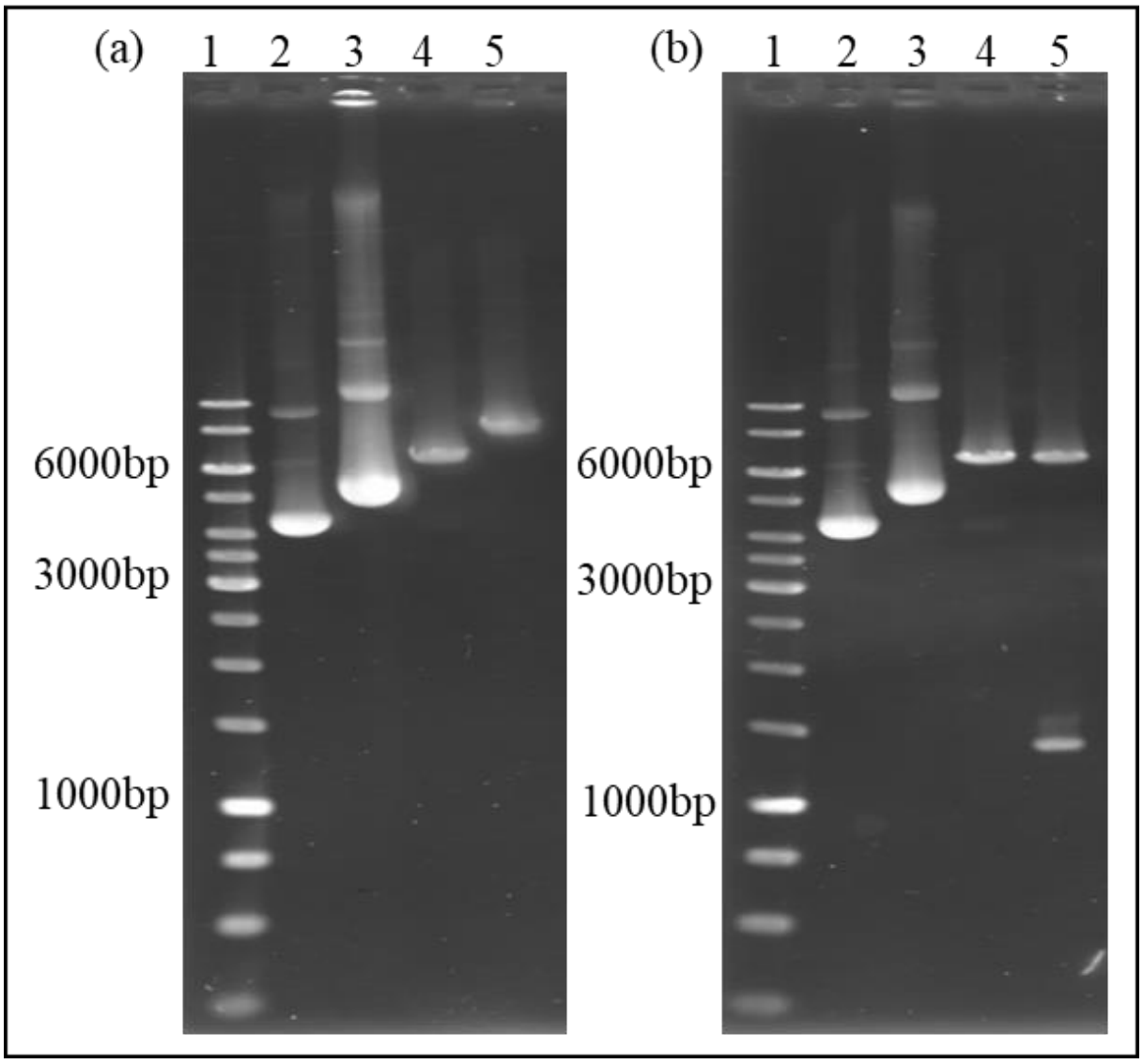


Figure 2.13: 0.8% Agarose gels showing mutant gene insertion orientations 1 (a) and 2 (b). (1) 1 Kbp DNA ladder (2) pUB190 plasmid (3) pPlasBΔ*phaB* plasmid (4) pUB190 PstI-HF restriction digest (5) pPlasBΔ*phaB* PstI-HF restriction digest.

2.3.1.2 Transformation of *P. thermoglucosidasius*

Despite numerous attempts to facilitate electroporation both under the conditions specified by Cripps *et al.*, (2009) and various optimisations of the protocol, we were unable to transform *P. thermoglucosidasius* DSM 6285 using either pPlasBΔ*phaB* or pPlasBΔ*phbE* plasmid constructs. Initial attempts to transform *P. thermoglucosidasius* DSM 2542^{T} according to the protocol used by Cripps *et al.* (2009), also resulted in no colony development following the outgrowth step after recovery. Successful growth of transformants was observed following

electroporation of 500 ng plasmid at 1.8 kV and recovery at 200 rpm for 2 h, with subsequent outgrowth for 48-72 h. However, only uncharacteristically very small transformant colonies were detected by 72 h of outgrowth. Notably, transformation efficiencies were much higher using a 2.5 kV applied voltage than 1.8 kV, although the higher voltage lowered the time constant achieved (Table 2.13). Transformation efficiencies of the pPlasBΔ*phaB* and pPlasBΔ*phbE* plasmids were relatively high, reaching 2.54 x 10^5 and 4.35 x 10^5 cfu/µg DNA. Transformant colonies that developed during outgrowth, however, exhibited comparably poorer growth than wild type colonies cultivated on the same medium (without kanamycin).

Table 2.13: Transformation efficiencies achieved and electroporation parameters used for *P. thermoglucosidasius* DSM 2542^T transformation with pPlasB constructs.

Plasmid	Electroporation Voltage (kV)	Time constant (ms)	Transformation efficiency (cfu/µg DNA)
pPlasBΔ*phaB*	1.8	5.1	1.06 x 10^4
pPlasBΔ*phaB*	2.5	4.3	2.54 x 10^5
pPlasBΔ*phbE*	1.8	4.9	5.30 x 10^3
pPlasBΔ*phbE*	2.5	4.2	4.35 x 10^5

2.3.1.3 Generation of primary and secondary *P. thermoglucosidasius* integrants

Initial attempts to generate primary integrants through dilution series (undiluted to 10^{-7} dilutions) plating of transformed *P. thermoglucosidasius* failed to exhibit growth at 68˚C by 72 h. Subsequent attempts to conduct this step using concentrated cell pellets were successful. Plates with cell pellets generated from >2 mL of liquid cultures developed comparably smaller colonies by 24 h post-inoculation compared to the wild type control (grown under the same conditions without kanamycin). Furthermore, colony size did not increase substantially with an additional 24-48 h growth period. Primary integration was confirmed by the presence of both the wild type and mutant gene sequences of *phaB* and *phbE*. This was identified through

PCR amplification and agarose gel electrophoretic screening of developed colonies after cultivation at 68˚C. Successful primary integrants displayed three bands, which indicated genomic integration of the plasmid alongside curing of autonomously replicating plasmid copies. These bands presumably constituted one wild type amplicon band, one mutant gene amplicon band and an additional band with mismatched amplicons, formed by annealing of single stranded wild type DNA amplicons with single stranded mutant DNA amplicons (Figure 2.14). A second banding pattern, wherein a single band in the same position as the predicted mutant gene band, was observed, likely the result of retention of autonomously replicative plasmid (Figure 2.14). With potentially higher copy numbers than the genomically located wild type gene, it was likely that amplification of the more abundant mutant gene harboured on retained plasmids were preferentially amplified, resulting in the formation of a single band. As a precaution, only colonies identified with the three-amplicon banding pattern were used in further experiments.

After serial subculturing of three presumptive primary integrants for each target gene, and subsequent plating on non-selective medium, colonies were replica plated to identify presumptive secondary integrants by way of kanamycin sensitivity. A total of 452/1,500 and 518/1,500 potential *phaB* and *phbE* mutant gene-containing colonies were kanamycin sensitive. Notably all colonies had resumed growth morphologies characteristic of the wild type *P. thermoglucosidasius*. PCR amplification and agarose gel electrophoretic analysis of the target gene sequences in kanamycin sensitive colonies only identified the presence of wild type target gene amplicons (Figure 2.15). Subsequent PCR amplification of the same gene targets in a subset of kanamycin resistant colonies also identified only wild type target gene amplicons (Figure 2.16). Amplification with plasmid specific primers in both the kanamycin sensitive and kanamycin resistant colonies demonstrated no banding patterns for either target gene, indicating the absence of either free or genomically integrated plasmid. Similarly, when using primers designed to amplify potential fragments between the kanamycin resistance gene and the target genes also failed to produce banding patterns for either target gene.

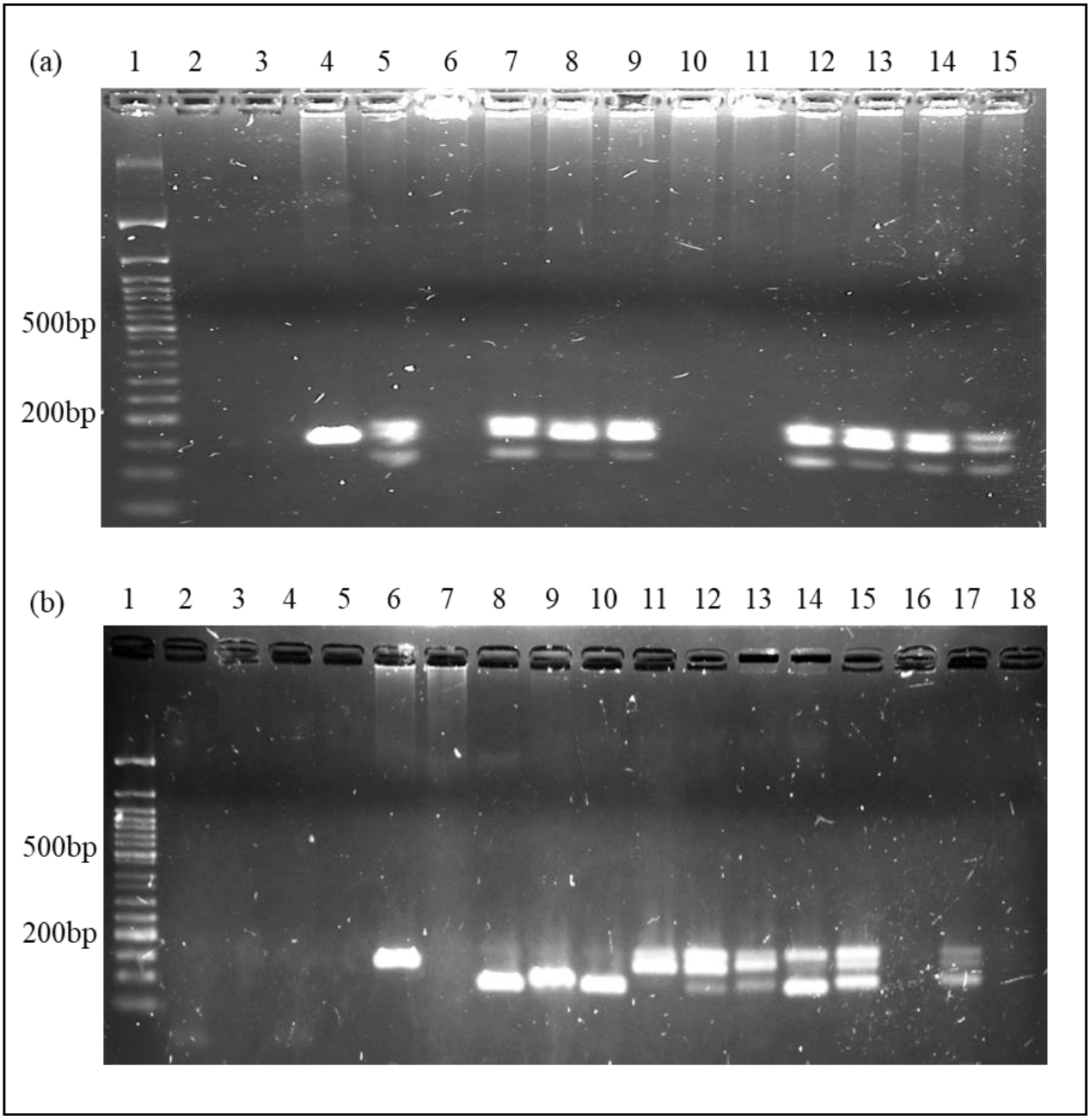


Figure 2.14: 3% Agarose gels showing wild type and mutant *phaB* (a) and *phbE* (b) gene amplicons from kanamycin resistant, presumed primary integrant *P. thermoglucosidasius* colonies. (a1) 50 bp DNA ladder (a2) DNA extraction -ve control (a3) PCR reaction -ve control (a4) wild type *phaB* amplicon (a5-15) presumed primary integrant colonies (b1) 50 bp DNA ladder (b2-3) DNA extraction -ve controls (b4-5) PCR reaction -ve controls (b6) wild type *phbE* amplicons (b7-18) presumed primary integrant colonies.

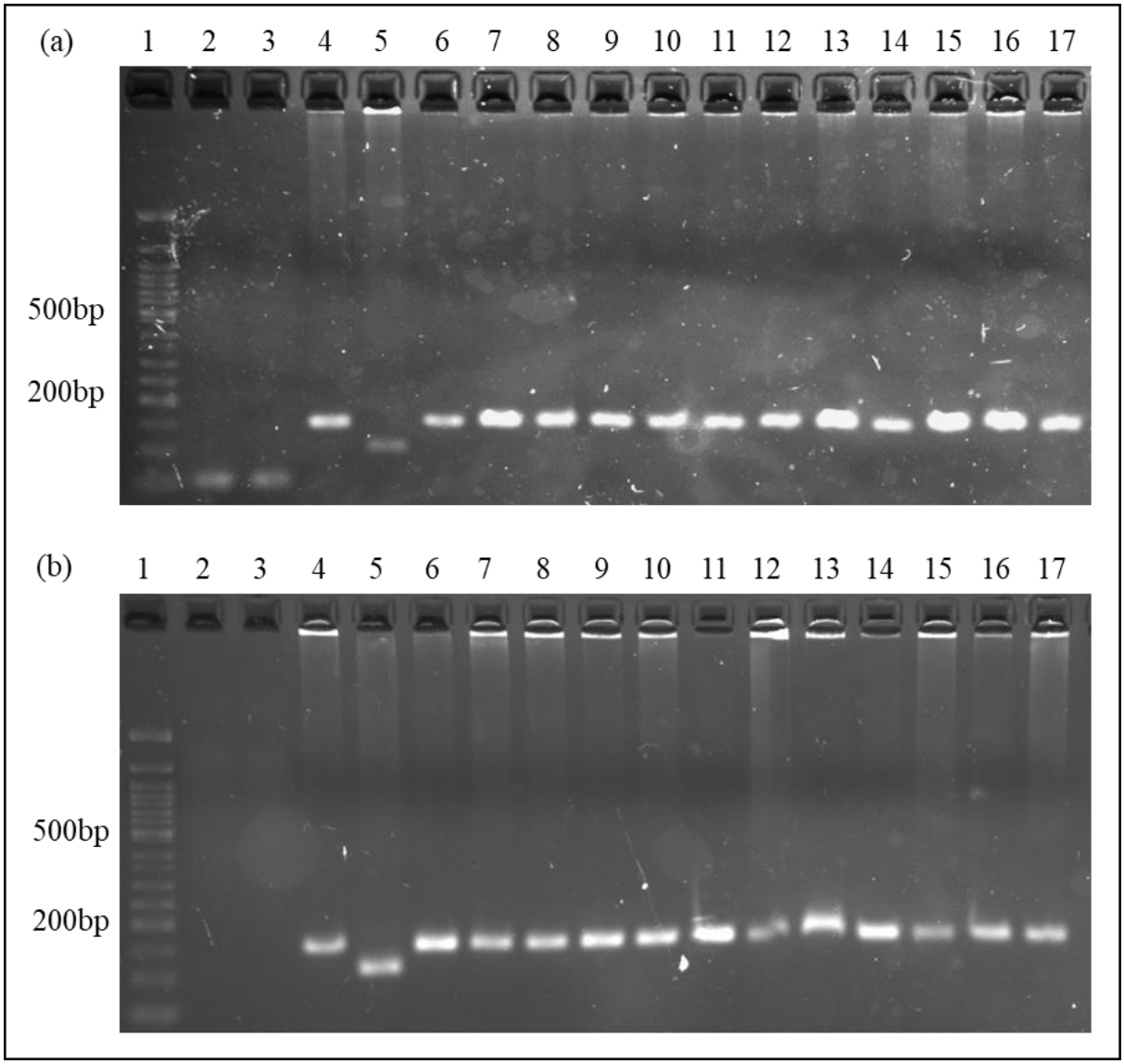


Figure 2.15: 3% Agarose gel showing wild type and mutant *phaB* (a) and *phbE* (b) gene amplicons from kanamycin sensitive, presumed secondary integrant *P. thermoglucosidasius* colonies. (a1) 50 bp DNA ladder (a2) DNA extraction –ve control (a3) PCR reaction –ve control (a4) wild type *phaB* amplicon (a5) pPlasBΔ*phaB* transformant *phaB* amplicon (a6-17) presumptive *phaB* mutant colonies. (b1) 50 bp DNA ladder (b2) DNA extraction –ve control (b3) PCR reaction –ve control (b4) wild type *phbE* amplicon (b5) pPlasBΔ*phbE* transformant *phbE* amplicons (b6-17) presumptive *phbE* mutant colonies.

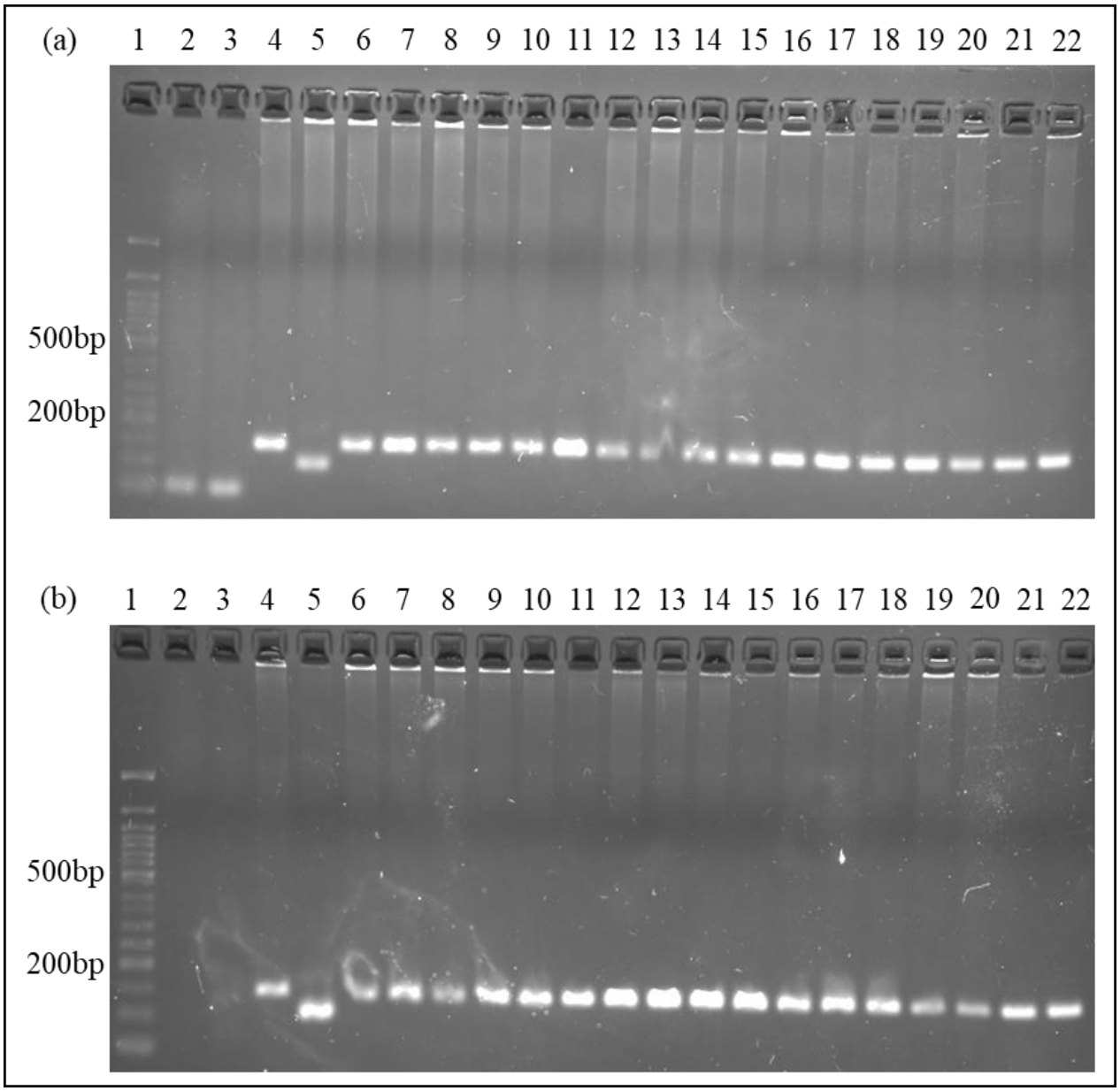


Figure 2.16: 3% Agarose gel showing wild type and mutant *phaB* (a) and *phbE* (b) gene amplicons from kanamycin resistant *P. thermoglucosidasius* colonies. (a1) 50 bp DNA ladder (a2) DNA extraction –ve control (a3) PCR reaction –ve control (a4) wild type *phaB* amplicon (a5) pPlasBΔ*phaB* transformant *phaB* amplicons (a6-17) kanamycin resistant colonies. (b1) 50 bp DNA ladder (b2) DNA extraction –ve control (b3) PCR reaction –ve control (b4) wild type *phbE* amplicons (b5) pPlasBΔ*phbE* transformant *phbE* amplicons (b6-17) kanamycin resistant colonies.

2.3.2 Genetic engineering protocol adapted from Bacon *et al.* (2017)

2.3.2.1 Transformation of *E. coli* 10β with pG2K-derived vectors

After assembly of the pG2kΔ*phbE and* pG2kΔ*phaB* transformation vectors, the plasmid constructs and pG2K positive control were successfully used to transform *E. coli* 10β. Similar to the electroporations conducted previously (Section 2.3.1.2), lower than ideal time constants ranging between 3.1-3.8 ms were observed. This, however, did not appear to impact transformation efficiency, as evident by the high transformation efficiencies obtained across all electroporations (Table 2.14).

Table 2.14: Transformation efficiencies achieved for the various plasmid constructs used to transform *E. coli* 10β

Plasmid	Transformation efficiency (cfu/µg DNA)
pG2K	1.18×10^9
pG2KΔ*phaB*	3.56×10^8
pG2KΔ*phbE*	5.19×10^8

2.3.2.2 Transformation of *P. thermoglucosidasius* with pG2KΔ*phaB* and pG2KΔ*phbE*

Despite numerous attempts to facilitate electroporation and subsequent genomic integration of plasmids into *P. thermoglucosidasius* DSM 6285, we were unable to successfully transform *P. thermoglucosidasius* DSM 6285 using either pG2KΔ*phaB* or pG2KΔ*phbE* plasmid constructs. Time constants for electroporations ranged between 4.8-5.1 ms. Transformation of *P. thermoglucosidasius* DSM 6285 with pG2K as a positive control was only achieved when culturing the strain above 52˚C, with transformation efficiencies of up to 3.18×10^5 achieved, indicating that the strain was receptive to the electroporation protocols.

2.4 Discussion

P. thermoglucosidasius is receiving increasing attention for its capacity to produce H_2 via the water-gas shift reaction. This activity is linked to a chromosomally co-localized set of genes coding for a carbon monoxide dehydrogenase (CooCSF) and group 4a [NiFe]-hydrogenase (PhcA-L). In addition, two predicted [NiFe]-uptake hydrogenases are encoded on the chromosome of the species, namely a group 1d (PhaA-K) and group 2a (PhbA-P) [NiFe]-uptake hydrogenase (Mohr *et al.*, 2018a). H_2 uptake hydrogenases are widely distributed amongst bacterial taxa and variably associate H_2 oxidation with both aerobic and anaerobic respiration processes. (Vignais and Billoud, 2007; Schwartz *et al.*, 2013; Greening *et al.*, 2016). Considering the potential capacity for the two uptake hydrogenases encoded by *P. thermoglucosidasius* to act as H_2 sinks and negatively affect H_2 production, we endeavoured to further characterise the roles of these hydrogenases in *P. thermoglucosidasius* DSM 6285. This was attempted through targeted knock-out mutagenesis of the genes coding for the major catalytic subunits (PhaB and PhbE respectively) of the two uptake hydrogenases. However, despite the application of two distinct genetic engineering methodologies and evaluation of various optimisation parameters, viable mutants for either target genes could not be generated.

P. thermoglucosidasius DSM 2542^T was successfully transformed with both pPlasBΔ*phaB* and pPlasBΔ*phbE*, however, transformants displayed noteworthy reduction in colony development compared to the wild type strain. Similarly, growth at 68˚C during selection for primary integrants was also very poor compared to wild type growth under the same conditions. This, in combination with the low plasmid integration rates observed, likely necessitated the need for greater concentrations of cells to identify successful integrants (Levin and Bergstrom, 2000; Carroll and Wong, 2018). Low-integration rates may also account for transformation failure observed using pG2K-derived transformation vectors in the genetic engineering protocol adapted from Bacon *et al.* (2017). Additionally, without a thermophilic origin of replication, plasmid counts in the electroporated cells were likely very low following the recovery period, further reducing the

likelihood for integration through homologous recombination (Levin and Bergstrom, 2000; Thomas and Nielsen, 2005). Notably, however, resumption of wild type-like colony growth, coincided with loss of detection of either mutant gene, both under kanamycin selective and non-selective conditions. The observation that growth was still impacted following removal of autonomously replicating plasmids suggested that this could not be exclusively the result of potential fitness costs imposed on the cells by maintenance of the plasmids (Gama *et al*., 2018).

It has been suggested that expression of plasmid-borne resistance genes are responsible for the bulk of plasmid fitness costs (Rajer and Sandegren, 2022). Given that those colonies that retained kanamycin resistance demonstrated comparable growth to those that reverted to the apparent wild type genotype, suggested that this was likely not the cause of observed growth reduction. Retention of kanamycin resistance in conjunction with no detection of the transformation vector or mutant gene sequences may be the result of recombination events occurring between the transformation vector and the naturally occurring plasmids in the strain (Rodríguez-Beltrán *et al.*, 2021). *P. thermoglucosidasius* DSM 2542^{T} harbours two native plasmids, one of 76 kb and one of ~46 kb (Sheng *et al.*, 2016). Production of chimeric plasmids is a characteristic feature of plasmids harboured amongst Firmicutes, which frequently display high degrees of plasmid mosaicism (Lanza *et al.*, 2015). Host defence systems such as restriction-modification and CRISPR-Cas systems, elements of which are present on the genome of *P. thermoglucosidasius* DSM 2542^{T} (Chapter four: Section 4.3.4.3), have been shown to increase the frequency of plasmid recombination events. As such, recombination between the native *P. thermoglucosidasius* plasmids and the transformation vectors resulting in retention of the kanamycin resistance gene is not implausible (Bernheim and Sorek, 2020; Bezalel-Buch *et al.*, 2020).

As neither plasmid fitness nor resistance gene maintenance were likely causes of the deleterious growth effects observed, an alternative explanation for the observed reductions in growth may be the consequence of disruption of the two target genes. Both uptake hydrogenase loci are universal among *P. thermoglucosidasius* (Mohr *et al.*, 2018a). While the roles of both uptake hydrogenases in *P.*

thermoglucosidasius are unclear, the roles of other group 1d and group 2a hydrogenases have been characterised in various taxa, with prominent roles in mediating respiratory H_2 uptake associated with energy conservation, H_2 recycling and maintaining pH homeostasis (King and Przybyla, 1999; Pandelia *et al.*, 2012; Greening *et al.*, 2016).

H_2 oxidation by group 1d [NiFe]-uptake hydrogenases is proposed to be coupled with aerobic respiration or respiratory reduction of anaerobic oxidants (Greening *et al.*, 2016). Electrons liberated through H_2 oxidation may subsequently be integrated into the quinone pool through membrane-integral *b*-type cytochrome components (Sawers *et al.*, 1986; Brugna-Guiral *et al.*, 2003; Burgdorf *et al.*, 2005; Fritsch *et al.*, 2011). Prominent taxa harbouring these hydrogenases include the aerobic chemolithoautotroph *Wautersia eutropha* and the facultatively anaerobic heterotroph *E. coli* (Greening *et al.*, 2016). Previous mutational studies targeting group 1d [NiFe]-uptake hydrogenases in these taxa have yielded variable results. *E. coli* encode four distinct [NiFe]-hydrogenases, belonging to groups 1c (Hyd-2), 1d (Hyd-1) and 4a (Hyd-3) (Søndergaard *et al.*, 2016), with uncertainty surrounding the role and classification of the Hyd-4 hydrogenase (Sargent, 2016). Although the group 1d hydrogenase (Hyd-1) has been identified as not essential for growth in *E. coli* (King and Przybyla, 1999), inactivation of the hydrogenase by (Maeda *et al*, (2007), resulted in a minor reduction in cell growth. Alternatively, *W. eutropha* harbours four [NiFe]-hydrogenases, belonging to groups 1d, 1h/5, 2b and 3c respectively (Sargent, 2016; Søndergaard *et al.*, 2016). Inactivation of the group 1d [NiFe]-uptake hydrogenase of *W. eutropha* resulted in reduced growth when coupled with lithoautotrophic dependent growth conditions (Hogrefe *et al.*, 1984; Søndergaard *et al.*, 2016). Lithoautotrophic growth was further halted entirely when combined with inactivation of its soluble group 3d [NiFe]-uptake hydrogenase (Bernhard *et al.*, 1996).

In contrast, inactivation of group 2a [NiFe]-uptake hydrogenases have been shown to strongly influence growth patterns of mutant bearing strains. Group 2a [NiFe]-uptake hydrogenases have similarly been implicated in aerobic respiration, although, are also commonly associated with recycling of H_2 produced through nitrogenase activity or fermentative pathways in aerobic soil bacteria and

Cyanobacteria (Houchins and Burris, 1981; Greening *et al.*, 2014a; Koch *et al.*, 2014). *Mycobacterium smegmatis*, an obligate aerobic heterotroph, capable of atmospheric H_2 scavenging, is a prominent and well-studied bacterium harbouring a group 1c and 2a [NiFe]-hydrogenase (Greening *et al.*, 2014a; Greening *et al.*, 2016). Inactivation of either of these uptake hydrogenases resulted in reduced growth and long-term viability of mutant harbouring strains, suggesting that they play important and distinct roles in cell function (Berney and Cook, 2010; Berney *et al.*, 2014; Greening *et al.*, 2014b; Cordero *et al.*, 2019; Islam *et al.*, 2019).

It is plausible that the Pha and Phb uptake hydrogenases in *P. thermoglucosidasius* participate in presently unknown but important roles in maintaining cell viability, and that disruption of genes involved in the synthesis of these hydrogenases adversely impacts cell growth. Following plasmid curing in the first applied protocol, it is possible that a proportion of the cell populations harboured the desired chromosomally integrated mutant phenotypes, however, these population subsets may simply have been outcompeted by more viable and actively growing wild type revertants (Figure 2.2). Were this to be the case, application of knock-out mutants of either hydrogenase in downstream applications would not have been ideal, due to highly limiting growth potential/viability. While this study cannot confirm the necessity or function of these hydrogenases in *P. thermoglucosidasius*, application of alternative recombinant independent strategies, such as CRISPR-Cas mediated down-regulation of the target genes (Otoupal and Chatterjee, 2018), or through posttranslational gene silencing (Nakashima and Miyazaki, 2014), may shed further light on their relevance.

2.5 Conclusions

In order to investigate the role of two putative [NiFe]-uptake hydrogenases, Pha and Phb, encoded on the genome of *P. thermoglucosidasius* in H_2 metabolism, we attempted to functionally inactivate the catalytic subunits of these hydrogenases through knockout mutagenesis of their respective genes, *phaB* and *phbE*. We were not able to produce stable strains bearing functional knockouts of either target gene despite using two different approaches for homologous recombination-based knockout mutagenesis. Following introduction of transformation vectors, bearing

presumed non-functional mutant copies of the two target genes, into *P. thermoglucosidasius* DSM 2542^{T}, growth of the strain was markedly reduced. Notably, neither the fitness costs of the transformation vectors nor their encoded antibiotic resistance markers could account for reductions in host viability. Resumption of wild type-like growth of transformants coincided with loss of detection of mutant genes encoded on the transformation vectors, suggesting deleterious impacts to growth were the result of the presence of these mutant genes. The Pha and Phb hydrogenases harboured by *P. thermoglucosidasius* may play an important role in maintaining cell viability. Further exploration of the roles that these hydrogenases play in *P. thermoglucosidasius* using alternative genetic engineering techniques may assist in resolving their functional relevance.

CHAPTER THREE: THE EFFECTS OF SYNTHETIC SYNGAS FEEDSTOCKS AND OXYGEN PERTURBATION ON HYDROGEN PRODUCTION BY *P. THERMOGLUCOSIDASIUS*

Abstract

The facultatively anaerobic thermophile *Parageobacillus thermoglucosidasius* is able to produce hydrogen gas (H_2) under anoxygenic atmospheres, through the oxidation of carbon monoxide via the water-gas shift (WGS) reaction. To date this process has been evaluated under controlled conditions, with gas feedstocks comprising carbon monoxide and variable proportions of air, nitrogen and hydrogen. Ultimately, an economically viable hydrogenogenic system would make use of industrial waste/synthesis gases that contain high levels of carbon monoxide, but which may also contain contaminants such as hydrogen, oxygen and variable chain-length hydrocarbons, which may be toxic to *P. thermoglucosidasius* or hamper the WGS reaction process. Here we evaluated the effects of syngas mimetic feedstocks on WGS reaction -driven H_2 gas production by *P. thermoglucosidasius* DSM 6285 in small-scale fermentations. Furthermore, as the WGS reaction is an anoxygenic process, we evaluated the influence of oxygen perturbation on *P. thermoglucosidasius* hydrogenogenesis. Improved H_2 gas production yields and faster onset towards hydrogen production were observed when anaerobic synthetic syngas feedstocks were used, at the expense of biomass accumulation. Oxygen supplementation improved biomass accumulation, but reduced hydrogen yields in accordance with the level of oxygen supplied. However, H_2 gas production was observed at low oxygen levels. Oxygen supplementation also induced rapid acetate consumption, likely to supplement growth. These capacities further support the development of *P. thermoglucosidasius* towards continuous hydrogenogenic operation making use of industrial syngas feedstocks.

3.1 Background

Increasing energy demands and dwindling fossil fuel reserves are driving the development and implementation of environmentally friendly energy production technologies, using renewable resources (IRENA, 2019). H_2 gas is regarded as an ideal energy carrier, with carbon neutral combustion characteristics and high energy density (Elbeshbishy *et al.*, 2017). While H_2 gas is currently produced almost exclusively through the reformation of non-renewable fossil fuels, various alternative renewable H_2 gas production processes have and are currently being developed (Abdalla *et al.*, 2018).

Numerous microbial taxa are capable of producing H_2 gas from a wide spectrum of organic and inorganic feedstocks in an environmentally friendly and energetically favourable manner (Hosseini and Wahid, 2016; van Niel, 2016). These are subdivided into light dependent and independent processes with those incorporating dark-, photo-, and hybrid-fermentative processes having dominated mainstream research (Hosseini and Wahid, 2016). Despite yielding promising results, these processes are hindered by a number of limitations, including cost, feedstock availability/suitability, turnover rate/ratio and scalability, that restrict their industrial scale application (Abdalla *et al.*, 2018). Less well-explored hydrogenogenic processes, such as the biological WGS reaction, are promising alternatives to the aforementioned systems.

The reversible WGS reaction involves the oxidation of carbon monoxide to carbon dioxide and the reduction of water to molecular hydrogen at an equimolar ratio ($CO + H_2O \rightarrow CO_2 + H_2$ $\Delta G^{O'} = -20$ kJ/mol CO; Alfano and Cavazza, 2018). The potential to conduct the WGS reaction has been identified in more than twenty-five bacterial species spanning the phyla Proteobacteria, Firmicutes and Dictyoglomi (Fukuyama *et al.*, 2020). The biological WGS reaction has been best characterised in *Rhodospirillum rubrum*, *Carboxydothermus hydrogenoformans* and *Parageobacillus thermoglucosidasius* (Haddad *et al.*, 2014; Hosseini *et al.*, 2015; Mohr *et al.*, 2018a). The reaction proceeds via the activity of a carbon monoxide dehydrogenase (CODH)/H_2 evolving hydrogenase complex (Adachi *et al.*, 2020; Greening *et al.*, 2016; Schoelmerich and Müller, 2019). Both the CODH and H_2

evolving hydrogenase enzymes are highly sensitive to O_2, greatly restricting their industrial applicability (Merrouch *et al.*, 2015; Lu and Koo, 2019). Although several hydrogenases demonstrate greater O_2-tolerance, they exhibit poorer H_2 gas turnover efficiencies (Lu and Koo, 2019). As such, the majority of bacterial taxa capable of conducting the WGS reaction are strict anaerobes. One exception, however, is the facultatively anaerobic thermophile *P. thermoglucosidasius*, which carries a Ni-CODH/ group 4a [NiFe]-hydrogenase complex and is capable of the WGS reaction following oxygen depletion (Mohr *et al.*, 2018a; Mohr *et al.*, 2018b; Aliyu *et al.*, 2020; Imaura *et al.*, 2023). By optimising growth medium composition and pH, fermentation temperatures and CO concentration, H_2 gas production yields have been improved (Mohr *et al.*, 2019).

While promising developments are being made towards optimizing the use of the *P. thermoglucosidasius* WGS reaction in single batch applications, further development is necessary. Key advancements include progression of the process towards more continuous operation and evaluation of the system efficiency when utilising gas feedstocks more reflective of industrially/commercially available waste gases. Industrial waste gasses or synthesis gases (syngas) are highly variable, although, are typically comprised of CO, H_2, CO_2 and CH_4, as well as trace amounts of O_2 (Ni *et al.*, 2006; Benevenuti *et al.*, 2021). Here we investigated the effects of gas feedstocks with compositions more reflective of industrial waste gases on H_2 gas production by *P. thermoglucosidasius* through bioassays in conjunction with GC and HPLC analysis of select gases and metabolites. Furthermore, as optimal *P. thermoglucosidasius* growth is observed under aerobic atmospheres, balancing aerobic conditions conducive to sustained biomass accumulation and anaerobic conditions obligatory for H_2 production is necessary for more advanced continuous/semi-continuous hydrogenogenic operation. Therefore, the impacts of temporal and volumetric O_2 perturbations (as would occur with industrial waste gases and syngas) on *P. thermoglucosidasius* DSM 6285 hydrogenogenesis were further evaluated in this study.

3.2 Methods and materials

3.2.1 Strains and media

P. thermoglucosidasius DSM 6285, acquired from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Braunschweig, Germany), was maintained and cultured routinely in modified Luria Bertani medium (mLB) composed of 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 1.25 mL/L of 10 g/L NaOH and 1 mL/L of filter-sterilized stock solutions of nitrilotriacetic acid (1.05 M), $MgSO_4 \cdot 7H_2O$ (0.59 M), $CaCl_2 \cdot 2H_2O$ (0.91 M) and $FeSO_4 \cdot 7H_2O$ (0.04 M) added post autoclaving. Experimental cultures were grown in a previously optimised modified ammonium sulphate medium (mASM) (Aliyu *et al.*, 2021), comprised of 8.7 mM citric acid, 20.2 mM $MgSO_4$, 10 mM K_2SO_4, 22.6 mM NaH_2PO_4, 0.8 mM $CaCl_2$, 25 mM $(NH_4)_2SO_4$, 4.16 mM glucose, 1 mL/L of filter-sterilized trace elements solution containing 6 mM H_2SO_4, 0.1 mM $CuSO_4$, 0.2 mM $CoSO_4$, 0.5 mM $ZnSO_4$, 2.29 mM $FeSO_4$, 0.3 mM $NiSO_4$, 0.9 mM $MnSO_4$ and 0.1 mM H_3BO_3 and 0.02 mM biotin (Carl Roth, Karlsruhe, Germany), 20 mL/L MEM amino acids solution (50x), 10 mL MEM non-essential amino acids solution (100x) and 10 mL MEM vitamin solution (100x; Thermo Scientific, Schwerte, Germany).

3.2.2 Experimental setup

Pre-cultures were grown in 100 mL mLB in 500 ml baffled shake flasks, inoculated with 40 µL of glycerol stock and incubated aerobically at 55˚C for 14 h at 120 rpm in an Infors HT Thermotron (Infors AG, Bottmingen, Switzerland). Following incubation, appropriate volumes of pre-culture were used to inoculate experimental cultures to an initial OD_{600} of approximately 0.1, which has previously been established as an optimal starting inoculum to minimise the lag period between inoculation and commencement of H_2 production (Mohr *et al*., 2019). Experimental cultivations were conducted in 250 ml serum bottles sealed with a rubber stopper, containing 50 mL ASM with differing gas headspace compositions at atmospheric pressure.

To investigate the influences of variable industrial syngas mixtures on the *P. thermoglucosidasius* DSM 6285 WGS reaction, the strain was cultured in several

gas headspace atmospheres (Table 3.1). Four sets of experimental cultures were cultivated in triplicate. The first set served as an aerobic control grown under an optimal gas atmosphere composed of 50% air and 50% CO. The second set served as an anaerobic control, with a gas headspace composed of 50% N_2 and 50% CO. This was achieved by flushing bottle headspaces with N_2 for 5 minutes at a flow rate of 0.2 L/minute, followed by injection of the bottles with CO to an overpressure of 1 bar, with subsequent equilibration to atmospheric pressure. The third culture set was cultivated under a 100% atmosphere of a synthetic syngas mixture (Air Liquide, Düsseldorf, Germany). The fourth culture set was cultured under an initial gas headspace comprising 50% N_2 and 50% synthetic syngas mixture. This was achieved by flushing bottle headspaces with N_2 for 5 minutes at a flow rate of 0.2 L/minute, then injecting the bottles with the synthetic gas mixture to an overpressure of 1 bar, followed by subsequent equilibration to atmospheric pressure. Following inoculation, bottles were incubated at 55˚C for 134 h at 120 rpm.

Table 3.1: Composition of synthetic syngas stock and initial gas headspace mixtures

	Gas Component (%)								
Atmosphere	CO_2	CO	CH_4	N_2	H_2	C_4H_{10}	C_2H_6	C_3H_8	O_2
Synthetic syngas stock	~43.99	36.90 ± 0.07	6.22 ± 0.12	5.07 ± 0.10	3.64 ± 0.07	2.42 ± 0.05	1.03 ± 0.02	0.73 ± 0.01	0
Aerobic Control	0.42 ± 0.03	48.61 ± 0.11	0	39.73 ± 0.10	0	0	0	0	9.93 ± 0.73
Anaerobic (N_2) Control	0.31 ± 0.05	45.51 ± 0.36	0	52.90 ± 0.44	0	0	0	0	0
Syngas 50% mixture*	17.08 ± 1.64	17.28 ± 0.77	3.91 ± 0.37	59.37 ± 2.02	1.05 ± 0.01	~1.21	~0.52	~0.37	0
Syngas 100% mixture*	41.65 ± 1.60	38.01 ± 0.92	6.30 ± 0.09	9.68 ± 0.50	3.11 ± 0.00	2.42 ± 0.05	1.03 ± 0.02	0.73 ± 0.01	0

*Theoretical composition inferred from dilution of stock gas.

To investigate the influence of O_2 addition on the hydrogenogenic activity of the *P. thermoglucosidasius* DSM 6285 WGS reaction, two experiments were conducted. In all cases, the serum bottle gas headspace of each stoppered replicate was reconstituted to an ideal initial atmosphere of 50% air and 50% CO prior to inoculation (Aliyu *et al.*, 2021). This was achieved by flushing bottles with CO to an overpressure of 1 bar, then equilibrating the headspace to atmospheric pressure. O_2 was injected into the gas headspace of bottles containing actively fermenting cultures at various stages of the fermentation. Four time points (post-inoculation) were selected for O_2 addition, reflective of distinct periods of the WGS reaction fermentation profile. These included the mid aerobic phase (14 h), the commencement of the anaerobic phase (24 h), the early-mid hydrogenogenic phase (38 h) and the late hydrogenogenic phase (48 h) as previously identified (Aliyu *et al.*, 2020). The experiment was carried out in two batches, comprising three sets of cultures, one control and two experimental cultures, respectively. In each batch, cultures were cultivated in triplicate with one set serving as a control where no O_2 was added at any timepoint. The two remaining culture sets in each batch were designated a time point where a defined volume of O_2 was injected. O_2 volumes injected were ~5% of the initial serum bottle gas headspace volume, calculated as $v_{O_2} = v_t - v_m - v_i$, where v_{O_2} = volume O_2 added, v_T = total serum bottle volume, v_m = volume of the liquid culture medium and v_i = inoculum volume. Following inoculation, bottles were incubated at 55˚C for 120 h at 120 rpm, with O_2 being added to the appropriate replicate set at the designated time points.

Finally, the influence of different volume O_2 additions on the WGS system at a single time point was investigated. A 38-hour post-inoculation timepoint, representative of the peak hydrogenogenic period was selected based on the fermentation profiles of the previous experiments. Five sets of experimental cultures were cultivated in triplicate under a gas atmosphere of 50% air and 50% CO, with one set serving as a control where no O_2 was added. Each of the four remaining culture sets were designated an O_2 volume to be injected at 38 h, either ~1, 3, 5, or 10% of the initial serum bottle gas headspace volume, calculated as described above. Following inoculation, bottles were incubated at 55˚C for 120 h

at 120 rpm, with the appropriate O_2 volumes being added to the respective replicate sets at 38 h.

3.2.3 Sampling and analytics

Liquid culturing medium and gas headspace sampling was conducted at various time points for each experiment. For the experiment where different syngas mixtures were evaluated, samples were taken at 0, 14, 24, 38, 48, 62, 72, 86, 96, 110, 120 and 134 hours post-inoculation. For the first oxygen supplementation experiment, where a fixed O_2 volume was supplemented at various timepoints in the WGS reaction fermentation, sampling points were 0, 10, 14, 24, 34, 38, 48, 58, 72, 84, 96, 110 and 120 hours post-inoculation. Sampling points for the final experiment, where variable O_2 volumes were supplemented during the peak hydrogenogenic period were 0, 14, 24, 38, 41, 44, 48, 62, 72, 86, 96, 110, 120, and 134 hours post-inoculation. In both O_2 addition experiments, prior to the addition of O_2 at applicable time points, both a liquid and gas sample were taken and immediately followed by a second gas sampling post-O_2-addition. A volume of 3 mL gas headspace was drawn at each time point and analyzed using a 300 Micro GC Fusion Gas Analyzer (Inficon, Switzerland), with argon and helium as carrier gases, connected with Molsieve 5Å and PLOT Q columns for data acquisition. Headspace pressure was determined before sampling using a manometer (GDH 14 AN, Greisinger electronic, Regenstauf, Germany) and temperature was maintained at 55˚C.

Gas composition was calculated based on the ideal gas law ($n = P * \frac{V}{R} * T$); where n = number of mols of gas; P = pressure of gas (1.013 bar + gas mixture over/under-pressure (p) * 10^5); V = volume of the gas (Start V - number of mL removed prior to each GC measurement); R = universal gas constant (8.314 J * $Kelvin^{-1}$ * mol^{-1} or kg * $m^2sec^{-2}mol^{-1}Kelvin^{-1}$); T = temperature of sample (333.15 K (55 °C)). Calculation of the gas amount was done using the following equation; n_{GAS} = (1.013 + p) * 10^2 * V / 2769.8091 – n_{H2O}; where n = number of mols of gas, p = over/under-pressure of the gas, V = volume of the gas (Start V - number of mL removed prior to each GC measurement). Final gas concentration was calculated as n_s = n_{GAS} *

GC % of specific gas (or mol % of specific gas) /100; where n_s = number of mols of a specific gas. (Mohr *et al.*, 2018a, 2019b).

A volume of 1 mL of liquid culture medium was sampled at each sampling point for growth and used for subsequent HPLC analysis. Growth was estimated using absorbance (OD_{600}) values using an Ultrospec 1100 pro spectrophotometer (Amersham Biosciences, Uppsala, Sweden) and pH was measured using a Profilab pH 597 (Xylem Analytics, Weilheim, Germany). Following absorbance and pH measurement, samples were centrifuged at 16,000 x g for 10 minutes and supernatants transferred to a clean 1.5 mL microcentrifuge tube for storage at -20˚C. Thawed supernatants (100 µL) were dispensed in 1.5 mL HPLC autosampler vials fitted with micro-inserts. Glucose and twelve central carbon metabolites, 1-propanol, acetate, butyrate, ethanol, formate, glyoxylate, iso-butyrate, iso-valerate, lactate, propionate, succinate and valerate were monitored over 0–2.0 g/L concentration ranges. Samples were analyzed using an Agilent 1100 series HPLC system (Agilent Technologies, Waldbronn, Germany) connected to a wavelength detector and refractive index detector with a 50 mm long pre-column (model Rezex ROA-Organic Acid H^+ 8% Guard Column) and a 300 mm long separation column (model Rezex ROA-Organic Acid H^+ 8%). Analysis was conducted using a 5 mM H_2SO_4 mobile phase, 50˚C column temperature, 0.6 mL/min flow rate for 40 min per sample and injection volume of 20 µL. Data acquisition and handling was conducted using Chemstation (Agilent Technologies, Waldbronn, Germany) and datasets were summarized and figures plotted using R v. 4.3.2 (Wickham *et al.*, 2019; R Core Team, 2021).

3.2.4 Statistical analysis

Statistical analysis for the Gas chromatography and HPLC was performed in R v. 4.3.2 (Wickham *et al.*, 2019; R Core Team, 2021). To determine significant changes in gas headspace and metabolite profile compositions between the controls and experimental cultures, one way analysis of variance (ANOVA) was conducted between the respective constituents within each experiment, followed by post-hoc Tukey honest significant difference (HSD) tests. To determine significant changes in H_2 following O_2 addition in both the temporal and volumetric experiments, H_2

levels at each sampling point within culture sets were subjected to ANOVA, followed by post-hoc Tukey HSD tests.

3.3 Results

3.3.1 Anaerobic synthesis gas mimetic feedstocks promote earlier onset of hydrogenogenesis at the expense of biomass accumulation

Previous studies have characterised the fermentation profile of *P. thermoglucosidasius* DSM 6285 under various aerobic and anaerobic CO enriched headspaces, citing a 50% CO/50% air atmosphere as an ideal balance between suitable growth conditions and facilitating downstream WGS reaction conditions (Mohr *et al.*, 2019; Aliyu *et al.*, 2020; Aliyu *et al.*, 2021; Imaura *et al.*, 2023). Here we observed distinct fermentation profiles between the aerobic and anaerobic (anaerobic control, syngas (50%) and syngas (100%) atmospheres (Figure 3.1).

The aerobic control growth profile corresponded with previously observed patterns. This included an initial sharp increase in absorbance during the aerobic phase, peaking at 0.327 ± 0.008 by 48 h (coincidental with peak WGS reaction activity), followed by gradual decreases towards the end of fermentation (Figure 3.1). All anaerobic atmospheres exhibited significantly poorer growth throughout the fermentations (Appendix B Supp. Table S3.1). Slight, albeit much less pronounced increases were observed for the anaerobic control (50%) and syngas (50%) atmospheres during the first 14 h of the experiment. This growth was not, however, shared by the syngas (100%) atmosphere replicates, which increased negligibly, plateauing similarly to the N_2 and synthetic syngas (50%) atmosphere replicates at 0.111 ± 0.007. Growth of the syngas (100%) atmosphere replicates was also atypical, exhibiting a granular flocculant-type of growth compared to the turbid/cloudy appearance observed in the other atmospheres (results not shown).

In the early phases of each fermentation, pH values decreased across all atmospheres to varying extents (Figure 3.1). The aerobic control, reached a low pH of 5.82 ± 0.02 by 14 h, accompanied and likely accounted for by concerted acetate production observed in this period. Anaerobic atmospheres demonstrated similar pH profiles to the aerobic control, however, reaching significantly lower pH values

of 5.31 ± 0.04 by 38 h (Appendix B Supp. Table S3.1). These atmospheres exhibited strong concurrent production of lactate, formate and acetate which plateaued in conjunction with pH stabilisation, and were likely responsible for the observed pH fluctuations. Following initial decreases, pH stabilized across all atmospheres, with the aerobic control stabilizing at a final value of 5.87 ± 0.02 and the anaerobic atmospheres at 5.37 ± 0.03.

H_2 production was observed within 14 h in the anaerobic atmospheres (Figure 3.1), while the aerobic control only showed H_2 production by 24 h. Notably, H_2 production in the aerobic control was observed with concurrent O_2 levels of 0.95 ± 0.06%. Peak H_2 productivity for the aerobic control was much higher than the anaerobic atmospheres with peak productivity observed between 38-48 h at 0.0137 ± 0.0048 $mmol.h^{-1}$. For the anaerobic control and syngas (100%) atmospheres, lower peak productivity was observed earlier between 24-38 h at 0.0063 ± 0.0005 $mmol.h^{-1}$ and 0.0047 ± 0.0016 $mmol.h^{-1-}$ respectively, while the syngas (50%) atmosphere exhibited the earliest peak productivity between 14-24 h at 0.0055 ± 0.0005 $mmol.h^{-1}$. Notably for the anaerobic atmospheres, H_2 production appeared to coincide with mixed acid fermentation, suggesting concurrent operation of these processes (Figure 3.1).

H_2 mol yields per CO mol consumed (H_2/CO) across the fermentations varied between the points where CO was completely consumed and the endpoints of the fermentations at 134 h. Each atmosphere, except for the syngas (100%) atmosphere showed complete CO consumption within 134 h. Unsurprisingly, due to its comparatively low initial CO levels, the syngas (50%) atmosphere demonstrated the earliest CO depletion (84 h), followed by the aerobic control (110 h), then the anaerobic control (120 h). By 134 h, 1.06 ± 0.63% CO remained in the syngas (100%) atmosphere gas headspace. At the points where CO was completely consumed, the aerobic and anaerobic controls and the syngas (50%) atmosphere exhibited significantly different H_2/CO yields from one another (Appendix B Supp. Table S3.2), of 89.48 ± 0.83%, 85.40 ± 0.14% and 93.62 ± 0.31% respectively. However, by the end point of the fermentations, yields fluctuated further, with both the aerobic control and syngas (50%) atmospheres having exhibited decreases of 4.92 ± 0.47% and 6.73 ± 0.85% respectively. Conversely, the anaerobic control

H_2/CO yield increased by 8.48 ± 0.43%. By 134 h, despite not having depleted CO, the H_2/CO yield of the syngas (100%) atmosphere stood at 83.83 ± 1.42%, similar to that of the aerobic control but significantly lower than those of the anaerobic control and syngas (50%) atmospheres (Appendix B Supp. Table S3.2).

Analysis of the metabolite profile with HPLC showed that lactate, formate and ethanol were, in descending order, the most prevalent metabolites produced under the anaerobic atmospheres, while only negligible amounts of formate and ethanol were observed for the aerobic atmosphere (Figure 3.1). Lactate, formate and ethanol levels across the anaerobic atmospheres stabilized at 9.3442 ± 0.2808 mmol, 7.6075 ± 0.4817 mmol and 5.2032 ± 1.0371 mmol after 38 h levels despite the syngas (100%) atmosphere having exhibited significantly lower productivity of each acid between 0-14 h (Appendix B Supp. Table S3.1).

Acetate was the central metabolite detected through HPLC in the aerobic atmosphere, rising rapidly and peaking at 12.1255 ± 0.2620 mmol by 38 h (Figure 3.1). Slight decreases in acetate levels were observed between 38-62 h to 11.0224 ± 0.1368 mmol but remained relatively stable thereafter. Acetate accumulation in the anaerobic atmospheres was significantly lower than the aerobic atmosphere (Appendix B Supp. Table S3.1), accumulating to only a third of that of the aerobic control. Acetate production was almost identical between the anaerobic control and syngas (50%) atmospheres, rising rapidly within the first 14 h. The syngas (100%) atmosphere, however, as observed for the lactate, formate and ethanol levels showed significantly lower production rates between 0-14 h and consequently significantly lower levels at 14 h post-inoculation compared to the anaerobic control and syngas (50%) atmospheres. Acetate levels in the anaerobic control and both syngas atmospheres remained relatively stable thereafter at 3.0566 ± 0.6325 mmol. In addition to acetate, low levels of succinate were detected in the aerobic fermentation. The aerobic control exhibited significant increases in succinate production between 38-72 h (Appendix B Supp. Table S3.1), reaching a final level of 2.5111 ± 0.0606 mmol, while the anaerobic atmospheres showed negligible succinate accumulation.

Glucose was completely consumed within 14 h post-inoculation in the aerobic control. Low levels of glucose, below the 0.2 $mg.L^{-1}$ HPLC method detection limit, were observed throughout all three anaerobic fermentations. However, most of the initially available glucose (>99%) was consumed by 24 h in the N_2 control and syngas (50%) atmospheres and by 38 h in the syngas (100%) atmosphere. Low levels of glucose, iso-butyrate, iso-valerate, below the 0.2 $mg.L^{-1}$ HPLC method detection limit, were observed throughout all three anaerobic fermentations. No notable fluctuations in butyrate, glyoxylate, propionate, 1-propanol or valerate were observed in any of the fermentations.

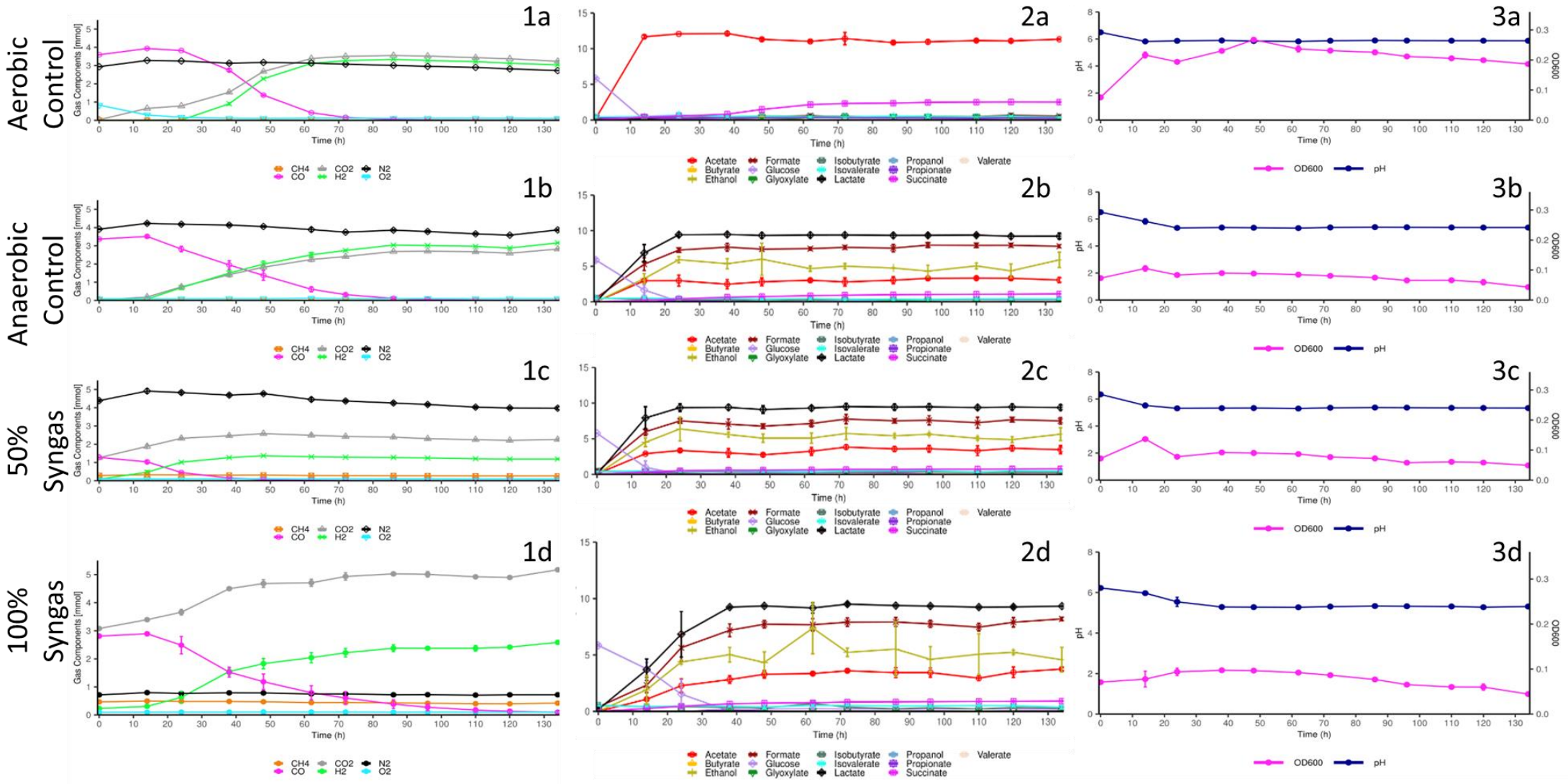


Figure 3.1: Gas composition (1), metabolite (2) and growth and pH (3) profiles for *P. thermoglucosidasius* DSM 6285 cultured under 50% CO/50% Air- (a) 50% CO/50% N_2- (b) 50% synthetic syngas- (c) and 100% synthetic syngas- (d) -gas headspaces.

3.3.2 Pre-hydrogenogenic phase O_2 perturbations extend the lag phase prior to H_2 production

Oxygen addition at the various timepoints resulted in distinct fermentation profiles. Oxygen additions at 14 h and 24 h (Figure 3.2) exhibited markedly different influences on the various fermentation profiles compared to the control and 38 h and 48 h O_2 additions (Figure 3.3).

Across controls, OD_{600} absorbance values increased sharply during the aerobic growth phase and peaked by 48 h at 0.267 ± 0.010 (Figure 3.2) and 0.218 ± 0.003 (Figure 3.3), coinciding with the onset of the WGS, diminishing thereafter. For all O_2 additions, significantly higher peak and final OD_{600} values were observed compared to the controls following O_2 addition (Appendix B Supp. Tables S3.3 and S3.4). In the cases of the 38 h and 48 h O_2 additions, which coincided with the hydrogenogenic phase, severe reductions in H_2 production were observed, implying the growth observed was primarily the result of O_2 addition and not supplemented through the WGS reaction. While the control OD_{600} values declined following peak levels at 48 h, the 38 h and 48 h O_2 additions exhibited further significantly elevated peak OD_{600} values of 0.332 ± 0.011 and 0.302 ± 0.006 at 48 h and 58 h respectively following oxygen addition (Appendix B Supp. Tables S3.3 and 4).

Steep declines in pH to lows of 5.78 ± 0.01 (Figure 3.2) and 5.93 ± 0.02 (Figure 3.3) within 14 h were observed across the fermentation batches in the aerobic phase, reflective of the observed acetate, formate and lactate accumulation. In the controls, these declines were followed by slight increases alongside near complete and complete consumption of formate and lactate, stabilizing at final pH values of 5.88 ± 0.02. Similar profiles were observed until either the addition of O_2 in the 14 h and 24 h O_2 addition replicates or 34-38 h in the 38 h and 48 h O_2 addition replicates. O_2 addition resulted in the significant increase in pH compared to the controls (Appendix B Supp. Tables S3.3 and S3.4), until the final sampling point, likely the result of decreases in observed acetate levels.

H_2 gas was first detected across all samples within 38 h. The highest O_2 level detected with concurrent H_2 production was 2.49 ± 0.06 % O_2. O_2 additions elevated

O_2 levels an average of 4.51 ± 0.35%. Peak H_2 productivities were experienced between 34-38 h at 0.0297 ± 0.0004 mmol.h^{-1} (Batch 1 control), 0.0205 ± 0.0045 mmol.h^{-1} (Batch 2 control) and 0.0130 ± 0.0052 mmol.h^{-1} (38 h O_2 addition). The 48 h O_2 addition replicates demonstrated peak H_2 productivity between 38-48 h at 0.0143 ± 0.0006 mmol.h^{-1}, whilst both the 14 h and 24 h O_2 additions showed more delayed peak productivities of 0.0133 ± 0.0004 mmol.h^{-1} and 0.0123 ± 0.0002 mmol.h^{-1} respectively, between 48-58 h. H_2 productivity decreased gradually as CO was depleted, eventually resulting in the net consumption of H_2, with a maximal H_2 consumption rate of 0.0030 ± 0.0002 mmol.h^{-1} across all replicate sets.

Significant increases in H_2 levels were observed in both controls between 24-72 h (Appendix B Supp. Table S3.5). In each instance where O_2 was added, significant disruptions to H_2 production were observed (Appendix B Supp. Tables S3.3 and S3.4). H_2 production was delayed following O_2 addition at 14 h and 24 h and resulted in significantly lower final H_2 levels (2.7351 ± 0.0058 mmol and 2.9094 ± 0.0127 mmol respectively) compared to the control (3.0260 ± 0.0112 mmol). H_2 productivity was significantly lower between both O_2 additions and the control between 24-38 h. Notably, the 24 h O_2 addition replicates showed earlier and more intensive H_2 production than the 14 h O_2 addition replicates, with comparably higher H_2 levels and H_2 productivities during the majority of the fermentation (Figure 3.2; Appendix B Supp. Tables S3.3 and S3.4). Following O_2 addition for the 38 h and 48 h O_2 addition replicate sets, both H_2 levels and H_2 productivities were significantly lower than the control for the subsequent 10 h period (Appendix B Supp. Tables S3.3 and S3.4). However, weak H_2 production was still observed across this period, with H_2 productivities of 0.0011 ± 0.0004 mmol.h^{-1}. Following this period, however, significant increases in H_2 levels were observed for the 38 h and 48 h O_2 addition replicates between 48-84 h and 58-84 h respectively (Appendix B Supp. Table S3.5). However, H_2 levels for both O_2 additions remained significantly lower than the control following O_2 addition.

Both controls and pre-hydrogenogenic phase 14 h and 24 h O_2 additions showed complete utilisation of CO, with trace amounts of CO remaining in the 38 h and 48 h O_2 addition replicates (Figure 3.2). H_2/CO yields at the points of complete CO depletion were 83.26 ± 1.19% (Controls), 74.35 ± 0.07% (14 h O_2 addition) and

79.10 ± 0.26% (24 h O_2 addition), respectively. These were all significantly different from one another (Appendix B Supp. Table S3.2). Yields of 80.35 ± 0.13% and 81.33 ± 0.30% were observed for the 38 h and 48 h O_2 addition replicates by 120 h. Both yields were significantly lower than the respective control but not significantly different from one another (Appendix B Supp. Table S3.2).

HPLC analysis detected fluctuations in acetate and succinate profiles following oxygen addition. Acetate levels across all samples increased dramatically from the onset of the fermentations and peaked either prior to O_2 addition or within 34 h post-inoculation (Figure 3.2). In the controls, after reaching peak acetate levels of 11.4257 ± 0.5833 mmol, a low-level net consumption of acetate was observed, gradually reducing acetate levels to a final concentration of 10.4132 ± 0.6191 mmol. In each case where O_2 was added, significantly lower acetate levels were observed compared to the respective controls, throughout the remainder of the fermentations (Appendix B Supp. Tables S3.3 and S3.4). Correspondingly, significantly higher acetate consumption levels were observed between all O_2 addition replicates and the controls immediately following O_2 addition. This was observed between 14-34 h (14 h O_2 addition), 24-38 h (24 h O_2 addition), 38-82 h (38 h O_2 addition) and 38-72 h (48 h O_2 addition), respectively post-O_2 addition.

Although succinate levels were observed to increase gradually from the start of all fermentations, prominent increases were only observed after 24 h in the controls, peaking at 2.544 ± 0.1405 mmol (Figure 3.2). O_2 addition at 14 h and 24 h delayed succinate accumulation, presenting significantly lower succinate levels than the control between 34-58 h and 38-58 h respectively (Appendix B Supp. Table S3.3). Despite this, the 14 h O_2 addition replicates accumulated significantly higher final succinate levels compared to both the control and 24 h O_2 addition replicates, peaking at 3.1261 ± 0.1344 mmol. Following O_2 addition at 38 h and 48 h, significantly lower levels of succinate were observed compared to the control between 48-58 h (Appendix B Supp. Table S3.4). Despite this delay in succinate production, there were no significant differences between the final accumulated succinate concentrations of the control, 38 h and 48 h O_2 additions, which averaged 2.7774 ± 0.9710 mmol.

Minor increases in propionate levels were observed across all conditions, peaking after 34 h (Figure 3.2) at an average value of 1.7308 ± 0.5558 mmol and reaching a final average value of ± 0.5558 mmol. However, there was no significant difference in propionate levels between conditions at any timepoint. Glucose levels across all samples were almost completely consumed (>97%) 14 h post-inoculation and were completely consumed within 34 h. Large formate 12.9833 ± 1.0624 mmol and lactate 6.5760 ± 0.5515 mmol peaks were detected at 14 h post-inoculation across each replicate set. Both compounds were almost completely consumed thereafter with only minor increases in formate below the detection limit (<0.2 g/L) of the HPLC method. Similarly, minor glyoxylate accumulation below the detection limit was noted across all sampling sets. Additionally consistent low levels (below the detection limit) of, iso-butyrate and iso-valerate were observed in each sampling set throughout the fermentation. No notable fluctuations in 1-propanol, butyrate, ethanol, iso-butyrate, iso-valerate, propionate, or valerate were observed.

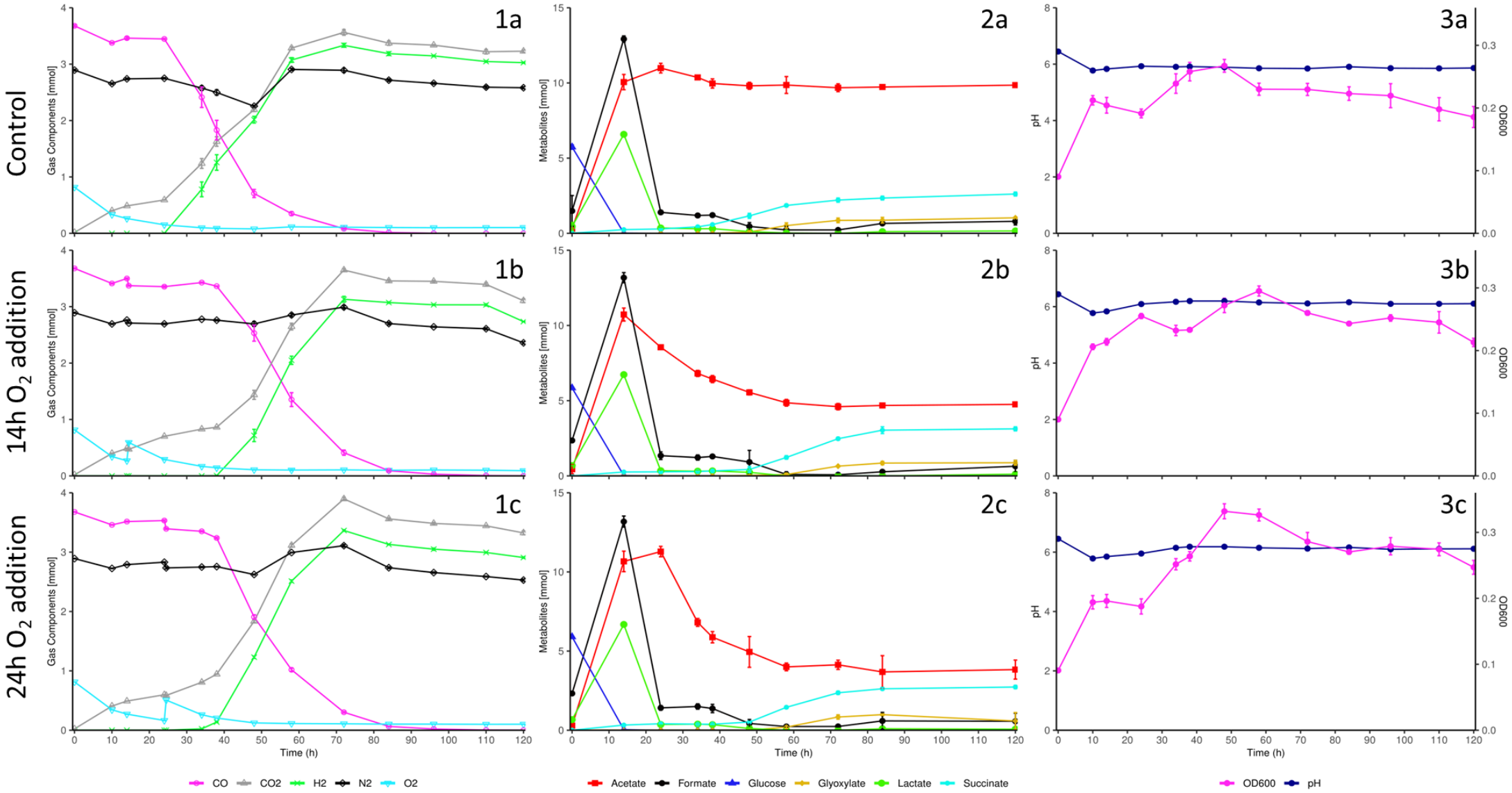


Figure 3.2: Gas composition (1) metabolite (2) and growth and pH (3) profiles for *P. thermoglucosidasius* DSM 6285 cultured under a 50% CO/50% Air gas headspace with no O_2 addition (a), 14 h O_2 addition (b) and 24 h O_2 addition (c).

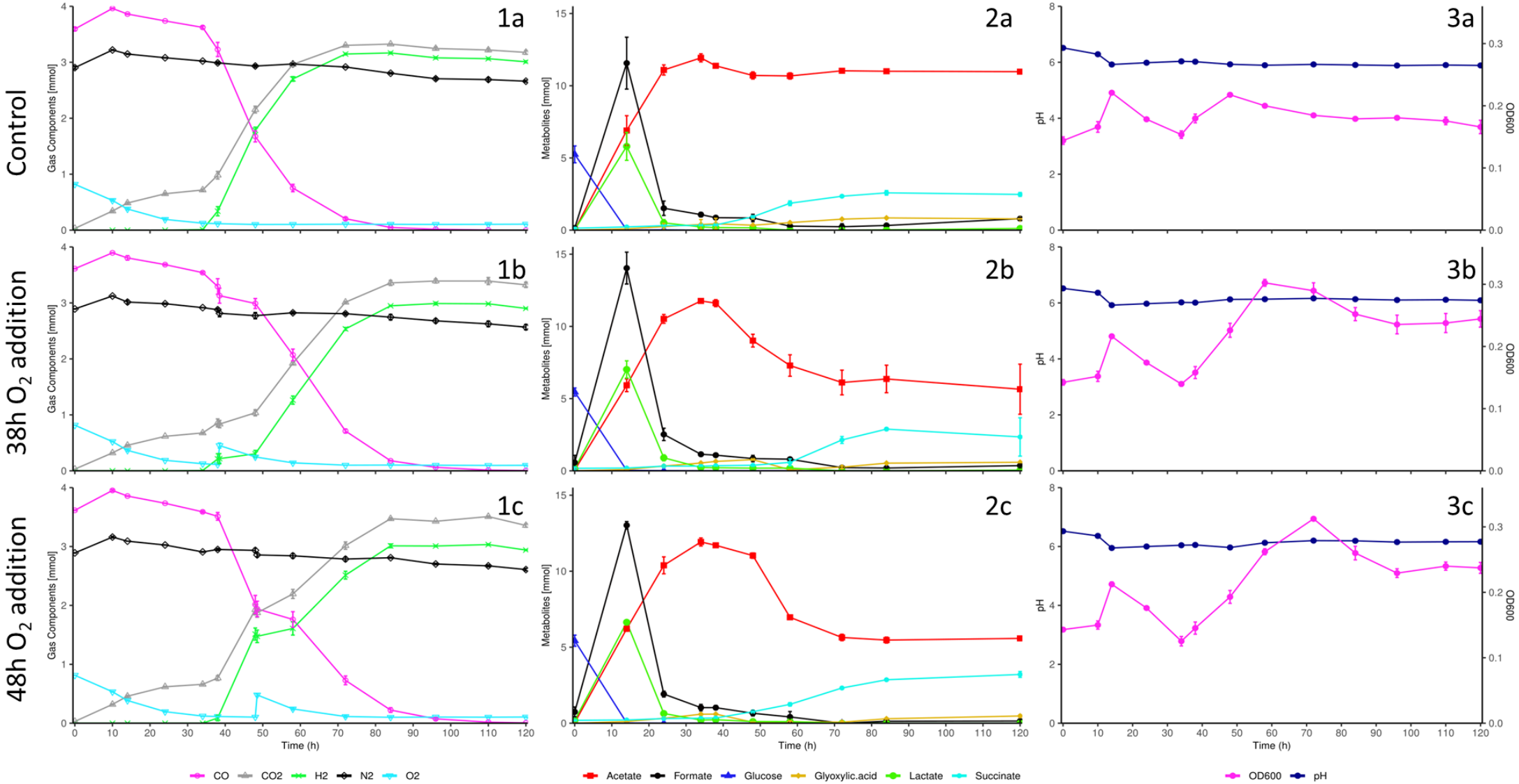


Figure 3.3: Gas composition (1), metabolite (2) and growth and pH (3) profiles for *P. thermoglucosidasius* DSM 6285 cultured under a 50% CO/50% Air gas headspace with no O_2 addition (a), 38 h O_2 addition (b) and 48 h O_2 addition (c).

3.3.3 Minor peak-hydrogenogenic phase O_2 perturbation minimally disrupts H_2 production in *P. thermoglucosidasius* DSM 6285

Prior to oxygen addition, the 38 h varied O_2 volume addition fermentations demonstrated the same trends in growth, pH, gas composition and HPLC metabolite profiles as observed previously (Figure 3.2; Figure 3.3). Following varied oxygen additions, however, disruptions in these profiles were scaled accordingly with the volumes of oxygen supplemented (Figure 3.4).

Following O_2 addition OD_{600} values increased in accordance with volumes of O_2 added, with larger O_2 volume additions resulting in greater peak OD_{600} values achieved (Figure 3.4). Expectedly, the 1% O_2 addition prompted the weakest response, exhibiting significantly higher values compared to the control from 48 h onwards (Appendix B Supp. Table S3.5). The 3% O_2 addition showed a slightly more delayed increase in OD_{600} values, with significantly elevated OD_{600} values compared to the control from 62 h onwards. The 5% and 10% O_2 additions both exhibited a slight decrease in OD_{600} values following O_2 addition between 38-41 h, however, showed a steep increase thereafter. From 44 h onwards, the 5% O_2 addition maintained significantly higher OD_{600} values compared to the control. The 5% O_2 addition also maintained significantly higher OD_{600} values than those of the 1% and 3% O_2 additions over the majority of the fermentation and was similar to the 10% O_2 addition. Although the 10% O_2 addition exhibited the greatest increase in OD_{600} values, peaking at 0.395±0.006 at 72 h, OD_{600} values appeared to decrease sooner than those of the 5% O_2 addition.

Similarly, larger O_2 volume additions resulted in the greater pH increments, likely the result of the observed reduction in acetate levels following O_2 addition (Figure 3.4). Apart from the 1% O_2 addition, which displayed no significant difference from the control, all other O_2 additions prompted significant increases in pH values following 6 h post-O_2 addition (Appendix B Supp. Table S3.5). Like the control, 1% O_2 addition pH values were significantly lower than those of the other O_2 addition conditions following O_2 addition. Following O_2 addition, pH values were lower for the 3% O_2 addition than 5% or 10% O_2 additions. The 5% O_2 addition also resulted in lower pH values than 10% O_2 addition. pH values stabilized in each

case after ~62 h, alongside plateaus in acetate consumption and OD_{600} peaks, with the 10% O_2 addition almost returning to starting pH values, reaching as high as 6.35 ± 0 (Figure 3.4).

H_2 was first detected at 38 h across all replicate sets at elevated levels (Figure 3.4). O_2 additions resulted in O_2 levels being raised by 0.76 ± 0.15% (1% O_2 addition), 2.67 ± 0.02% (3% O_2 addition), 4.44 ± 0.13% (5% O_2 addition) and 8.77 ± 0.10% (10% O_2 addition), respectively. O_2 levels post-O_2 addition were 0.90 ± 0.22% (1% O_2 addition), 2.83 ± 0.07% (3% O_2 addition), 4.70 ± 0.14% (5% O_2 addition) and 9.08 ± 0.14% (10% O_2 addition), respectively. The control exhibited a H_2 productivity of 0.0460 ± 0.0013 $mmol.h^{-1}$ between 38-41 h. Comparatively, H_2 productivities of the O_2 addition replicates were all significantly lower than the control in this period (1%: 0.0137 ± 0.0062 $mmol.h^{-1}$; 3%: 0.0002 ± 0.0003 $mmol.h^{-1}$; 5%: 0.0005 ± 0.0012 $mmol.h^{-1}$; 10%: 0.0009 ± 0.0003 $mmol.h^{-1}$; Appendix B Supp. Table S3.5).

In the case of the 1% O_2 addition, no significant difference in H_2 productivity between the control was detected following 41 h (Appendix B Supp. Table S3.6), with H_2 productivity between 38-48 h (0.0101 ± 0.0013 $mmol.h^{-1}$), the peak H_2 productivity period, only slightly lower than that of the control (0.0150 ± 0.0006 $mmol.h^{-1}$). However, unlike the control, no significant increase in H_2 levels were observed between 38-41 h (Appendix B Supp. Table S3.5). Significant increases in H_2 levels were only detected in the 1% O_2 addition replicates between 41-62 h (Appendix B Supp. Table S3.5), when O_2 levels were below 0.79 ± 0.13%. Despite this, no significant difference in H_2 concentration was detected between the 1% O_2 addition and the control over this period (Appendix B Appendix B Supp. Table S3.6). However, following complete CO consumption, the 1% O_2 addition replicates exhibited significantly lower final H_2 levels (2.8999 ± 0.0788 mmol) than the control (3.1277 ± 0.0874 mmol; Appendix B Supp. Table S3.6).

Following the reduction in H_2 productivity experienced by the 3%, 5% and 10% O_2 addition replicates between 38-41 h, H_2 consumption was observed for each of these O_2 additions between 41-44 h at rates of 0.0002 ± 0.0005 $mmol.h^{-1}$, 0.0019 ± 0.0002 $mmol.h^{-1}$ and 0.0018 ± 0.0005 $mmol.h^{-1}$ respectively (Figure 3.4). H_2 productivity

for these replicates remained significantly lower than the control between 38-48 h for 3% and 5% O_2 addition replicates and 38-62 h for the 10% O_2 addition replicates (Appendix B Supp. Table S3.6). However, H_2 productivity increased following 44 h with the 10% O_2 addition replicates demonstrating significantly higher H_2 productivity than the control between 62-96 h (Appendix B Supp. Table S3.6). Consequently, significant increases in H_2 levels were only detected between 48-62 h for the 3% and 5% O_2 addition replicates and between 62-72 h for the 10% O_2 addition replicates (Appendix B Supp. Table S3.5). Following O_2 addition, O_2 levels at the first point where a significant increase in H_2 was observed (Appendix B Supp. Table S3.6), were 0.03 ± 0.06% (3% O_2 addition), 0.17 ± 0.09% (5% O_2 addition) and 0.97 ± 0.1% (10% O_2 addition). After O_2 addition, significantly lower levels of H_2 were detected between the 3%, 5% and 10% O_2 addition replicates and the control till the endpoint of the fermentations (Appendix B Supp. Table S3.6). Lower final H_2 levels were observed accordingly with higher volumes of O_2 added (Control: 3.0525 ± 0.0110 mmol, 1% O_2 addition: 2.8290 ± 0.0097 mmol, 3% O_2 addition: 2.8038 ± 0.0084 mmol, 5% O_2 addition: 2.7613 ± 0.0085 mmol, 10% O_2 addition: 2.7657 ± 0.0096 mmol). However, no significant differences in final H_2 levels were observed between the 1% and 3%- and between the 5% and 10%-O_2 addition replicates (Appendix B Supp. Table S3.6).

CO was completely consumed across each condition by 134 h (Figure 3.4). As could be expected, the control exhibited CO depletion first (110 h) alongside the 1% O_2 addition replicates, then the 3% O_2 addition replicates (120 h) and the 5% and 10% O_2 addition replicates (134 h). The control exhibited the highest H_2/CO yield both at the first point of CO depletion (88.56 ± 0.62%) and by the end of the fermentation (83.86 ± 0.62%). In both instances yields were significantly higher than all O_2 additions yields (Appendix B Supp. Table S3.2), despite exhibiting a decrease in yield following CO depletion. Although producing significantly different yields at the first points of CO depletion (82.26 ± 0.25% and 85.26 ± 0.30% respectively), the 1% and 3% O_2 additions showed no significant difference in H_2/CO yield by the endpoint of the fermentation (Appendix B Supp. Table S3.2), with final yields of 77.82 ± 0.00%. In both cases, however, reductions in H_2/CO yields of 4.70 ± 0.41% and 8.30 ± 0.94%, respectively, were observed between the two points. The 5%

followed by 10% O_2 additions exhibited the lowest H_2/CO yields at 134 h, with yields of 76.51 ± 0.20% and 75.89 ± 0.24%, respectively.

As observed in the previous experiment, acetate levels across all samples increased dramatically from the onset of the fermentations and peaked prior to O_2 addition (Figure 3.4). Peak acetate levels of 12.5025 ± 0.1713 mmol were observed by 24 h, followed by net reduction, with the control reaching a final acetate concentration of 11.1013 ± 0.6522 mmol. The 1% O_2 addition showed almost no differences in acetate levels compared to the control (Appendix B Supp. Table S3.6). In the cases of the 3%, 5% and 10% O_2 additions, significantly lower acetate levels were observed after 48 h, 44 h and 48 h, respectively, with final acetate levels of 8.5019 ± 0.5227 mmol (3% O_2 addition), 5.3856 ± 0.3675 mmol (5% O_2 addition) and 2.2098 ± 1.0733 mmol (10% O_2 addition). Greater degrees of acetate consumption were observed in accordance with greater volumes of O_2 added, with the most noteworthy decreases in acetate observed by the 10% O_2 addition, followed by the 5%, 3% and finally 1% O_2 additions. Correspondingly, significantly higher acetate consumption levels were observed between the 3%, 5% and 10% O_2 addition replicates and the control (Appendix B Supp. Table S3.6). No significant differences in acetate productivity were observed between the 1% O_2 addition and the control.

Succinate levels increased gradually from the start of all fermentations (Figure 3.4), although prominent increases in succinate were staggered. Whilst there was no significant difference in succinate productivity immediately following O_2 addition between the 1%, 3% and 5% O_2 additions and the control, the 10% O_2 addition exhibited net succinate consumption following O_2 addition, between 38-48 h and significantly lower succinate productivity until 62 h post-inoculation (Appendix B Supp. Table S3.6). All O_2 additions exhibited significantly higher succinate productivity compared to the control between 62-84 h. Compared to the control which exhibited final succinate levels of 2.7534 ± 0.1376 mmol, the 5% O_2 addition replicates accrued significantly higher levels of 3.2737 ± 0.2034 mmol by the end of the fermentation.

Glucose levels across all samples were almost completely consumed (>97%) within 14 h and were completely consumed within 34 h. Formate and lactate peaks were detected at 14 h post-inoculation across each replicate set, however, both compounds were almost completely consumed thereafter with only minor increases in formate below the detection limit (<0.2 g/L) of the HPLC method. Minor glyoxylate accumulation below the detection limit was noted across all conditions. Additionally consistent low levels (below the detection limit) of propionate, iso-butyrate, iso-valerate were observed in each condition throughout the fermentation. No notable fluctuations in 1-propanol, butyrate, ethanol, iso-butyrate, iso-valerate, propionate, or valerate were observed.

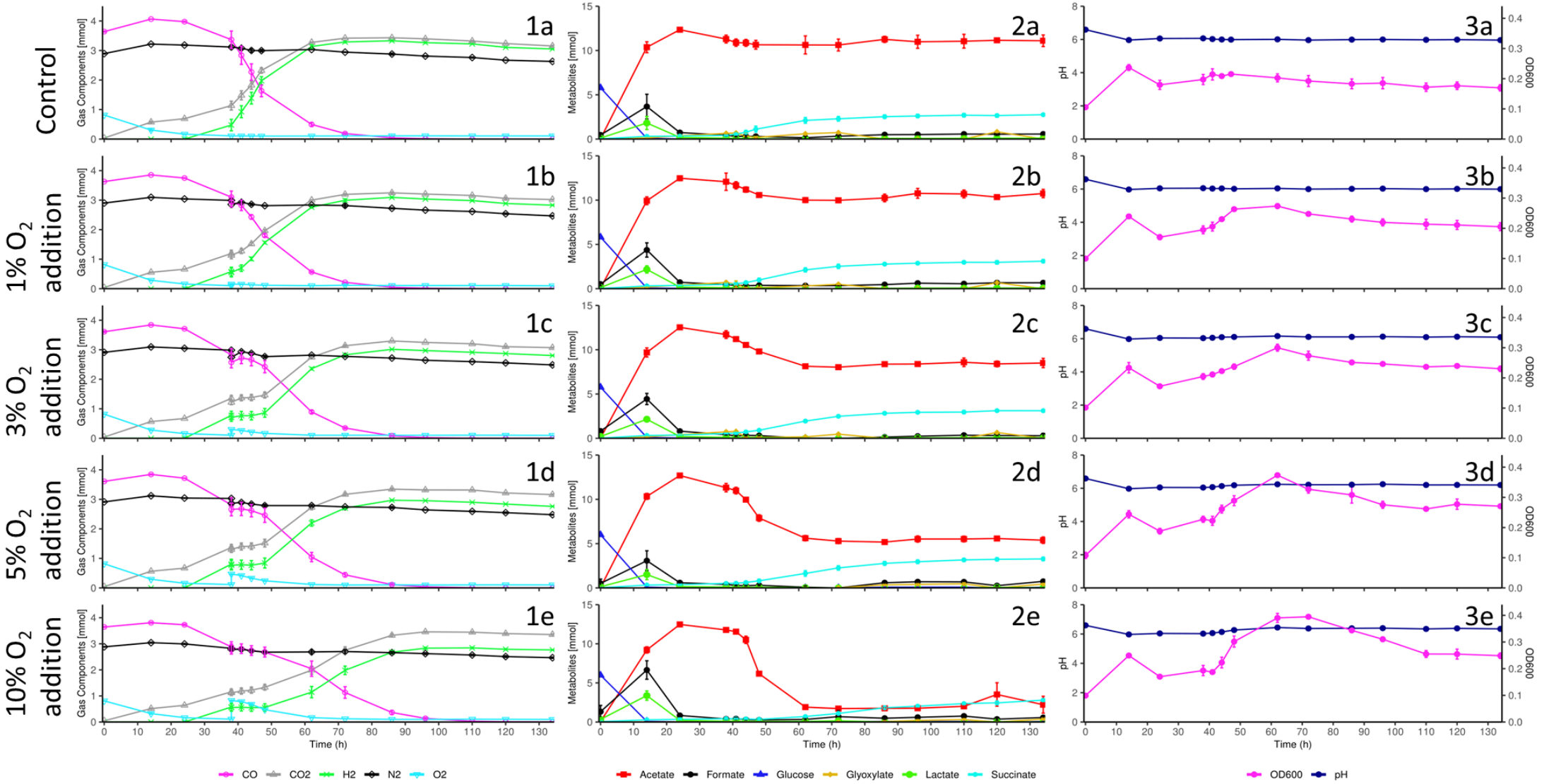


Figure 3.4: Gas composition (1) metabolite (2) and growth and pH (3) profiles for *P. thermoglucosidasius* DSM 6285 cultured under a 50% CO/50% Air gas headspace with no O_2 addition (a), 1%- (b), 3%- (c), 5%- (d) and 10%- (e) -O_2 addition at 38 h.

3.4 Discussion

The capacity to metabolise CO containing gases by several *P. thermoglucosidasius* strains via the WGS reaction has been characterised using a range of both aerobic and anaerobic gas atmospheres (Mohr *et al.*, 2018a; Mohr *et al.*, 2019b; Aliyu *et al.*, 2020; Aliyu *et al.*, 2021; Imaura *et al.*, 2023). These studies have tended towards the use of strictly/relatively pure, well defined CO containing gases, generally including an initial oxygen component to supplement aerobic growth, that have facilitated the characterisation of the *P. thermoglucosidasius* WGS reaction in a pragmatic and systematic manner. A crux of the implementation of this system, however, resides in its advancement towards utilizing crude or minimally refined gas feedstocks. In particular, industrial waste gases such as synthesis gas (syngas), are an appealing option. The composition of syngases however, are highly variable, dependant on their initial feedstock and production process (Chiche *et al.*, 2013; Ramachandriya *et al.*, 2016). Primary constituents of syngases include CO, H_2 and CO_2, alongside impurities including tars, particulate carbon, light hydrocarbons, inorganic compounds (sulfur-, nitrogen-, halogen- and metal-incorporated compounds) and O_2 (Abubackar *et al.*, 2019). In an attempt to emulate the use of less refined gas feedstocks, here we used a synthetic syngas mixture in both an undiluted and diluted state and explored the impacts of this on the *P. thermoglucosidasius* WGS reaction. Our analyses showed that *P. thermoglucosidasius* DSM 6285 cultivated in an anaerobic synthetic syngas atmosphere was able to produce hydrogen gas, despite exposure to highly variable CO, CO_2 and N_2 levels and the presence of impurities such as methane, ethane, propane and butane. Cultivation under anaerobic syngas-containing conditions, however, did come at the cost of biomass accumulation.

P. thermoglucosidasius has undergone extensive study exploring ideal gas compositions to facilitate H_2 production through the water-gas shift reaction. This has generally included an initial oxygen component to supplement aerobic growth, as it has been established that oxygen is necessary for the growth of *P. thermoglucosidasius* (Hussein *et al.*, 2015; Mohr *et al.*, 2018a; Mohr *et al.*, 2019b; Aliyu *et al.*, 2021; Mol *et al.*, 2021). Initially anaerobic atmospheres have also been explored in the presence and absence of supplemented alternative electron acceptors, however, no significant implications to growth were observed (Imaura *et al.*, 2023). It was evident here, by comparing the growth profiles of the aerobic control to those of the anaerobic control and two anaerobic syngas atmospheres that growth was severely hampered under

anaerobic conditions. This could be expected and accounted for by the radically different energetics associated with aerobic and anaerobic metabolism respectively.

Metabolite profiles aligned in two consistent groups governed by the aerobic or anaerobic status of the gas headspace. Although there was minor variability between the anaerobic control and syngas atmospheres, highly similar trends were observed between them that were distinct from the aerobic control. The aerobic control exhibited pronounced acetate production prior to commencement of H_2 production, alongside concerted succinate production which coincided with O_2 depletion and onset of the WGS reaction, in concert with similar previous studies (Mohr *et al.*, 2018a; Mohr *et al.*, 2019b; Aliyu *et al.*, 2021). Predictably, anaerobic atmospheres exhibited distinct mixed acids fermentation profiles favouring lactate, formate, ethanol and acetate production, which could account for the more marked reductions in pH observed across these conditions compared to the aerobic control.

Lactate, formate and ethanol were produced in abundance in the anaerobic atmospheres but scarcely or not detected at all in the aerobic control. Lactate is typically produced by *P. thermoglucosidasius* under anaerobic conditions through conversion of pyruvate by lactate-dehydrogenase, as a product of mixed acids fermentation (Tang *et al.*, 2008; Cripps *et al.*, 2009). Lactate production is commonplace among the sister genera of *Geobacillus* and *Parageobacillus* and has been explored to some extent both towards the reduction and enhancement of lactate production under variable culturing conditions (Cripps *et al.*, 2009; Najar and Thakur, 2020; Liu *et al.*, 2023). As expected, lactate was a central fermentation end product produced during mixed acids fermentation in the anaerobic conditions. Formate featured as the second most abundant fermentation product detected across the anaerobic conditions, whereas it was only minimally detected in the aerobic control. Formate production in *P. thermoglucosidasius* is carried out through the conversion of pyruvate to acetyl-coA and formate by pyruvate formate lyase under anaerobic or microaerobic conditions (Tang *et al.*, 2008; Cripps *et al.*, 2009). Previous aerobic *P. thermoglucosidasius* cultivations have similarly observed minimal formate production, sparing occasional ephemeral detection as an overflow metabolite (Aliyu *et al.*, 2021). Similarly, elevated ethanol levels were observed across the anaerobic conditions. Ethanol production in *P. thermoglucosidasius* proceeds through acetyl-CoA reduction to ethanol. This process competes with acetate production, relying on the pool of available acetyl-CoA and consumes reducing equivalents with no net energy generation (Tang *et al.*, 2008; Cripps *et al.*, 2009). This is another well explored fermentative process of *Parageobacillus*, with several studies having investigated the enhancement of the

ethanologenic capacities of the genus (Tang *et al.*, 2008; Cripps *et al.*, 2009; van Zyl *et al.*, 2014; Zhou *et al.*, 2016). Gradual accumulation of succinate in the aerobic control was congruent with previous studies, and attributed to the switch from the oxidative to the reductive branch of the TCA cycle where the glyoxylate shunt is observed (Figure 3.5), a result of the switch from aerobic to anaerobic conditions, (Aliyu *et al.*, 2021). The same degree of succinate accumulation observed in the aerobic conditions was understandably not observed in the anaerobic conditions which featured predominantly fermentative metabolism.

Under an aerobic atmosphere, *P. thermoglucosidasius* DSM 6285 exhibited particularly strong acetogenic activity, with acetate comprising the largest component of the metabolites detected. Acetate is produced through the activity of a two-step conversion in *P. thermoglucosidasius*, which reversibly converts acetyl-coA into acetyl-p and then acetate by acetyl phosphotransferase and acetate kinase respectively, although requiring ATP investment for the reverse reaction (Wolfe, 2005; Pinhal *et al.*, 2019). Acetyl-coA, the precursor to acetate, is produced through several pathways in *P. thermoglucosidasius*. Perhaps the most pertinent, considering the metabolites observed in this study include aerobic/microaerobic conversion of pyruvate to acetyl-coA by pyruvate dehydrogenase and microaerobic/anaerobic pyruvate conversion to formate and acetyl-coA by pyruvate formate lyase (Sawers and Böck, 1988; Wolfe, 2005; Tang *et al.*, 2008; Cripps *et al.*, 2009; Mol *et al.*, 2021). Acetate accumulation is also known to have inhibitory effects on cell growth, potentially causing cell damage (Russell, 1992). In the aerobic controls, minor reductions in acetate levels were observed following progression towards the anaerobic phase. This has previously been correlated with the expression of acetyl-CoA synthetase during the microaerobic phase, which catalyses the conversion of acetate into acetyl-CoA in *P. thermoglucosidasius* and may facilitate integration of acetate into the central carbon metabolism through energy supplementation derived from the WGS reaction (Aliyu *et al.*, 2020; Aliyu *et al.*, 2021). As well as serving as a means for balancing cell redox levels and potentially as an alternative carbon pool, it has been suggested that elevated acetate levels increase the CO uptake capacity of CO consumers (Allaart *et al.*, 2023). Notably, two-fold greater H_2 production rates were observed under an aerobic atmosphere, where acetate levels were also several fold higher than anaerobic/microaerobic atmospheres. Direct attribution of this only to elevated acetate levels would be misguided, given the variations in growth and metabolic state of the various conditions.

It appeared as though varying CO levels, especially under the three anaerobic conditions influenced growth and metabolism to a large extent, impacting downstream H_2 production.

Notably, the syngas (50%) atmosphere which comprised the lowest CO level exhibited a pronounced earlier onset of fermentative metabolism and the WGS reaction, most likely the result of the more pronounced biomass accumulation observed. The syngas (100%) atmosphere, however, comprising the next highest CO level showed the most delayed metabolic response and weakest growth response. From previous aerobic studies it would have been expected that atmospheres with enriched CO levels would have facilitated a more prolific growth response and correspondingly greater hydrogenogenic activity (Mohr *et al.*, 2019; Aliyu *et al.*, 2021). However, under the strictly anaerobic atmospheres, responses to enriched CO levels were variable.

In contrast to previous studies which observed faster glucose depletion with elevated CO levels (Aliyu *et al.*, 2021), the syngas (50%) atmosphere comprising the lowest CO component (~17%) exhibited the most rapid glucose utilisation among the anaerobic atmospheres. Additionally, the syngas (50%) atmosphere also exhibited the highest levels of lactate, formate, ethanol, succinate and H_2 production across all conditions in the early stage of the fermentation and the greatest degree of biomass accumulation across the anaerobic atmospheres prior to CO limitation. It is likely that the low CO atmosphere facilitated accumulation of more catalytically active biomass, enabling higher gross metabolic activity, enhancing glucose consumption and mixed acids production rates. Across the different anaerobic atmospheres, despite exhibiting pronounced H_2 evolution and mixed acid fermentation end-products, albeit slightly delayed, the syngas (100%) atmosphere exhibited lower and delayed fluctuations in biomass accumulation. This may in part be the influence of the granular growth exhibited by the cultures which may have led to inconsistent OD_{600} estimation (McBirney *et al.*, 2016). However, slower glucose utilisation and fermentation product accumulation by this atmosphere supported the notion that biomass accumulation was indeed hampered. This may be attributed to the influence of the gas headspace composition and most likely the CO_2 or CO levels.

CO_2 levels have previously been reported to have variable influences on aerobic taxa, although generally hinder growth (Daniels *et al.*, 1985). The influence of elevated CO_2 levels has not previously been described to influence the growth of *Parageobacillus* spp., however, it is not unlikely that it may have contributed to the reduced growth patterns observed. CO_2 levels were varied between the three anaerobic atmospheres, although the insubstantial initial CO_2 levels present in the anaerobic control which like the syngas (100%) atmosphere exhibited poor biomass accumulation, suggested it was not the principle contributing factor. Initial CO levels of the syngas (50%) atmosphere were considerably lower than both the anaerobic control and

the syngas (100%) atmospheres. Previous anaerobic cultivations of *P. thermoglucosidasius* DSM 6285 under various CO containing gas compositions have observed similar reductions in biomass accumulation with increasing CO levels, with 50% CO showing optimal biomass accumulation (Díaz *et al.*, 2022). However, this study did not consider lower CO compositions between 0-50%. Here, the syngas (50%) atmosphere only constituted approximately 17% CO, much lower than the optima described previously (Mohr *et al.*, 2019). CO is well established to be toxic to a range of enzymes, preferentially binding to catalytic metal clusters over native substrates, potentially resulting in complete reduction of the ferredoxin pool or inactivation of key metalloproteins that utilise low valence metals such as hydrogenases and hemeproteins (Menon and Ragsdale, 1996; Ragsdale, 2004; Allaart *et al.*, 2023). It may be possible that increasing CO levels exhibit increasing degrees of toxicity to key anabolic anaerobic pathways, reducing the capacity for biomass accumulation. It has already been established that CO induces certain metabolic changes in *P. thermoglucosidasius* by altering cell redox balance (Aliyu *et al.*, 2021). However, the holistic extent of the metabolic and cell changes have not been fully characterised and represent a further potential field of study.

Anaerobic culturing of *P. thermoglucosidasius* exhibited more rapid onset of H_2 production at the expense of biomass accumulation, although with H_2 production appearing limited only by CO availability. H_2 production in the anaerobic control and both syngas atmospheres commenced within the first 14 h of the fermentations, showing a considerable shortening of the lag phase between inoculation and onset of the WGS reaction. Previous anaerobic cultivations have exhibited reductions in this lag period to within 4 h post-inoculation, compared to ~24 h under an aerobic atmosphere, both timeframes similar to what was observed in this study (Imaura *et al.*, 2023). Notably, the syngas (50%) showed greater H_2 production rates in the early stages of the fermentation, compared to the other anaerobic atmospheres. As both the anaerobic control and the syngas (100%) atmosphere exhibited similar H_2 evolution profiles, despite having largely variable CO_2 and N_2 proportions and distinct but not considerably variable levels of gas impurities (H_2, CH_4, C_2H_6, C_2H_8, C_4H_{10}), it was most likely that the considerably higher CO levels in these atmospheres were responsible for their delays in concerted H_2 evolution. Gas mixtures containing 25% CO: 75% air and 36% CO: 64% air have previously been compared to greater CO comprising gas mixtures and deemed less suitable than 1:1 ratios, however, these ratios were evaluated under aerobic atmospheres and may not be reflective of optimal conditions for anaerobic operation (Mohr *et al.*, 2019). Furthermore, the syngas (50%) showed elevated OD_{600} absorbance values in this period

compared to the other anaerobic atmospheres. If indeed CO levels were responsible for these observations, it is likely that lower levels of CO in the syngas (50%) atmosphere facilitated greater biomass accumulation compared to the anaerobic control and syngas (100%) atmosphere and consequently more rapid H_2 evolution, in contrast to what is observed under aerobic operation.

Previous fermentations using *P. thermoglucosidasius* DSM 6285 have attained a 1:1 ratio of H_2 produced/CO consumed under initial aerobic cultivations with a 50% CO: 50% air atmosphere (Mohr *et al.*, 2018a). Here we observed distinct conversion efficiencies between the first noted points of complete CO consumption and the end-points of the fermentations. At the points of CO depletion, the aerobic control exhibited a slightly higher conversion yield than the anaerobic atmospheres. However, the syngas (50%) atmosphere exhibited more efficient conversion rates. In both instances for the aerobic control and the syngas (50%) atmosphere, reductions in yield between the points of CO consumption and the end points of the fermentation were observed. Yields by the end of the fermentation were similar across these conditions and the syngas (100%) atmosphere. Given the absence of known CO producing pathways, decreases in conversion efficiency were most likely the result of H_2 consumption by one or both of the two putative group 1d and 2a [NiFe]-uptake hydrogenases encoded by *P. thermoglucosidasius* DSM 6285 (Discussed in Chapter two; Mohr *et al.*, 2018a). It is likely that one or both of these hydrogenases were involved in coupling H_2 oxidation with an aerobic or anaerobic electron transport chain or respiratory reduction processes, and may account for the observed H_2 level reductions, although this requires validation (Greening *et al.*, 2016; Søndergaard *et al.*, 2016; Mohr *et al.*, 2018a; Cordero *et al.*, 2019). Conversely, the anaerobic N_2 control exhibited a large increase (8.48 ± 0.14%) in this ratio following complete CO consumption, it remains unclear as to why this was observed.

Evidently there were stark metabolic responses to aerobic and anaerobic CO fermentation by *P. thermoglucosidasius* DSM 6285. While anaerobic conditions may facilitate an accelerated H_2 productivity phase in cases like the syngas (50%) atmosphere, it also greatly reduced the potential for biomass accumulation, inadvertently reducing H_2 producing potential as observed for the syngas (100%) atmosphere. It can be reasoned that biomass accumulation and consequently O_2 availability are important factors to consider for enhancing this system. To develop this system towards more continuous operation, it would be beneficial to balance maximally anaerobic conditions to benefit minimal disruption or delay of WGS reaction onset, but maintain supply of O_2 to sustain elevated levels of catalytically active biomass. As such we

further explored the impacts of O_2 perturbation of the *P. thermoglucosidasius* DSM 6285 WGS reaction system at different key phases. Furthermore, we explored the impacts of different scale O_2 disruptions to the WGS reaction system during peak H_2 gas production.

O_2 addition across the different phases of the *P. thermoglucosidasius* DSM 6285 fermentation exhibited distinct changes in fermentation profiles and pronounced effects on H_2 production, with O_2 addition negatively impacting H_2 production at each stage. O_2 addition prior to commencement of H_2 production, during the aerobic growth phase delayed H_2 production to a greater extent than O_2 addition during the anaerobic phase. The extension of the aerobic phase may have facilitated extended respiratory metabolism, the more energetically efficient route in heterotrophic metabolism than anaerobic respiration or fermentation (Unden, 1998). This extended the lag phase prior to WGS reaction commitment to a greater extent, than the later anaerobic phase O_2 addition. At the point of O_2 addition in the anaerobic phase, cells would have progressed further towards fermentative metabolism with a more limited pool of reducing agents in higher oxidation states (Vignais and Billoud, 2007). This may have prompted the more rapid progression and commitment to the WGS as a means of conserving energy, as observed for the WGS system in *Rubrivivax gelatinosus* (Maness *et al.*, 2005; Mohr *et al.*, 2018a). This photosynthetic purple non-sulfur bacterium harbours a group 4c [NiFe]-hydrogenase, similar to the group 4a [NiFe]-hydrogenase present in *P. thermoglucosidasius* DSM 6285, and is responsible for coupling CO and H_2 metabolism with energy conservation through establishment of an ATP-yielding transmembrane proton gradient (Maness *et al.*, 2005). Similarly, other group 4 [NiFe]-hydrogenase harbouring taxa have been observed to conserve energy through proton or sodium ion membrane translocation coupled processes (Welte *et al.*, 2010; Pisa *et al.*, 2007; Schoelmerich and Müller, 2019). However, no evidence has yet been provided to declaratively state a similar process in *P. thermoglucosidasius*.

Later O_2 additions in the hydrogenogenic phase showed stark reductions in H_2 production alongside strong and extended gains in biomass accumulation. This could be indicative of the redirection of the *P. thermoglucosidasius* metabolism towards the use of O_2, which offers a higher oxidation state and the potential for greater energy yields. Greater volume additions of O_2 saw more pronounced impacts on H_2 production. Following O_2 disruption H_2 production resumed, albeit at much lower rates even after complete O_2 depletion, likely the combined result of the extended effects of O_2 induced inhibition of the WGS components and the shift towards maximizing use of the oxidative respiratory chain (Fernandez *et al.*, 1985; Vignais and Billoud, 2007; Merrouch *et al.*, 2015; Kim *et al.*, 2022).

The Ni-CODH and group 4a [NiFe]-hydrogenase responsible for conducting the *P. thermoglucosidasius* DSM 6285 WGS are both highly O_2 sensitive (Vignais and Billoud, 2007; Merrouch *et al.*, 2015; Greening *et al.*, 2016; Mohr *et al.*, 2018a). However, neither enzyme from *P. thermoglucosidasius* DSM 6825 have been fully characterised, so comparisons are often drawn between homologous enzymes harboured by other taxa (Soboh *et al.*, 2002; Finney and Sargent, 2019). The mechanism and occurrence of [NiFe]-hydrogenases O_2 sensitivity has received the bulk of investigations, although group 4a hydrogenases, also termed formate hydrogenlyases, are less well characterised, often drawing from studies of the two putative formate hydrogenlyases present in *E. coli*, as well as other better characterised group 4 [NiFe]-hydrogenases (Finney and Sargent, 2019; Lu and Koo, 2019).

O_2 tolerances of group 4 hydrogenases vary widely from highly sensitive to oxygen tolerant as observed for the group 4c hydrogenase from *Rubrivivax gelatinosus*, capable of maintaining activity under atmospheric O_2 levels (Maness *et al.*, 2002; Finney and Sargent, 2019; Shomura, 2019). The nature of the tolerance and recoverability of the *P. thermoglucosidasius* WGS system to oxygen disruptions have not been investigated. Here we observed the initial commencement of whole-cell WGS activity in the presence of 2.49 ± 0.06% O_2 in the gas headspace and resumption of H_2 production following O_2 disruption, at increasingly reduced rates in accordance with increasing volumes of O_2 addition. Significant increases in H_2 post-O_2 addition were only observed following sufficient reduction in O_2 levels, with significant increases in H_2 observed concurrently with a maximum O_2 level of 0.97 ± 0.1% or 0.665 ± 0.0081 mmol. [NiFe]-hydrogenases, unlike their counterpart FeFe-hydrogenases, generally display reversible inactivation upon O_2 exposure (Stiebritz and Reiher, 2012). O_2-based hydrogenase inactivation typically leads to the formation of an Ni-B (ready) state, capable of almost immediate reactivation by H_2 in O_2 tolerant hydrogenases. In O_2 sensitive hydrogenases, this state is further progressed to the Ni-A (unready) state, which exhibits much slower reactivation kinetics that can take several hours (Fernandez *et al.*, 1985; Fernandez *et al.*, 1986; Teixeira *et al.*, 1989; Stiebritz and Reiher, 2012).

Following O_2 addition during the active WGS phase, we observed varying degrees of reduction in H_2 production following O_2 depletion. It is possible that this reduction in productivity was the lingering result of O_2-derived inactivation/reduction in activity of the WGS machinery as CO was still present and unlikely to be limiting. While the lowest volume O_2 addition (1%), resulted in reduced productivity between the subsequent samplings following O_2 addition, H_2 productivity remained positive throughout and demonstrated a H_2 profile highly similar to that

of the control. Despite the minor reduction in H_2 productivity following the three hours post-O_2 addition, H_2 was still produced, a feature observed across each volume O_2 addition to greater effect in accordance with increasing O_2 volumes added. This could be explained as the effect of O_2 inhibition of the hydrogenogenic enzymes facilitating the WGS reaction, rationally with more pronounced inhibition with greater exposure to O_2 (Merrouch *et al.*, 2015; Lu and Koo, 2019; Kim *et al.*, 2022). Notably, however, unlike the 1% O_2 addition, between 3-6 h following O_2 addition, reductions in H_2 levels were observed in the 3-10% O_2 additions. Again, this was likely the influence of at least one of the two putative [NiFe]-uptake hydrogenases encoded on the genomes of strain, through coupling H_2 oxidation with aerobic respiration, although validation of this is still required (Greening *et al.*, 2016; Søndergaard *et al.*, 2016; Mohr *et al.*, 2018a; Cordero *et al.*, 2019).

The 5% and 10% O_2 additions experienced a minor decrease in observed OD_{600} in the 3 h post O_2 addition, prior to rapid increases in absorbance. This coincided with relatively low increases in CO_2 despite the rapid consumption of O_2. Possibly the sudden pronounced oxidative stress experienced by these cultures upon O_2 exposure contributed to the observed lag in growth (Partridge *et al.*, 2006; Ezraty *et al.*, 2017; Imlay, 2019; Aliyu *et al.*, 2021). A similar pattern of response following aeration of anaerobically cultured facultatively anaerobic *Klebsiella aerogenes* and *E. coli* has previously been observed. Consumption and high respiration rates of fermentatively produced metabolites were accompanied by lower-than-expected CO_2 production, severely hampered cell yields and apparent uncoupling between growth and energy conservation mechanisms, resulting in a switch towards energetically wasteful oxidation pathways (Harrison and Loveless, 1971). Considering this metabolic shift towards energetically wasteful processes under fluctuating O_2 levels, in conjunction with the O_2 susceptibility of the WGS machinery, it was unsurprising to have observed considerably lower H_2 production and yields following both temporally and volumetrically diverse O_2 perturbations.

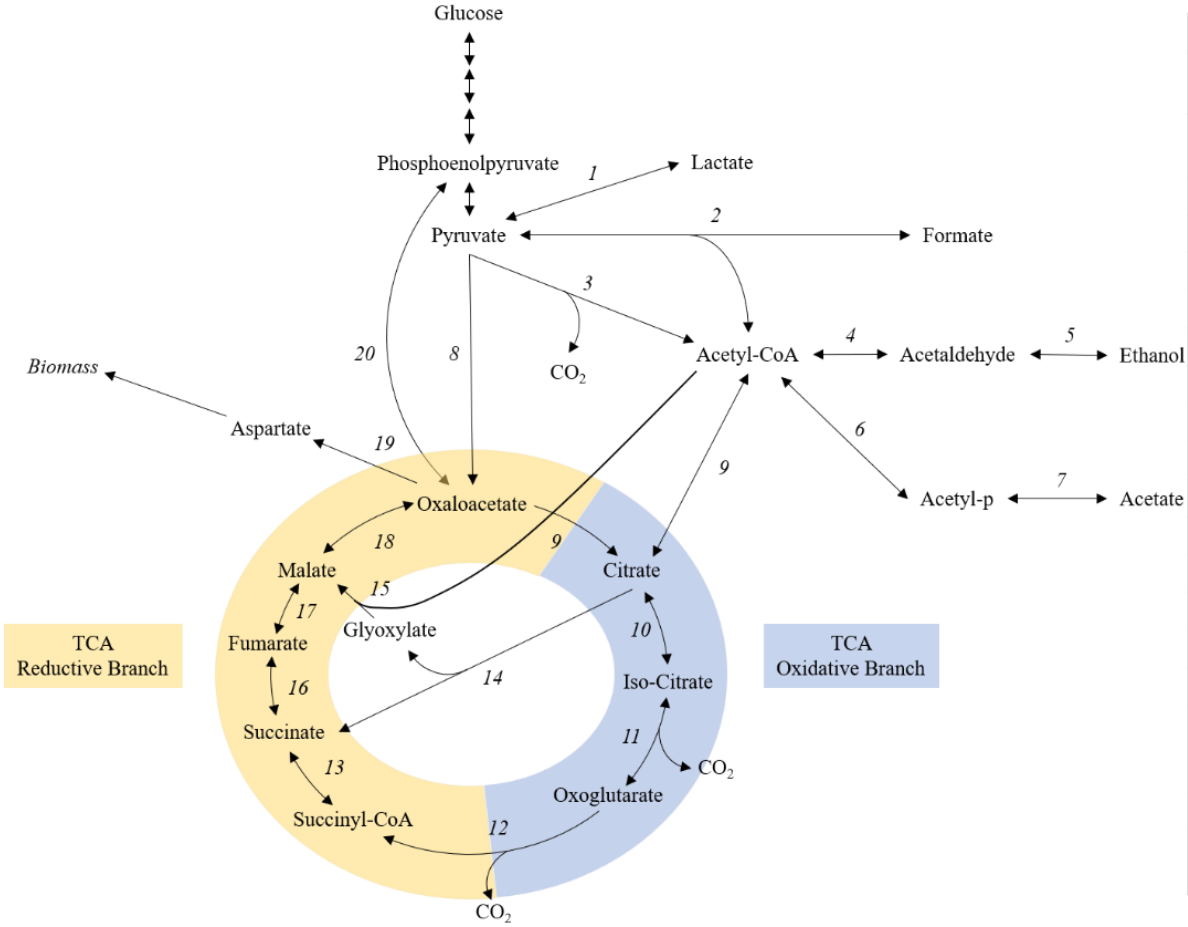
Metabolite analysis further exposed several pertinent observations of the impacts of O_2 disruption of the *P. thermoglucosidasius* WGS and metabolic responses upon O_2 perturbation in the central carbon metabolism (Figure 3.5). Across all conditions, large increases in acetate, formate and lactate were detected early in the fermentations and can account for the dramatic decrease in pH observed in this period. Initial accumulations of mixed acids, such as lactate served to reduce the backlog of NADH generated from aerobic metabolism with disrupted access to terminal oxidases by CO, as observed by Aliyu *et al*, (2021). Subsequently shortfalls

in ATP generation are compensated by ATP generating acetate production by acetate kinase. In conjunction, accumulation of these acids has been described as a common phenomenon, attributed to the development of microoxic conditions developed in the media during O_2 consumption (Liang *et al.*, 2022; Metcalfe *et al.*, 2022). These organic acids are considered over-flow metabolites, serving as electron sinks and being passaged extracellularly when glucose consumption proceeds at rates surpassing the cells capacity to oxidise NADH to NAD^+ via the respiratory chain (Campbell *et al.*, 2018).

Initially, during the early formate and lactate peaks, pyruvate was likely diverted from a saturated TCA cycle carbon pool towards production of these typically fermentative products, as a means to balance cell redox at the expense of energetic efficiency (Tang *et al.*, 2008; Aliyu *et al.*, 2021). The subsequent consumption of the lactate may have been made possible through the activity of the putative LutA-C protein complex encoded by *P. thermoglucosidasius*, proteins associated with aerobic lactate utilisation in *B. subtilis* (Chai *et al.*, 2009; Aliyu *et al.*, 2021). Formate levels decreased similarly and were essentially completely consumed by the onset of the WGS. This was noteworthy in itself as formate consumption is associated with concomitant CO_2 and H_2 production. No H_2 production was observed during this stage of formate consumption, suggesting diversion of electrons towards another oxidant or electron transport chain. It remains unclear as to exactly how formate is metabolised in these circumstances in *P. thermoglucosidasius* and represents an area for further investigation.

Strong acetate production in the aerobic phase alongside lactate and formate resembled previous *P. thermoglucosidasius* fermentations (Aliyu *et al.*, 2021). It is likely that acetate production served to compensate for the detraction of potential ATP production from lactate and formate production while the cell TCA cycle carbon load was saturated in the aerobic phase (Tang *et al.*, 2008; Aliyu *et al.*, 2021). Following this, acetate production was likely further bolstered through mixed acids fermentation to the detriment of ethanol directed fermentation, which aligns with the insubstantial levels of ethanol detected here and corroborates previous studies (Tang *et al.*, 2008). However, following O_2 addition at all timepoints strong acetate consumption accompanied delayed growth responses, increasingly so for larger volume O_2 additions. O_2 perturbation of steady-state anaerobically cultivated *E. coli* has previously been conducted by Yasid *et al*, (2016). However, whereas they observed no change in acetate, we observed acetate consumption following O_2 addition, with the assumption that a large proportion of the acetate was directed towards the acetyl-CoA pool.

Acetyl-CoA can be integrated into several pathways in *P. thermoglucosidasius*. During acetate consumption periods following O_2 addition in the various timepoint O_2 additions, succinate and glyoxylate increases were observed to increase towards the tail-end of acetate consumption, where O_2 levels were approaching depletion, except for the late 48 h O_2 addition. Notably, in the variable volume O_2 additions, greater degrees of CO_2 were evolved between O_2 addition and the resumption of WGS activity, correspondingly with greater O_2 additions, that could not be attributed to the WGS reaction, as H_2 was not being produced at similar equimolar amounts. Furthermore, despite appearing at low levels, following the 10% O_2 addition, there was a significant decrease in succinate, suggestive of respiratory succinate consumption. Together these factors suggest that at least a portion of the acetyl-CoA derived from the acetate consumed was likely integrated into the TCA cycle following conversion to acetyl-CoA, in conjunction with other biomass generating pathways. Following O_2 consumption, acetate consumption ceased, indicating a necessity for O_2 for concerted acetate consumption in *P. thermoglucosidasius*.



1	Lactate dehydrogenase
2	Pyruvate formate lyase/ Formate hydrogen lyase
3	Pyruvate dehydrogenase complex
4	Acetaldehyde dehydrogenase
5	Alcohol dehydrogenase
6	Acetyl phosphotransferase
7	Acetate kinase
8	Pyruvate carboxylase
9	Citrate synthase
10	Aconitate hydratase
11	Isocitrate dehydrogenase
12	2-oxoglutarate dehydrogenase
13	Succinyl-CoA ligase
14	Isocitrate lyase
15	Malate synthase
16	Succinate dehydrogenase
17	Fumarate hydratase
18	Malate:quinone oxidoreductase/Malate dehydrogenase
19	Aspartamate transaminase
20	Phosphoenolpyruvate carboxykinase

Figure 3.5: Predicted glucose metabolism pathway for *P. thermoglucosidasius*. Adapted from Mohr et al, (2018a); Aliyu et al, (2021); Mol et al, (2021).

3.5 Conclusions

Cultivation of *P. thermoglucosidasius* DSM 6285 under previously unexplored, industrially relevant anaerobic synthetic syngas atmospheres induced faster onset of H_2 production than previously identified ideal aerobic gas mixtures, however, this was at the expense of biomass accumulation. This in turn severely hampered H_2 turnover. O_2 supplementation at all stages of the fermentation hindered H_2 production to greater extents accordingly with the volume of O_2 supplied, however, also supported reinvigorated biomass accumulation. Furthermore, upon depletion of the supplemented O_2, hydrogenogenic activity was rapidly resumed following a negligible lag period. Supplementation with 1% O_2 negligibly impacted H_2 production, however, still gave some support to growth. Notably, with elevated (10%) O_2 supplementation, H_2 consumption was observed briefly. This was potentially due to the two putative uptake hydrogenases encoded by *P. thermoglucosidasius* DSM 6285. H_2 conversion efficiency was also reduced accordingly with the level of O_2 administered. The tolerance of the system to O_2 perturbations appeared to be linked with the metabolism of accumulated acetate. O_2 addition appeared to induce strong acetate consumption to support cell growth at all stages of the fermentation, including during the aerobic/microaerobic phase. It may be possible to incorporate this capacity for O_2 induced acetate consumption and the relative flexibility of the *P. thermoglucosidasius* DSM 6285 WGS reaction to O_2 perturbation to facilitate cyclical/fed batch hydrogenogenic fermentations on syngas at a larger scale. Additionally, focusing the strain towards concerted acetate production through inactivation of formate or lactate production pathways may facilitate concentrated acetate production to further supplement downstream acetate dependent growth.

CHAPTER FOUR: ELUCIDATING THE BIOTECHNOLOGICAL POTENTIAL OF THE GENERA *PARAGEOBACILLUS* AND *SACCHAROCOCCUS* THROUGH COMPARATIVE GENOMIC AND PAN-GENOMIC ANALYSIS

Abstract

The genus *Geobacillus* and its associated taxa have served as a focal point of numerous thermophilic biotechnological investigations, both at the whole cell and enzyme level. However, comparatively little research has been conducted on its recently delineated sister genus, *Parageobacillus.* Here we performed phylogenomic and pan-genomic analyses on a subset of publicly available *Parageobacillus* genomes to elucidate their biotechnological potential. Phylogenomic analysis delineated the compared taxa into two distinct genera, *Parageobacillus* and *Saccharococcus*. Both genera present open pan-genomes, with the species *P. toebii* being characterized by the highest novel gene accrual. Diversification of the two genera appears to be driven through the activity of plasmids, bacteriophages and transposable elements. Both genera, *Parageobacillus* in particular, present a range of potentially biotechnologically relevant features, including a source of novel antimicrobials, thermostable enzymes including DNA-active enzymes, carbohydrate active enzymes, proteases, lipases and carboxylesterases. Both genera also present a number of metabolic pathways pertinent to degradation of complex hydrocarbons and xenobiotics including, aromatics, long-chain hydrocarbons and heavy metals. Furthermore, in addition to the applicability of a subset of *Parageobacillus* to H_2 gas production, the genus also harbours several alternative pathways of appeal towards green energy production. Comparative genomic analyses suggest that *Parageobacillus* and *Saccharococcus* can be delineated as two distinct genera, both of which can serve as a rich source of biotechnologically and industrially relevant secondary metabolites, thermostable enzymes and metabolic pathways that warrant further investigation.

4.1 Background

The genus *Geobacillus* has served as an epicenter for biotechnological exploitation of thermophilic taxa (Zeigler, 2014; Hussein *et al.*, 2015; Finore *et al.*, 2019). First described following the 16s rRNA gene-based reclassification of previously recognised thermophilic clustering (group 5) *Bacillus* spp. (Ash *et al.*, 1991), the genus currently comprises 12 validly described species (Parte *et al.*, 2020). Members are Gram-positive, aerobic or facultatively anaerobic, spore forming rods, characterised by their thermophilicity, being capable of growth at temperatures ranging between 37-80˚C (Najar and Thakur, 2020). Key taxa of biotechnological value include *Geobacillus stearothermophilus, G. thermoleovorans* and *G. thermodenitrificans* (Hussein *et al.*, 2015; Novik *et al.*, 2019). These and other taxa in the genus have been the topic of research and commercial development in a wide range of whole-cell applications, including bioremediation, crude oil recovery and refinement, textile processing, synthesis of nanoparticles, production of antibiotics and production of value-added chemicals such as biodiesel, lactate and ethanol (Hussein *et al.*, 2015; Novik *et al.*, 2019; Najar and Thakur, 2020; Khaswal *et al.*, 2022). Geobacilli further serve as a source of various thermostable enzymes which present comparably more cost-effective, rapid, non-toxic and environmentally friendly alternatives to whole-cell or abiotic processes that support diverse industries (Novik *et al.*, 2019; Najar and Thakur, 2020). The application of thermophile derived enzymes has become more prevalent due to their greater thermostability, pH tolerance, catalytic efficiency and reduced cost and contamination rates associated with thermophilic operation (Atalah *et al.*, 2019; Han *et al.*, 2019). Consequently, *Geobacillus*-derived enzymes including α-amylases, α-glucosidases, cellulases, lipases, pectinases, xylanases have received extensive interest for their applicability towards agricultural, biofuel, food, paper, petrochemical, pharmaceutical and textile industries (Hussein *et al.*, 2015; Novik *et al.*, 2019; Najar and Thakur, 2020; Bhalla *et al.*, 2022).

The application of whole genome phylogenetic approaches recently highlighted the clustering of *Geobacillus* taxa in two distinct clades, which were further distinguished based on G+C content, resulting in the establishment of the genus *Parageobacillus* (Aliyu *et al.*, 2016; Aliyu *et al.*, 2018). This genus currently comprises six validly described species (Parte *et al.*, 2020), which are readily isolated from diverse and globally distributed high temperature environments, including hot springs (Brumm *et al.*, 2015), oil wells (Nazina *et al.*, 2005), hot composts (Poli *et al.*, 2016) and geothermal sites and sediments (Omae *et al.*, 2019; Imaura *et al.*, 2023). However, they have also been isolated from moderate and low temperature

environments, like temperate soils (Marchant *et al.*, 2002), plant matter (Sung *et al.*, 2002) and marine sediments (Inoue *et al.*, 2019). In congruence with their wide and varied distribution, the group contains versatile metabolic potential, encoding a range of robust, thermostable and thermoactive enzymes, many of which may be of potential biotechnological value (Hussein *et al.*, 2015; Najar and Thakur, 2020; Khaswal *et al.*, 2022). While some research has probed the biotechnological potential of *P. thermoglucosidasius*, the inherent capacity of the genus as a whole, in comparison to its sister genus, *Geobacillus*, remains relatively underexplored. Here we have made of use of whole genome sequence data and phylogenomic approaches to establish the relationship of taxa within the genus and further dissociate them into two genera, *Parageobacillus* and *Saccharococcus*. Further, using comparative genomic and pan-genome analyses, we provided an in-depth characterization of the biotechnological potential of these key thermophilic taxa.

4.2 Methods

4.2.1 Genome assembly and annotation

The publicly available genome sequences of thirty-four *Parageobacillus* taxa*, Saccharococcus thermophilus* DSM 4749^T and *G. thermodenitrificans* DSM 465^T (used for comparative and outgroup purposes) were obtained from the NCBI genome assembly database (Appendix C Supp. Table S4.1; Kitts *et al.*, 2016). Genome assemblies were improved using the MeDuSa genome scaffolder (Bosi *et al.*, 2015). All genomes were structurally annotated using Prodigal v.2.6.3 (Hyatt *et al.*, 2010) and the proteome datasets functionally annotated and assigned to COG categories, using eggnog-mapper v. 2.1.12 (Cantalapiedra *et al.*, 2021) against the eggNOG v. 5.0 database (Huerta-Cepas *et al.*, 2019). The subcellular localisations of all proteins encoded on each genome were determined using PSORTb v. 3.0.3 (Yu *et al.*, 2010). Plasmids and transposable elements were identified on the basis of the eggNOG annotations, while phage elements were identified using the PHASTER server (Arndt *et al.*, 2016).

Biotechnologically relevant enzymes were identified and characterised using several pipelines. Secondary metabolite biosynthetic loci were identified in the *Parageobacillus* and *Saccharococcus* genomes using antiSMASH v. 7.0.1 (Blin *et al.*, 2023) and further confirmed and characterised using the BAGEL 4 (van Heel *et al.*, 2018) and RiPPMiner-Genome (Agrawal *et al.*, 2021) servers. CAZYmes were predicted from the protein datasets for each genome using the HMMer, Hotpep and DIAMOND tools of DbCAN3 (Zheng *et al.*, 2023) against the CAZYme database (Drula *et al.*, 2022), where only those predictions made by ≥ 2 tools were considered as positive hits. Proteases/peptidases and lipases were identified and characterised by aligning the proteome datasets for each compared *Parageobacillus* and *Saccharococcus* strain against the MEROPS v. 11.0 database (Rawlings *et al.*, 2018) and the Lipase Engineering Database (LED) v. 4.1.0 (Fischer and Pleiss, 2003) respectively, using DIAMOND v. 2.1.8 (Buchfink *et al.*, 2021). CRISPR/Cas associated proteins were predicted through the CRISPRCasFinder tool v. 1.1.2 - I2BC (Grissa *et al.*, 2007; Abby *et al.*, 2014; Couvin *et al.*, 2018). Other proteins of potential biotechnological relevance were identified by localized Blast analyses and alignment using Bioedit v. 7.7.1 (Hall, 1999). Restriction-modification genes were tentatively identified on the basis of the eggnog-mapper annotations and confirmed through Blastp analysis to the REBASE database (Roberts *et al.*, 2015). Heavy metal resistance genes were identified by aligning the proteome datasets for each compared

Parageobacillus and *Saccharococcus* taxon against the BacMet2 predicted resistance genes database using DIAMOND v. 2.1.8 (Pal *et al.*, 2014; Buchfink *et al.*, 2021).

4.2.2 Phylogenomic analyses

The proteome datasets for each comparator strain (and *G. thermodenitrificans* DSM 465^T as an outgroup) were compared and clustered into their orthologous groups using Orthofinder v. 2.5.5 (Emms and Kelly, 2019). Single copy orthologous (SCO) proteins conserved among all taxa (1,784 SCOs) were individually aligned using the M-Coffee implementation of T-Coffee v. 13.46.0.919e8c6b (Wallace *et al.*, 2006), concatenated and poorly aligned blocks were removed using GBlocks v. 0.91b (Talavera and Castresana, 2007). The trimmed concatenated alignment was used to construct a maximum likelihood (ML) phylogeny using IQ-Tree v. 2.2.0 (Nguyen *et al.*, 2015) with the optimal evolutionary model predicted using ModelFinder (Kalyaanamoorthy *et al.*, 2017). Branch support was provided using ultrafast bootstrap (UFBoot2) analysis (n = 1,000 replicates; Hoang *et al.*, 2018). Support for the core genome phylogeny and species delineation was provided by calculating the Average Nucleotide Identity (ANI) values with the OAT tool v. 0.9.1 (Lee *et al.*, 2016) and digital DNA-DNA hybridization values (dDDH) were determined using the Genome-to-Genome Distance Calculator (GGDC 3.0; Meier-Kolthoff *et al.*, 2022), where taxa sharing OrthoANI values > 96% and dDDH values > 70% were considered to belong to the same species (Lee *et al.*, 2016; Meier-Kolthoff *et al.*, 2022).

4.2.3 Pan-genome analyses

The Orthofinder output discussed above was used to identify the core (conserved among all taxa), accessory (shared by several but not all compared strains) and unique (to a single taxon) proteome fractions of the compared *Parageobacillus* and *Saccharococcus* taxa. The presence (1) or absence (0) of each orthogroup was scored and the pan-genome of different datasets (*Parageobacillus/Saccharococcus; P. thermoglucosidasius/P. toebii/S. caldoxylosilyticus*) were used to determine the pan-genome using the bacterial pan-genome analysis (BPGA) pipeline (Chaudhari *et al.*, 2016) and extrapolated (to 100 genomes/per set of taxa) using PanGP (Zhao *et al.*, 2014). The functions of the core, accessory and unique pan-genome fractions were determined by comparison of the pan-genome element-specific proteome datasets against the eggNOG v. 5.0 database (Huerta-Cepas *et al.*, 2019) using eggnog-mapper v. 2.1.12 (Cantalapiedra *et al.*, 2021).

4.3 Results and discussion

4.3.1 Phylogenomic analysis delineates *Parageobacillus* and *Saccharococcus* as two distinct genera

The genus *Parageobacillus* was resolved from the genus *Geobacillus* using phylogenomic analysis, and comprises six distinct species, including *P. caldoxylosilyticus* (Aliyu *et al.*, 2016). However, the taxonomic status of the latter species remains contentious, having first been assigned to the genus *Saccharococcus*, subsequently the genus *Geobacillus* and finally the genus *Parageobacillus* (Ahmad *et al.*, 2000; Fortina *et al.*, 2001; Aliyu *et al.*, 2016). In this study a core genome maximum likelihood (ML) phylogeny was constructed on the basis of 1,784 single-copy orthologous proteins conserved among 34 *Parageobacillus* strains, the *Saccharococcus thermophilus* DSM 4749^{T} genus type and the outgroup strain *Geobacillus thermodenitrificans* DSM 465^{T}. This phylogeny showed the clear delineation of the taxa in two distinct clades (Figure 4.1), with the nine *P. caldoxylosilyticus* strains and *Parageobacillus* genomosp. A NUB3621 clustering with *S. thermophilus* DSM 4749^{T}, indicating they belong to the genus *Saccharococcus*. This is further supported by the phylogenomic metrics ANI and dDDH, where intraclade ANI values of 92.22 and 96.16% and dDDH values of 59.19 and 73.72% were observed for the *Parageobacillus* and *Saccharococcus* clades, respectively, while interclade values were 83.57% (ANI) and 27.57% (dDDH; Appendix C Supp. Table S4.2). Two *Parageobacillus* strains, namely KH3-4 and W-2 dDDH (average 44.04%) and ANI (90.01%)

were below the 70% and 96% threshold values that constitute the species boundaries (Aliyu *et al.*, 2016) and as such, they form a novel genomospecies, *Parageobacillus* genomosp. A.

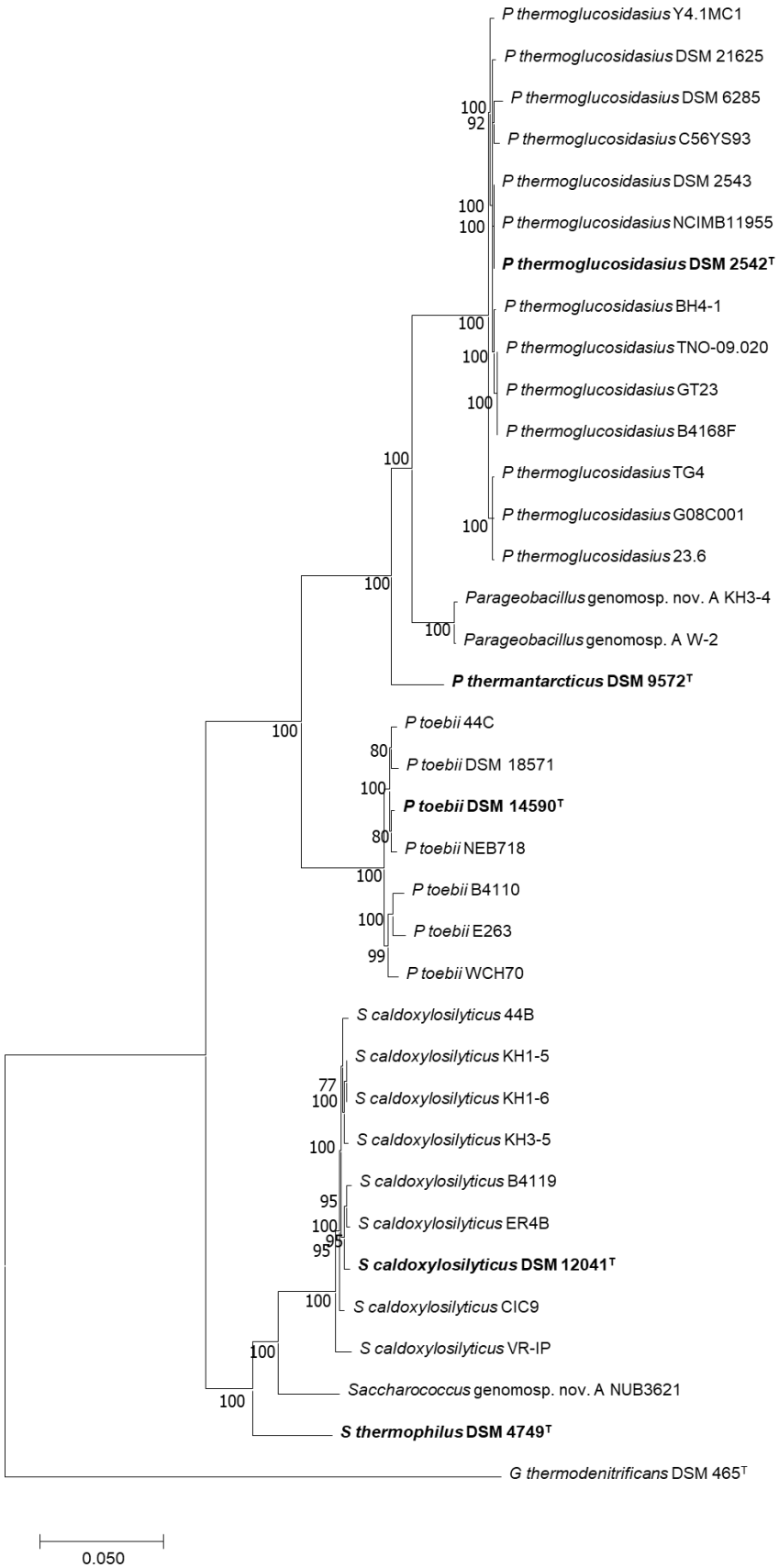
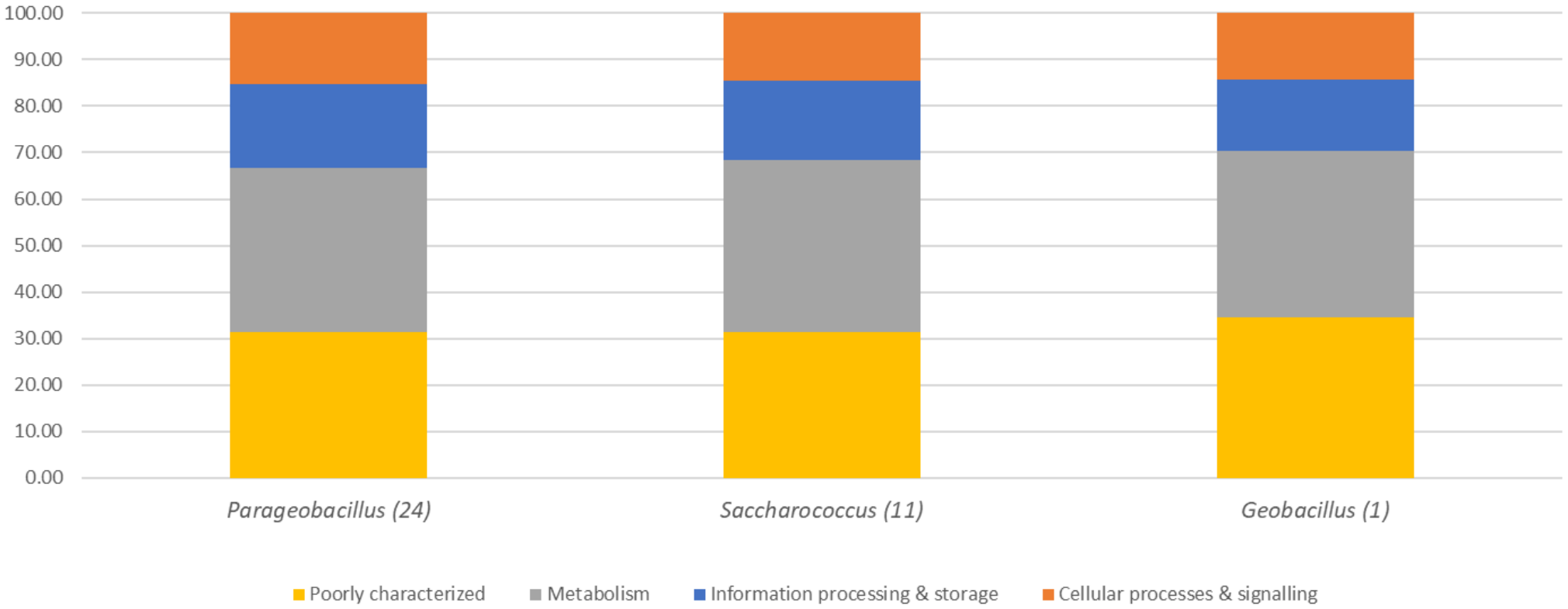


Figure 4.1: Core genome ML phylogeny of the genera *Parageobacillus* and *Saccharococcus*, with *G. thermodenitrificans* DSM 465^T as an outgroup. The phylogeny was constructed on the basis of 1,784 SCOs, with the concatenated alignment comprising 499,928 amino acid positions, of which 81,368 were deemed parsimony informative and 58,513 represented single sites. The ML phylogeny was constructed using the optimal evolutionary model JTT+F+I+G4 with ultrafast bootstrap support (n= 1,000 replicates).

The genomes of members of both *Parageobacillus* and *Saccharococcus* are similar in size (average: 3.763 and 3.742 Mb, respectively), while the G+C content of members of the genus *Saccharococcus* are on average 0.94% greater than their *Parageobacillus* counterparts (Supp. Table 4.1). The genomes of taxa in both genera code for a similar number of proteins (3,704 and 3,719, respectively), with the most proteins encoded on the genome of *S. caldoxylosilyticus* B4119, followed by three *P. thermoglucosidasius* strains (Appendix C Supp. Table S4.1). In general, fewer proteins are encoded on the genomes of *P. toebii* strains (average 3,461 proteins) The least proteins (3,085) are encoded on the genome of *S. thermophilus* DSM 4749^{T}, with a genome that is also ~ 650 kb smaller than the other comparator taxa (Appendix C Supp. Table S4.1). Analysis of the COG functions associated with the proteomes of each strain showed that slightly more proteins (~2% or 84 proteins on average) involved in metabolism are encoded on the genomes of *Saccharococcus* taxa, which can primarily be attributed to the COG categories amino acid (E), nucleotide (F) and lipid (I) transport and metabolism (Figure 4.2). By contrast, *G. thermodenitrificans* DSM 465^{T} codes for substantially fewer (2.5%) proteins involved in information processing and storage (primarily in COG category L - replication, recombination and repair) and a greater (3.2%) number of proteins of unknown function (Figure 4.2).



	Information processing & storage	Cellular processes & signalling	Metabolism	Poorly characterized
Parageobacillus	18.01	15.37	35.26	31.41
Saccharococcus	17.03	14.83	37.21	31.31
G. thermodenitrificans	15.25	14.33	35.82	34.60

Figure 4.2: COG supra-functional categorisation of proteins encoded on the genomes of *Parageobacillus*, *Saccharococcus* and *G. thermodenitrificans* DSM 465^{T}. The number of proteins assigned to COG functional categories are shown in brackets above each bar. The relative proportions for each COG category are given below the figure. Values > mean + standard deviation are highlighted in green, while those < mean – standard deviation are highlighted in red.

4.3.2 *Parageobacillus* and *Saccharococcus* have open pan-genomes with *P. toebii* as a key driver of novel gene accrual

The core (conserved among all taxa in a set), accessory (conserved among some taxa or unique to specific taxon in set) and pan-genome (combination of core and accessory fractions) for the genera *Parageobacillus* and *Saccharococcus* were determined. The overall pan-genome of both genera combined (taxa) comprises 9,082 orthogroups, of which 1,950 (21.5%) are core to all taxa (Figure 4.3A). A total 37.1% and 15.4% of the orthogroups are unique to the genera *Parageobacillus* and *Saccharococcus*, respectively. Analysis of the functions of the core and *Parageobacillus*- and *Saccharococcus*-unique fractions showed that carbohydrate transport and metabolism (COG category G) in particular is overrepresented in the genus-specific proteome datasets, suggesting distinct metabolic capacities for the two genera. Furthermore, the synthesis of secondary metabolite biosynthesis (COG category Q) and defense mechanisms (COG category V) are largely genus-specific traits (Figure 4.3B), with only 11 and 21 of orthogroups in this category core to all *Parageobacillus* and *Saccharococcus* taxa in each set, respectively. The *Parageobacillus*-unique proteins core to all taxa in this genus are dominated by transcription regulators (four proteins), while the *Saccharococcus*-unique core proteins include three proteins involved in amino acid transport and three proteins involved in copper resistance (CotA, CopC and YcnI; Figure 4.3B).

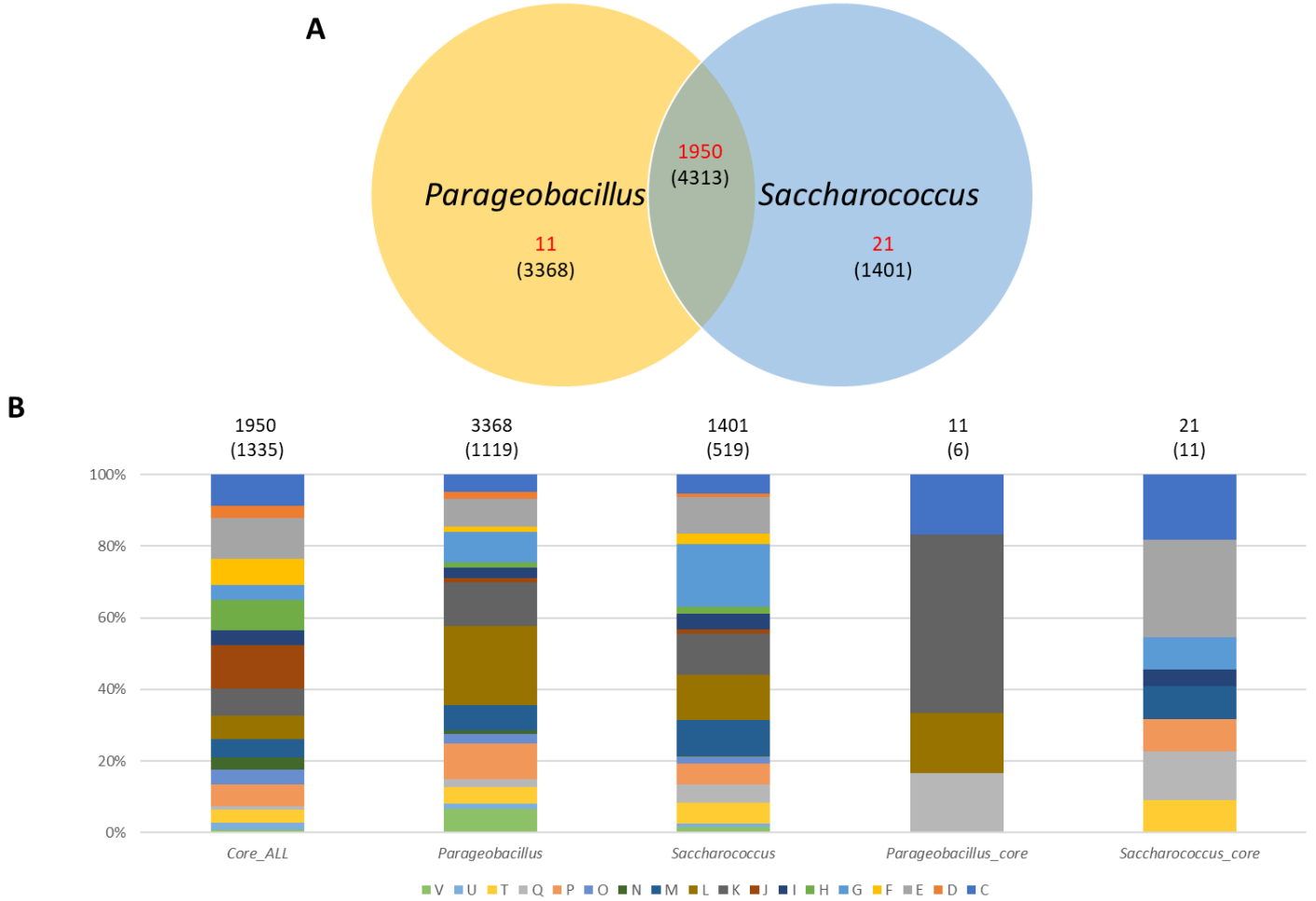


Figure 4.3: Protein conservation and function in the genera *Parageobacillus* and *Saccharococcus*. (A.) Proportion of conserved and genus-specific orthogroups for the two genera. Numbers in red indicate those proteins conserved in all taxa within each dataset, while those in brackets reflect those that are present in some, but not necessarily all taxa within each genus or the combination of the genera. (B) Conserved Orthologous Group functional categories for the proteins conserved among all taxa in both genera, those specific to each genus and those that are conserved among all taxa in each genus. The graphs represent the relative proportions as calculated for those proteins categorized for a COG function (with the exception of category S: function unknown). The number of proteins assigned to COG functional categories are shown in brackets above each bar.

Pan and core genome graphs were constructed for the genera *Parageobacillus* and *Saccharococcus* and extrapolated to encompass 100 genomes/genus (Figure 4.4A), to model and describe the potential genomic totality, shared/unique and accessory/core genes of the respective taxa. Both genera display an open pan-genome, i.e., a pangenome that increases in accordance with the addition of new genomes, with that of *Parageobacillus* being slightly larger than the genus *Saccharococcus*. This is indicative of a greater degree of genetic diversity among these taxa, which may harbour potentially novel genes or gene variant of biotechnological significance. Similar numbers of new genes (24.4 and 24.8) are predicted to be added to the pan-genome when the 100th genome of *Parageobacillus* and *Saccharococcus* are sequenced. When considering genome conservation, the core genome of *Saccharococcus* is predicted to be slightly larger (2,332) than that of *Parageobacillus* (2,171) across 100 genomes.

To evaluate the pan-genome dynamics of individual species within each genus, the pan- and core-genomes of three species for which ≥ 7 genomes are available (*P. thermoglucosidasius* – 14 genomes*, P. toebii* – 7 genomes and *S. caldoxylosilyticus* -8 genomes), were extrapolated (Figure 4.4B). Open pan-genomes were observed for all three species, with similar pan- and core-genome trends for *P. thermoglucosidasius* and *S. caldoxylosilyticus*. For both species, the core genome approached a predicted average of 3,012 orthogroups when 100 genomes are sequenced, while the 100th genome would add 15 novel proteins to the pan-genome of both species. By contrast, a much larger (~2,800 more orthogroups when considering 100 genomes) pan-genome is observed for *P. toebii* than the other species, with 24.4 novel proteins added by the 100th taxon genome included in the analysis. This species further has a substantially smaller core genome, with almost 600 core orthogroups less than the other two species. This suggests that *P. toebii* has a more unstable pan-genome than *P. thermoglucosidasius* and *S. caldoxylosilyticus* and that this species may be capable of greater ecological, metabolic and functional diversification than the two latter species (De Maayer *et al*., 2014a). This is further supported when considering the genomes incorporated in this study, where *P. toebii*-specific orthogroups (seven genomes) contribute 14.6%, while *P. thermoglucosidasius*-specific proteins (with double the number of

genomes analysed) contribute 18.5% (Figure 4.5A). The largest proportion of proteins involved in the supra-functional category information storage and processing (38.4%) is observed for the *P. toebii*-specific protein complement, while the *P. thermoglucosidasius*-specific proteins are primarily involved in metabolism (Figure 4.5B). These species-specific datasets are predominated by proteins involved in DNA replication, recombination and repair (25.9%) and carbohydrate (10.4%) as well as amino acid (10.1%) transport and metabolism, respectively (Figure 4.5C).

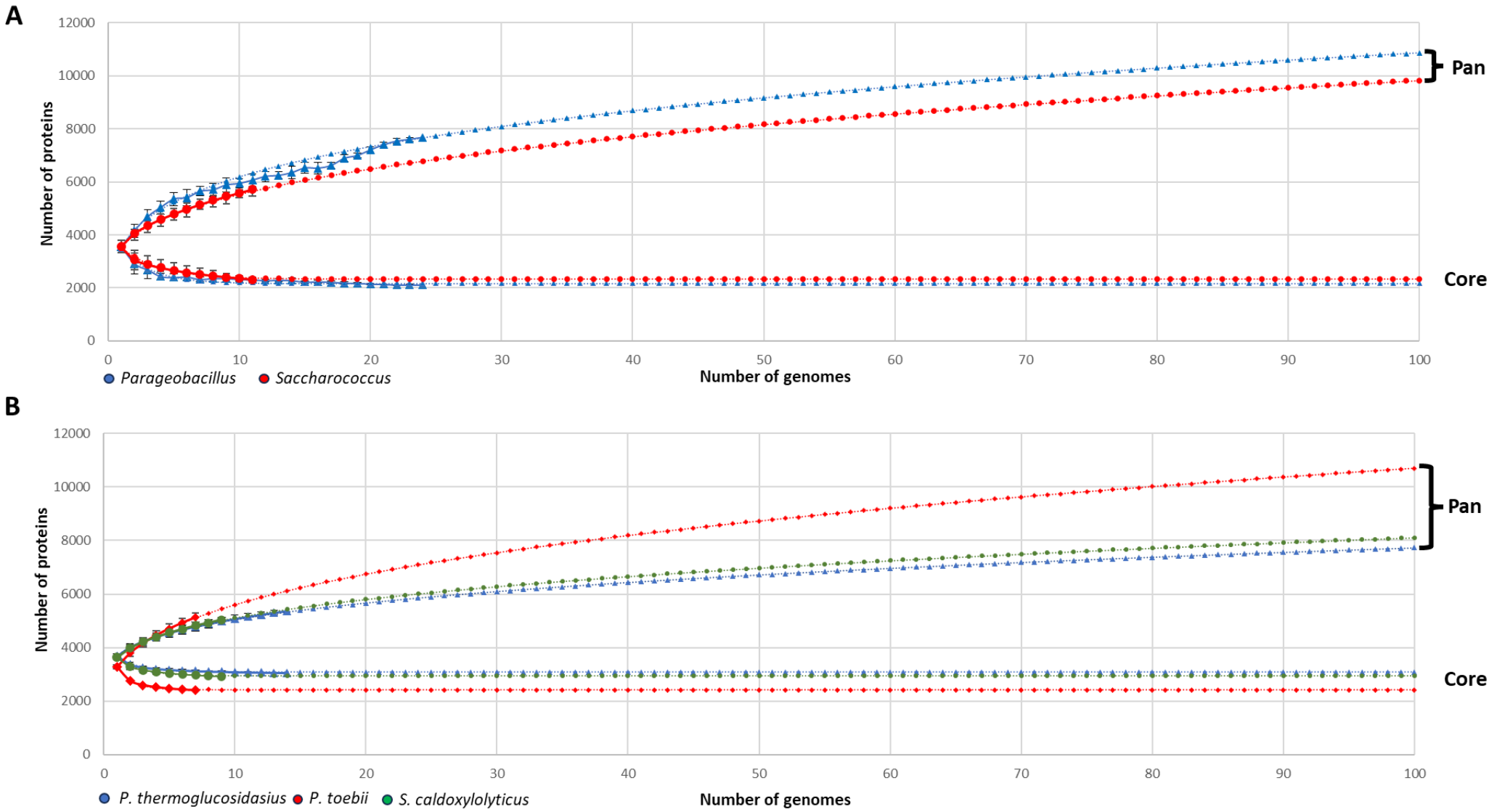


Figure 4.4: Pan- and core-genome graphs. A) Pan- and core genomes for the genera *Parageobacillus* (blue) and *Saccharococcus* (red). Larger shapes indicate the actual values, while smaller shapes depict the extrapolated values. B) The pan- and core genomes of *P. thermoglucosidasius* (blue), *P. toebii* (red) and *S. caldoxylosilyticus* (green).

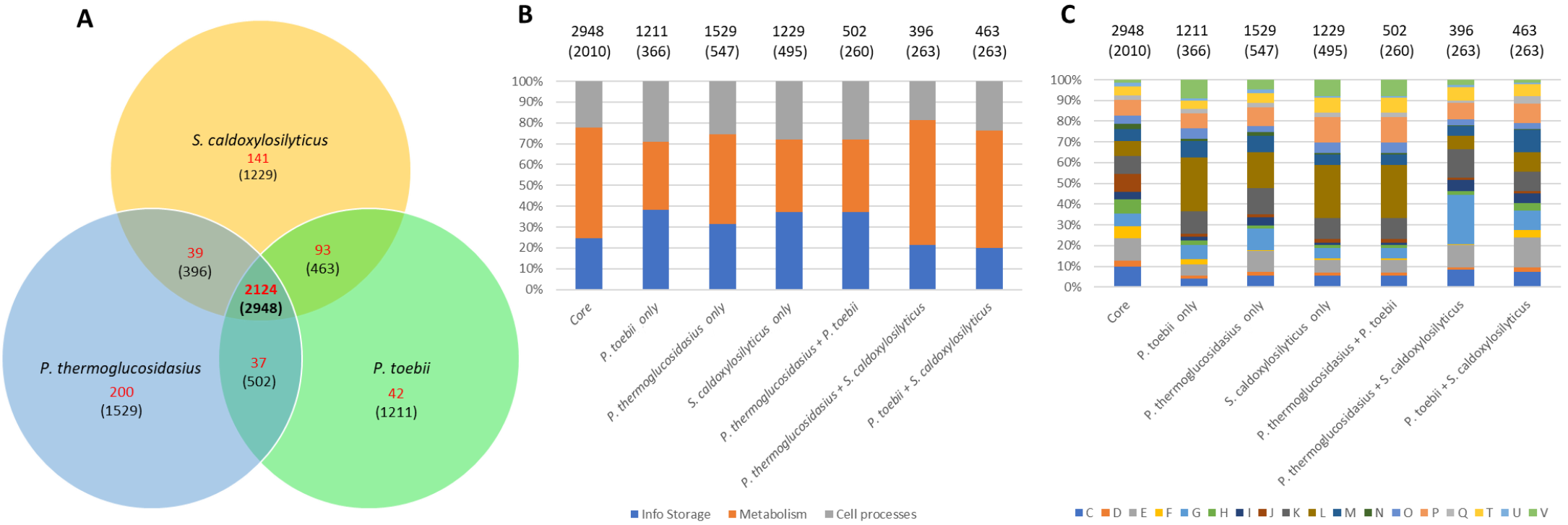


Figure 4.5: Protein conservation and function in the species *P. thermoglucosidasius*, *P. toebii* and *S. caldoxylosilyticus*. A) Proportion of conserved and genus-specific orthogroups for the two genera. Numbers in red indicate those proteins conserved in all taxa within each dataset, while those in brackets reflect those that are present in some, but not necessarily all taxa within each genus or the combination of the genera. B) Conserved Orthologous Group supra-functional categories for the proteins conserved among all taxa in all species, those specific to each species and those shared between 2/3 species. C) Functional categorisation of the pan-genome elements on the basis of the COG functional

categories. One-letter abbreviations for the functional categories: C, energy production and conversion; D, cell division and chromosome partitioning; E, amino acid metabolism and transport; F, nucleotide metabolism and transport; G, carbohydrate metabolism and transport; H, coenzyme metabolism; I, lipid metabolism; J, translation, including ribosome structure and biogenesis; K, transcription; L, replication, recombination and repair; M, cell wall structure and biogenesis and outer membrane; N, secretion, motility and chemotaxis; O, molecular chaperones and related functions; P, inorganic ion transport and metabolism; R, general functional prediction only T, signal transduction. The number of proteins assigned to COG functional categories are shown in brackets above each bar. The graphs represent the relative proportions as calculated for those proteins categorized for a COG function (with the exception of category S: function unknown). The number of proteins assigned to COG functional categories are shown in brackets above each bar.

4.3.3 Plasmids, bacteriophages and transposable elements are key drivers of *Parageobacillus* and *Saccharococcus* diversification

Plasmid replicons, prophages and transposable elements were predicted for the comparator *Parageobacillus* and *Saccharococcus* taxa. Plasmid replicons occur in 75% and 54.5% of the taxa in each genus, respectively (Appendix C Supp. Table S4.1). Half of the plasmid-bearing *Parageobacillus* taxa incorporate two plasmids, while *S. thermophilus* DSM 4749^{T} harbours two plasmids and *S. caldoxylosilyticus* B4119 is predicted to carry three distinct plasmids. The plasmids vary substantially in size, with the smallest (1,080 nucleotides) and largest (~ 105 kilobases) both occurring in *P. thermoglucosidasius* G08C001. These plasmids contribute up to 3.81% and 4.31% of the total genome and protein complement (highest for both observed in *P. thermoglucosidasius* 23.6; Appendix C Supp. Table S4.1). Prophage elements are more prevalent in both genera, with between one and eight (*Parageobacillus* genomosp. A. W-2) elements per genome (Appendix C Supp. Table S4.1). In most cases these prophage elements are predicted to be incomplete, but three complete phage elements are predicted on the genome of *Parageobacillus* genomosp. A W-2 and phage-proteins contribute 6.29% of the total proteins encoded on the genome of the latter strain.

Between 29 (*Saccharococcus* genomosp. A NUB3621) and 263 (*P. toebii* WCH70) transposases (belonging to 74 distinct orthogroups) were predicted per genome. Notably, *P. toebii* incorporate an average of 124 transposases per genome, while *P. thermoglucosidasius* and *S. caldoxylosilyticus* genomes incorporate an average of 50 and 63 transposases, respectively, indicating a key role for transposition in the diversification of *P. toebii*.

When considering plasmids, prophages and transposases in combination, these elements contribute 6.1% and 5.1% to the total genome protein contents for *Parageobacillus* and *Saccharococcus*, respectively, while for the comparator *G. thermodenitrificans* DSM 465^{T}, these elements encompass only 3.8% of the total proteome. Stand-out taxa include *P. thermantarcticus* DSM 9572^{T} and *P. toebii* WCH70, where these elements in combination, contribute 9.7% (primarily prophage elements) and 10.1% (primarily transposases) of the total protein content,

highlighting the combined role of these elements in shaping the highly versatile genera *Parageobacillus* and *Saccharococcus*. Given the genomic versatility and extensive core genome the *Parageobacillus* and *Saccharococcus* genome dataset was evaluated for proteins of potential biotechnological value.

4.3.4 Mining the *Parageobacillus* and *Saccharococcus* pan-genome for biotechnology

4.3.4.1 *Parageobacillus* as a source of novel antimicrobials

The emergence and rapid spread of antibiotic resistance among clinically relevant pathogens has driven the continued search for novel natural products to combat these pathogens (Rossiter *et al.*, 2017). To this extent, the geobacilli have been garnering attention, with several studies identifying bacteriocins and bacteriocin-like inhibitory substances effective against a range of different pathogens (Garg *et al.*, 2012; Egan *et al.*, 2018; Zebrowska *et al.*, 2022). antiSMASH predicted on average 5.3 and 6.4 secondary metabolite biosynthetic loci in members of the genera *Parageobacillus* and *Saccharococcus*, respectively. Included among these are loci for the synthesis of metallophores (three types), betalactones (three types), a ladderane, betalactones (two types), a spore-killing factor and eight distinct bacteriocin biosynthetic loci. The latter loci were further confirmed and characterised using the BAGEL 4 (van Heel *et al.*, 2018) and RiPPMiner-Genome (Agrawal *et al.*, 2021) servers.

A collection of six Class I and two Class II bacteriocins are distributed across the genome dataset. The Class I bacteriocins comprise four lantibiotic loci, a linear azole-containing peptide and a thiopeptide biosynthetic locus. The best-known *Geobacillus* antimicrobials are the lantibiotics geobacillin I and II from *G. thermodenitrificans*, effective against vancomycin-resistant *Enterococcus faecium*/methicillin-resistant *Staphylococcus aureus* and *Bacillus cereus*/*B. subtilis*, respectively (Garg *et al.*, 2012). The geobacillin I locus comprises ten genes, including *geoAI* which codes for the bacteriocin peptide, while the geobacillin II locus comprises three genes, with *geoAII* encoding the bacteriocin peptide (Bergonzi *et al.*, 2016). A complete geobacillin I locus was identified in a single

taxon in our dataset, namely *P. thermantarcticus* DSM 9572 (84.6% average amino acid identity across 10 proteins; 92.9% average AAI for GeoAI bacteriocin peptide to *G. thermodenitrificans* NG80-2; Figure 4.6). Of note, 19/23 of the other *Parageobacillus* strains encode orthologues of geobacillin I self-immunity (*geoEFGI*) and two-component regulatory systems (*geoKR*; Netz *et al.*, 2002), suggesting they have immunity to the geobacillin I lantibiotic but are unable to produce it themselves. Only a single taxon in the dataset, *P. toebii* B4110, incorporates a geobacillin II locus (Figure 4.6), which was previously shown to be more restricted in distribution than geobacillin I (only in two *G. thermodenitrificans* strains). The locus encodes all three proteins produced by the *G. thermodenitrificans* NG80-2 geobacillin II locus (99.9% average AAI). Downstream of the *P. toebii* locus are three genes coding for orthologues of erythromycin-like esterases (cd14728 – ere-like), which provide resistance to macrolides (Zieliński *et al.*, 2021), and may potentially serve as a self-immunity mechanism for geobacillin II.

Two further lantibiotic biosynthetic loci types, lantibiotic III and lV were predicted on the genomes of 4/7 *P. toebii* strains evaluated and *S. caldoxylosilyticus* B4119, respectively (Figure 4.6). The lantibiotic III cluster was previously identified *in silico* as lantibiotic cluster 4/5 (Egan *et al.*, 2018) while the predicted bacteriocin peptide is a predicted esterase/lipase (cd00312). The lantibiotic IV cluster, novel to this study, includes genes coding for a lantibiotic dehydratase (*lanB*), cyclase (*lanC*) and ABC transporter (*lanT*), showing limited homology to the subtilin biosynthetic proteins SpaBCT of *Bacillus subtilis* (P33115-6; P39774.2; 33.5% average AAI). Two predicted FDLD family class I lanthipeptides (sharing 69.3% AAI) are encoded upstream of the other biosynthetic genes (Figure 4.6).

Linear azole containing peptides (LAPs) contain heterocyclic rings of thiazole and (methyl)oxazole (Haft, 2009). With the exception of *P. thermantarcticus* DSM 9572 and *S. thermophilus* DSM 4749, all examined taxa incorporate a four gene LAP biosynthetic locus, coding for a cyclohydratase (*sagC*), a maturase (*sagD*) and a dehydrogenase (*sagB*) as well as a 74-113 amino acid bacteriocin peptide (71.5% AAI among the compared taxa) belonging to the heterocycloanthracin/sonorensin family (TIGR03601). Heterocycloanthracin was identified in *Bacillus cereus* and

Bacillus anthracis (Haft, 2009) and sonorensin from a marine *Bacillus sonorensis* isolate (Chora *et al.*, 2014). Sonorensin has been shown to effective against *Listeria monocytogenes* and *Staphylococcus aureus*, with anti-biofilm activity for the latter pathogen and could be used as a food biopreservative (Chopra *et al.*, 2014; Chopra *et al.*, 2015). *P. thermantarcticus* DSM 9572 and *Parageobacillus* genomosp. A W-2, incorporate a locus coding for a predicted sactibiotic (Figure 4.6). Sactipeptides incorporate post-translational modifications with intramolecular bridges of cysteine sulphur to α-carbon linkages (Mathur *et al.*, 2015). The identical 49 aa peptide in the *Parageobacillus* strains shares 67.4% AAI with the huazacin peptide in *Bacillus thuringiensis* serovar *huazongensis* BGSC 4BD1 (EEM79974.1), which shows activity against the food-borne pathogen *L. monocytogenes* (Hudson *et al.*, 2019).

Two distinct class II bacteriocin loci were also identified among the studied taxa. A 123 aa peptide (98.0% AAI) present in 15/24 *Parageobacillus* taxa (all *P. thermoglucosidasius* and *P. toebii* WCH70), but absent from all *Saccharococcus* strains is predicted to belong to the lactococcin 972 family (pfam09683), produced by *Lactococcus lactis* and active against closely related organisms (Martínez *et al.*, 1999). The second locus encodes a 48 aa peptide and is found on the plasmids of four *P. thermoglucosidasius* strains (100% AAI). It is predicted to belong to the aureocin A53 family (NF033881), which is produced by *S. aureus* and is active against *L. monocytogenes* (Netz *et al.*, 2002).

Another potential group of antimicrobials are lactonases, which degrade or quench *N*-acyl-homoserine lactones (AHLs) that serve as chemical signalling molecules in Gram-negative pathogens and thereby inhibit AHL-regulated functions such as the production of virulence factors and biofilms (Bergonzi *et al.*, 2016). One such lactonase, GcL (WP_017434252.1) was identified in *S. caldoxylosilyticus* DSM 14590^{T} (Bergonzi *et al.*, 2016). Orthologues sharing 96.4% AAI are found in all *Parageobacillus* and five *S. caldoxylosilyticus* strains. A second predicted *N*-acyl-homoserine lactonase is found in all 35 comparator taxa and these share 68.2% AAI with the quorum quenching lactonase YntP of *B. subtilis* 168 (O34760.2). The latter lactonase inhibits streptomycin production in *Streptomyces griseus* (Schneider *et al.*, 2012). Furthermore, orthologues (78.3% AAI) of a broad-substrate *N*-acyl-homoserine lactonase from *G. kaustophilus* HTA426 (GKL – 3OJG; Chow *et al.*,

2010) are encoded on 11/11 *Saccharococcus* genomes, as well as those of *P. thermantarcticus* DSM 9572^{T} and *P. toebii* DSM 14590^{T}. Given the increasing prevalence of antimicrobial resistance, thermostable *N*-acyl-homoserine lactonases produced by *Parageobacillus* and *Saccharococcus* should receive additional attention.

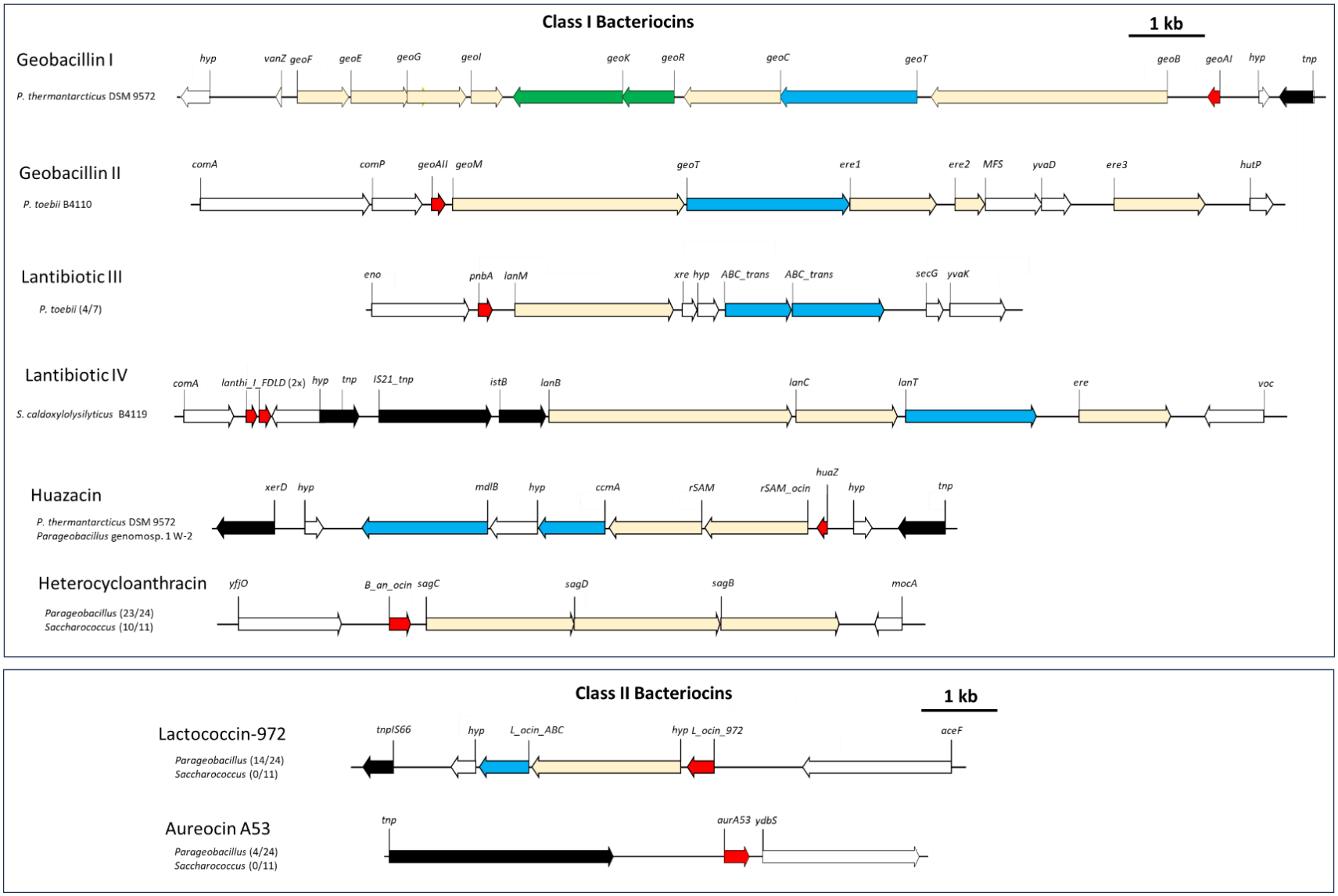


Figure 4.6: *Parageobacillus* and *Saccharococcus* antimicrobial biosynthetic loci. Schematic diagrams of the bacteriocin I and II loci present on the genomes of select *Parageobacillus* and *Saccharococcus* taxa. Genes coding for the active bacteriocin peptide are coloured in red, while those coding for accessory biosynthetic proteins, regulation and transport are shown in light yellow, green and blue, respectively. Genes coding for proteins with a potential role in self-immunity are indicated by orange arrows. A 1 kb scale bar is shown.

4.3.4.2 *Parageobacillus* and *Saccharococcus* as sources of bioindustrially relevant enzymes

With a projected market share of $ 16.9 billion by 2027 (Markets and Markets, 2022), enzymes and in particular their thermostable counterparts, form a cornerstone of a broad range of industries, including the production of food, detergents, textiles and bioenergy (Atalah *et al.*, 2019; Kumar *et al.*, 2019). Using a range of *in silico* tools, the *Parageobacillus* and *Saccharococcus* genomes were screened for thermostable enzymes of potential biotechnological value.

Carbohydrate-active enzymes

Bacteria produce a range of carbohydrate-active enzymes (CAZymes) to degrade complex carbohydrate polymers into monomeric sugars, which from a biotechnological perspective can be further fermented into biofuels and a broad range of value-added chemicals (Chettri *et al.*, 2020; Contesini *et al.*, 2021). A total of 2,130 CAZymes were predicted across the 35 compared genomes (average 61 CAZymes/genome; Appendix C Supp. Table S4.3). These were predominated by glycoside hydrolases (GH: 44.6%) that hydrolyse or rearrange glycosidic bonds in carbohydrate chains, glycosyltransferases (GT: 43.0%) that form bonds in carbohydrate chains, and carbohydrate esterases (CE: 11.5%) that deacetylate ester-substituted carbohydrates (Chettri *et al*., 2020). Biotechnological focus has been centred on GH and CE classes, as well as less represented polysaccharide lyases (PL: only presented on 4/35 genomes) that catalyse the non-hydrolytic cleavage of glycosidic bonds in carbohydrate chains (Appendix C Supp. Table S4.3; Chettri *et al*., 2020). A total of 930 GHs were identified on the 35 compared genomes, with 57 of these (6%) predicted to be extracellularly secreted. Substantially greater numbers of GHs are encoded on the genomes of members of the genus *Saccharococcus* (average GHs: 34.5/genome) than those of *Parageobacillus* (average GHs: 23.8/genome). This could largely be attributed to several strains of *S. caldoxylosilyticus*, in particular KH1-5 and KH1-6 which both code for 44 GHs (Appendix C Supp. Table S4.3).

GHs are further classified into 186 GH families (Drula *et al.*, 2022), each with their own hydrolytic mechanism and/or substrate. The *Parageobacillus* and *Saccharococcus* GHs cover 33 distinct GH families, eight of which are predicted to be secreted extracellularly. Of these families, two are uniquely represented in the genus *Parageobacillus*, while five families are restricted to members of the genus *Saccharococcus*. Between seven (*P. toebii* NEB718 and *S. thermophilus* DSM 4749) and twenty-six (*S. caldoxylosilyticus* DSM 12041 and KH3-5) of the 33 GH families are encoded on each individual strain genome, with only three GH families, namely GH13, GH18 and GH23 being core to all 35 compared taxa (Appendix C Supp. Table S4.3). The latter two families are involved in peptidoglycan hydrolysis and play a role in spore germination (Kodama *et al.*, 1999) and cell wall remodelling and recycling (Byun *et al.*, 2018), respectively. The GH13 α-amylase family, which degrades starch and its derivatives (e.g. amylopectin and pullulan; Stam *et al.*, 2006), is the most broadly represented of all GH families among the *Parageobacillus* and *Saccharococcus* taxa, with 260 members across the 35 genomes. Being the major storage carbohydrate of terrestrial plants, starch degrading enzymes are of value in the food, fermentation and pharmaceutical industries, in particular the thermostable variants as produced by *Geobacillus* and *Parageobacillus* species (Novik *et al.*, 2019; Najar and Thakur, 2020).

The majority of GHs encoded on the *Parageobacillus* and *Saccharococcus* genomes are involved in the degradation of lignocellulosic biomass. Lignocellulose, comprised of cellulose, hemicellulose, lignin and minor fractions of lipids, proteins, pectin and soluble sugars, forms the predominant component of plant biomass and is one of the most abundant renewable substrates on Earth (Chettri *et al.*, 2020; Bhalla *et al.*, 2022). In geobacilli, plant biomass degradation activity can be linked to the large, highly variable Hemicellulose Utilization System (HUS) locus, which incorporates hydrolytic enzymes, sugar transport systems and carboxylesterases to completely degrade and utilise the xylose backbone, arabinose, galactose and glucuronic acid side chains and methyl or acetyl group decorations (De Maayer *et al.*, 2014b). Highly variable HUS loci were found in 14/35 of the compared taxa, which could be further subdivided into five types (I-V; Figure 4.7). Type I and II are restricted to *P. thermoglucosidasius* and *P. thermantarcticus* DSM

9572 (Type I) and *P. thermoglucosidasius* only (Type II) and target xylans decorated with glucuronic acid and arabinofuranose side chains, respectively. Unique to the Type I HUS loci is a gene coding for a GH5 endoglucanase, indicating that these taxa may also target the cellulose component of biomass. Type III HUS loci were found on the genomes of the two *Parageobacillus* genomosp. A isolates and three *S. caldoxylosilyticus* strains, and are predicted to target arabinose and glucuronic acid-containing xylans. The Type IV HUS locus, unique to *S. caldoxylosilyticus* VR-IP, likely also targets this hemicellulose substrate, but further incorporates genes coding for enzymes for the hydrolysis and metabolism of galactose (GH36), mannose (GH38_1, GH38_2 and GH38_3), *N*-acetylglucosamine (GH84) and fructofuranose (GH100; Drula *et al.*, 2022), suggesting this strain can degrade more complex plant biomass substrates. Finally, the Type V HUS locus of *S. caldoxylosilyticus* KH1-5 and KH1-6 encodes the cellular machinery for the degradation of rhamnogalacturonan I, with pathways for the degradation of the backbone as well as arabinan and glucuronic acid side chains. This pectic polymer forms a major part of the primary cell wall and middle lamella of most higher plants (Kaczmarska *et al.*, 2022).

The propensity of *Parageobacillus* and *Saccharococcus* taxa to degrade distinct and variously decorated plant biomass constituents offers excellent opportunities for biocomposting of plant biomass, potentially as mixed cultures (Sarkar *et al.*, 2010; Wang *et al.*, 2020), or the production of value-added products such as oligosaccharides that could be used as prebiotics or food additives (Placier *et al.*, 2009). One component of plant biomass that affects the efficacy of enzymatic degradation is lignin. A lignin degrading laccase has been identified in *Geobacillus* sp. WSUCF1 (WP_011230630.1; Rai *et al.*, 2019). Orthologues of this laccase are encoded on the genomes all 35 of the studied taxa (61.4% AAI), suggesting that they further incorporate the machinery to degrade this plant biomass constituent.

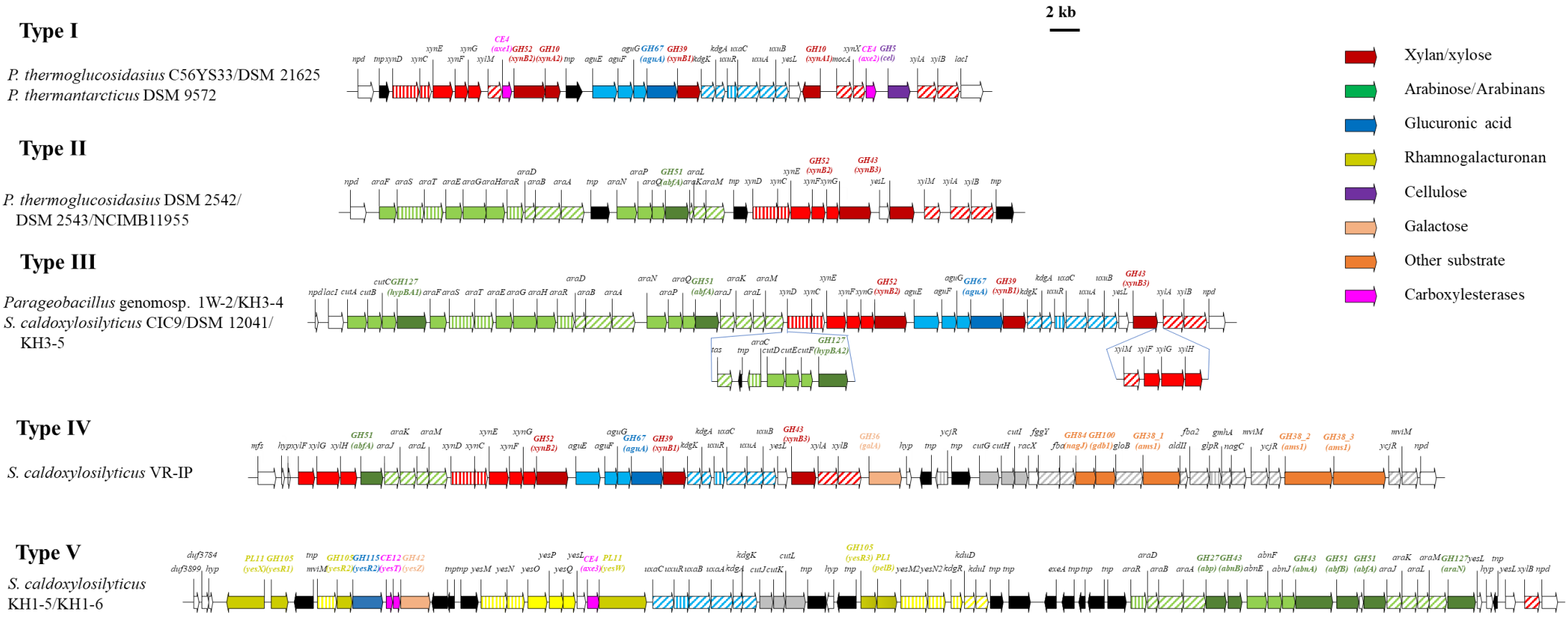


Figure 4.7: *Parageobacillus* and *Saccharococcus* Hemicellulose Utilisation (HUS) loci. Schematic diagrams of the Type I-V hemicellulose utilisation loci on the genomes of select *Parageobacillus* and *Saccharococcus* taxa. Arrows corresponding to genes are coloured according to their predicted carbohydrate target. Grey arrows represent those genes coding for proteins of unrelated or unknown functions, while black arrows represent predicted transposase-associated elements. Lighter-shared arrows represent genes involved in carbohydrate transport, while cross-hatched arrows are those genes with predicted regulatory roles in carbohydrate metabolism. A 2 kb scale bar is shown.

Lipases, carboxylesterases and proteases

Thermostable lipases and carboxylesterases are of growing interest in the food, pharmaceutical and fine-chemical industries, where their products of hydrolysis can be used for the synthesis of various chemicals (Hussein *et al.*, 2015; Novik *et al.*, 2019). Where lipases degrade water-insoluble long chain triglycerides, carboxylesterases hydrolyse ester bonds in shorter chain acyl derivatives (Novik *et al.*, 2019). Comparison of the *Parageobacillus* and *Saccharococcus* proteomes against the Lipase Engineering Database (LED; Fischer and Pleiss, 2003) identified orthologues for twenty-four distinct homologous family groups (Appendix C Supp. Table S4.3). Of these, thirteen constituted alpha/beta hydrolases (abhydrolases – cl 21494) for which no clear substrate/activity could be identified, while five distinct acetyl esterases are predicted to contribute to the removal of acetyl groups from lignocellulosic components (xylan and rhamnogalacturonan). Three distinct carboxylesterases are encoded on the genomes. *p*-Nitrobenzyl esterases need to be removed from oral beta-lactam antibiotics for their final synthesis, and the *p*-nitrobenzyl esterase (PbnA) of *B. subtilis* is effective in this activity. Orthologues of this enzyme (P37967.2; 44.2% AAI) are present in 11/11 *Saccharococcus* strains and *P. thermantarcticus* DSM 9572. Orthologues of two characterised carboxylesterases from *Geobacillus stearothermophilus* (Est30; Pdb = 1TQH; 90% AAI) and *G. thermodenitrificans* CMB-A2 (EstGtA2; AEN92268.1; 72% AAI) are present in all 35 analysed taxa. Both of these thermostable enzymes show activity against *p*-nitrophenyl esters of different chain length (Montoro-García *et al.*, 2009; Charbonneau *et al.*, 2010). All 35 *Parageobacillus* and *Saccharococcus* taxa also encode orthologues of a lysophospholipase (YpA; COG 2267) as well as two distinct copies of GDSL-like lipases (pfam 13472). However, the target triglycerides would need to be determined.

Microbial proteases and peptidases, in particular their thermostable counterparts, have a broad range of applications including the treatment of leather, as additives in detergents and in the food industry (Hussein *et al.*, 2015; Novik *et al.*, 2019). Comparison of the proteome datasets against the MEROPS database (Rawlings *et al.*, 2018) identified 4,765 distinct protein orthologues encoded on the 35 genomes.

On average, slightly more (138) are encoded on the *Saccharococcus* genomes than on the *Parageobacillus* genomes (135), while 130 are encoded on the genome of *G. thermodenitrificans* DSM 465. The highest number of proteases/peptidases are encoded on the genome of *S. caldoxylosilyticus* B4119 (152; Appendix C Supp. Table S4.3). The proteases/peptidases belong to 40 and 66 distinct MEROPS clans and families, respectively, with the highest numbers of families represented in *Saccharococcus* genomosp. A NUB3621 (60). The proteases/peptidases can be subdivided into 212 orthogroups, 91 of which (43%) are core to all compared taxa, while 33 (16%) occur only in a single taxon. A total of 40 and 21 protease/peptidase orthogroups are unique to either the genus *Parageobacillus* or *Saccharococcus*, respectively. Only a small proportion (23/212) of the protease/peptidase orthogroups are secreted extracellularly, with six each of these unique to *Parageobacillus* and *Saccharococcus* taxa, respectively.

Parageobacillus genomosp. nov. A KH3-4 and W-2 as well as *Saccharococcus* genomosp. nov. A NUB3621 (two copies) produce a predicted neutral thermolysin protease sharing 68.2% AAI (range 51.6-82.7%) with thermolysin from *G. stearothermophilus* (P43133.1). The latter protease NprS is commercially used to produce precursors for the artificial sweetener aspartame (Ke *et al.*, 2013). Serine proteases, particularly those of the subtilisin superfamily (S8), have a broad range of applications in the food, cosmetics and detergent industries, and in the treatment of sewage (Jang *et al.*, 1992; Chang *et al.*, 2021; Falkenberg *et al.*, 2023). A total of 171 S8 family proteases are encoded across the *Parageobacillus/Saccharococcus* genomes, belonging to 16 distinct orthogroups (12/16 extracellularly secreted). Orthogroups of five and one subtilisin protease are unique to single strains of *Parageobacillus* and *Saccharococcus*, respectively, while a further three orthogroups are represented in *Parageobacillus* species only. While the S8 proteases in these taxa share between 27.1 and 43.6% AAI with subtilisin J of *G. stearothermophilus* NCIMB 10278 (P29142.1; 27), the S8 protease orthogroups in this study share < 50% AAI among them, indicating a broad underexplored set of proteases of potential biotechnological value among the genera *Parageobacillus* and *Saccharococcus*.

4.3.4.3 Enzymes for the molecular laboratory

DNA polymerases

Thermostable DNA active enzymes encompass an expanding toolkit for numerous conventional molecular biotechnology applications, including PCR, genetic engineering, DNA sequencing, diagnostics and synthetic biology (Hamilton *et al.*, 2001; Terpe, 2013; Loenen *et al.*, 2014; Najar and Thakur, 2020). Several thermostable DNA polymerases have been derived and commercially developed from *Geobacillus* spp., most notably the Bst DNA polymerase, a family A DNA polymerase I with 5'-3' exonuclease activity isolated from *G. stearothermophillus* GIM1.543 (Aliotta *et al.*, 1996; Sandalli *et al.*, 2009; Oscorbin *et al.*, 2015; Ma *et al.*, 2016; Oscorbin and Filipenko, 2023). All strains analysed possessed DNA polymerases families A (DNA PolI), C (DNA PolIII - α, τγ, δ, δ′ and β subunits), Y (DNA PolIV) and X (DNA PolX) represented by one orthogroup each (Appendix C Supp. Table S4.4). In addition, a putative DNA polymerase family B (PolB) orthologue, is encoded on the genomes of *P. thermantarcticus* DSM 9572^{T} and *P. thermoglucosidasius* DSM 21625. These proteins show some homology (17.1% AAI) with the *Escherichia coli* PolII enzyme (PDB: 1Q8I) which is implicated in translesion and mutagenic DNA synthesis (Wang and Yang, 2009; Kazlauskas *et al.*, 2020). In addition to the DinB DNA polymerase IV orthologues (74.2% AAI; range 52.0-100%) encoded by all strains analysed, two *P. toebii* and eight *S. caldoxylosilyticus* strains encode putative UmuC DNA polymerase family Y (DNA PolV) orthologues (81.8% AAI) involved in UV-dependent and chemically-induced mutagenesis (Bumgardner *et al.*, 2015; Timinskas and Venclovas, 2019). These polymerases may have application in inducing random mutagenesis for the purpose of directed evolution (Al Mamun, 2007; Labrou, 2010)

Restriction modification and CRISPR/Cas systems

Thermostable restriction enzymes and their associated modification (RM) systems are used in various generic engineering strategies, sequencing and diagnostics (Sharma *et al.*, 2013). Comparison to the REBASE database (Roberts *et al.*, 2015) identified 61 orthogroups incorporating restriction-modification (RM) components

(Appendix C Supp. Table S4.5). These included twenty-seven Type I, eighteen Type II, nine Type III and six Type IV putative RM components. Most (59/61) of the identified RM components are encoded on the genomes of *Parageobacillus* spp., 43 of which are unique to the genus. Of these, twelve and twenty-three are specific to *P. thermoglucosidasius* and *P. toebii*, respectively. *Saccharococcus* genomes only encode 19/61 of the RM components, three of which are unique to *S. caldoxylosilyticus* strains. On average, ~7 and 3 RM components are encoded on the genomes of *Parageobacillus* and *Saccharococcus*, respectively, suggesting they, and in particular the former genus, represent a rich source for novel thermostable RM enzymes.

In addition to the native role CRISPR/Cas systems play in preventing foreign plasmid and nucleic acid transfer in prokaryotes (Barrangou *et al.*, 2007; Pougach *et al.*, 2010; Westra *et al.*, 2013), modified CRISPR/Cas systems have also been employed in various biotechnological and biomedical applications through targeted genome editing and gene regulation (Sander and Joung, 2014; Ding *et al.*, 2020; Wang *et al.*, 2022). Recently, several *Geobacillus* Cas proteins have also received attention due to their thermostability and greater specificities when compared to the more frequently utilised mesophilic Cas9 systems (Harrington *et al.*, 2017; Mougiakos *et al.*, 2017; Trasanidou *et al.*, 2023). Using CRISPRCasFinder (Grissa *et al.*, 2007; Abby *et al.*, 2014; Couvin *et al.*, 2018), 34 distinct orthogroups were identified as Cas proteins of type I and type III CRISPR-Cas systems (Appendix C Supp. Table S4.6). Substantially more Cas proteins were identified in *P. thermoglucosidasius* (average Cas proteins: 16.96/genome) compared to *Saccharococcus* spp. (average Cas proteins: 6.1/ genome).

4.3.4.4 Whole-cell biotechnological applications for *Parageobacillus* and *Saccharococcus*

<u>*Parageobacillus* potential in bioremediation</u>

Aside from the biotechnological potential of their enzymes, there has also been extensive interest in whole cell biocatalysis with thermophilic Geobacilli (Hussein *et al*., 2015). Numerous *Geobacillus* (and *Parageobacillus*) strains have been

investigated for their applicability towards various bioremediation applications, including degradation of xenobiotics, phenols and in particular long chain- and aromatic-hydrocarbons and petroleum hydrocarbons (Orellana *et al.*, 2018; Elumalai *et al.*, 2019; Lin *et al.*, 2019; Novik *et al.*, 2019).

Analysis of the comparator protein dataset identified 84 distinct orthogroups associated with degradation of various xenobiotic compounds (Figure 4.8; Appendix C Supp. Table S4.7). The genomes of *P. thermoglucosidasius* taxa typically encode substantially more orthologues (61.4/genome) than either *Saccharococcus* spp. (43.6/genome) or *P. toebii* spp. (28.9/genome). The highest number of proteins involved in xenobiotic degradation occur in *P. thermoglucosidasius* 23.6 (69). Phenol *meta*-cleavage pathway degradation loci (a twelve gene chromosomal and ten gene plasmid locus) have previously been identified in the genus *Parageobacillus* (Aliyu *et al.*, 2021; Sun *et al.*, 2023). The full chromosomal phenol degradation operon is conserved among all *P. thermoglucosidasius* strains, 2/7 *P. toebii*, both *Parageobacillus* genomosp. A strains and 5/8 *S. caldoxylosilyticus* strains (Appendix C Supp. Table S4.7). Furthermore, 8/14 *P. thermoglucosidasius* strains carry the complete plasmid-bound locus.

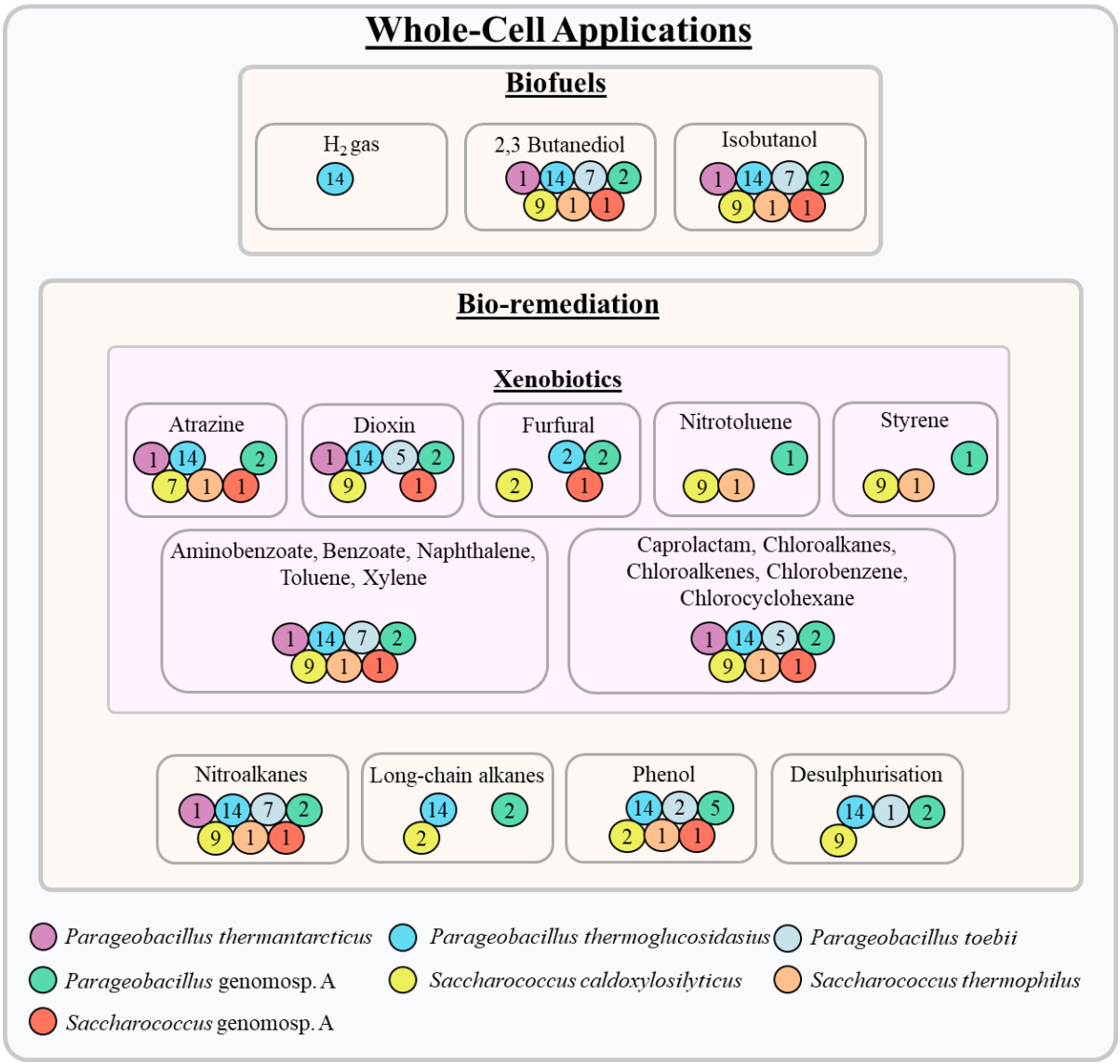


Figure 4.8: Schematic diagram showing potential whole-cell applications of the genera *Parageobacillus* and *Saccharococcus*. Coloured balls represent species containing at least one gene orthologue or pathway component for each respective system.

Crude and refined petroleum fractions may contain or release (upon combustion) high levels of organosulphur compounds, which are resistant to degradation and hazardous to the environment (Shi and Wu, 2021). Consequently, means of reducing levels of organosulfur compounds either preventatively in processed petroleum products or in the remediation of polluted systems, is desirable. Various thermophilic taxa have been observed capable of catabolising sulphur-rich

petroleum compounds via several mechanisms (Peng *et al.*, 2019; Peng *et al.*, 2023; Kalita *et al.*, 2022). Thirty distinct orthogroups were associated with sulphur metabolism by eggnog-mapper (Supp. Table 4.7; Cantalapiedra *et al.*, 2021). These included nine orthogroups similar (>86.84% AAI) to previously identified sulphonate transporters of *Parageobacillus* genomosp. A W-2, that were conserved amongst most taxa (Supp. Table 4.7; Peng *et al.*, 2019). Three distinct desulphurization-associated gene clusters (1, 2 and 3; Appendix C Supp. Table S4.7), incorporating distinct monooxygenases, have previously been described in *Parageobacillus thermoglucosidasius* (Peng *et al.*, 2019; Peng *et al.*, 2023). The genomes of all *P. thermoglucosidasius* and two *S. caldoxylosilyticus* strains incorporate all three complete desulphurisation clusters, while those of 7/9 *S. caldoxylosilyticus* strains and the two *Parageobacillus* genomosp. A isolates incorporate complete desulphurisation clusters 2 and 3 (Figure 4.8; Appendix C Supp. Table S4.7). *S. caldoxylosilyticus* VR-IP and *P. toebii* WCH70 harboured only a complete desulphurisation cluster 3, while none were observed in the other *P. toebii* strains (Figure 4.8; Appendix C Supp. Table S4.7).

Long-chain alkanes form a major component of crude oils. Several studies have identified the presence and activity of genes associated with variable length long-chain alkane catabolism in *Geobacillus* and *Parageobacillus* taxa (Feng *et al.*, 2007; Boonmak *et al.*, 2014; Tourova *et al.*, 2016; Tourova *et al.*, 2018). Orthologues of LadAα (ART30136: 66.62-67.69% AAI range) and LadAβ (ART30139: 70.85-71.86% AAI range) and LadB (ART30142: 61.81-66.54% AAI range) that contribute to C_{10}-C_{30} *n*-alkane utilisation in *P. toebii* B1024 (Tourova *et al.*, 2016) are encoded on the genomes of all *P. thermoglucosidasius* strains analysed, *Saccharococcus* genomosp. A KH3-5, *S. caldoxylosilyticus* ER4B. Both LadA orthologues, but no LadB orthologues are present in *Parageobacillus* genomosp. A KH3-4 and W-2 (Figure 4.8; Appendix C Supp. Table S4.7). In accompaniment, at least one putative aldehyde dehydrogenase (ABO68462; 78.89-94.52% AAI range) and alcohol dehydrogenase orthologues (ABO66657, ABO67118 and ABO68223; 76.77-79.42%, 79.76-86.10% and 86.7-91.3% AAI ranges, respectively), assumed to participate in LadA initiated metabolism of long-chain alkanes in *G. thermodenitrificans* NG80-2 (Feng *et al.*, 2007), were detected

across all strains except *S. thermophilus* DSM 4749, which did not encode orthologues of ABO68462 or ABO67118.

Nitroalkanes are another group of highly recalcitrant compounds, utilized as fuels, solvents, herbicides and pesticides which are also toxic and carcinogenic (Fiala *et al.*, 1989; Li *et al.*, 2011; Sun *et al.*, 2019). Given the potential hazards of these compounds and their release into the environment, control strategies to mitigate their effects are a necessity. Recently, three nitroalkane oxidizing enzymes (WP_064553126, WP_064551563 WP_064551165) were shown to variably degrade nitropropane and nitroethane in *Parageobacillus* genomosp. A W-2 (Sun *et al.*, 2019). Orthologues of each enzyme (Gt2929, Gt1378 and Gt1208; 55.20-100%, 76.96-100% and 88.73-100% AAI ranges, respectively) are encoded on the genomes of 34, 33 and all 35 of the analysed strains, respectively (Figure 4.8; Appendix C Supp. Table S4.7).

Members of the genera *Geobacillus* and *Parageobacillus* have also been associated with resistance to a range of heavy metals, that unlike organic xenobiotics cannot be completely degraded (Poli *et al.*, 2006; Cuebas *et al.*, 2011; Ben Fekih *et al.*, 2018; Giovanella *et al.*, 2020; Puopolo *et al.*, 2020). Comparison of the proteome datasets against the BacMet predicted heavy metal resistance gene database identified thirty-one unique orthologue groups similar to twenty-eight genes implicated in resistance to eighteen heavy metals (Figure 4.8; Appendix C Supp. Table S4.7), predominantly Arsenic (5), Copper (5), Iron (4), and Mercury (4; Pal *et al.*, 2014). Notably, *mdrL*/*yfmO* and *corA* genes associated with Cobalt, Chromium, Magnesium, Nickel resistance, were absent in *P. thermoglucosidasius* and *P. toebii*, however were present in 15/21 and 12/21 of the remaining strains (Appendix C Supp. Table S4.7). Conversely, genes associated with Mercury resistance were less widely distributed, only present in 3/7 *P. toebii*, *P. thermantarcticus* DSM 9572^{T}, *P. thermoglucosidasius* DSM 6285 and 2/9 *S. caldoxylosilyticus* (Appendix C Supp. Table S4.7).

4.3.4.5 *Parageobacillus* as a producer of green energy

In part due to its capacity for biomass degradation, as well as its fermentation pathways, *P. thermoglucosidasius* has received extensive interest for the production of biofuels (Figure 4.8). In particular, ethanol production has been widely researched, but as a mixed acid fermenter with limited ethanol tolerance, metabolic engineering of this species is required (Tang *et al.*, 2008; Cripps *et al.*, 2009).

Another *P. thermoglucosidasius* fermentation product of biotechnological interest is isobutanol, which can serve as biofuel, fuel additive or as a primer for the production of chemicals (Lin *et al.*, 2014; Novik *et al.*, 2019). The final step in isobutanol formation from isobutyraldehyde involves an isobutayraldehyde dehydrogenase, with two putative enzymes (AdhA and Geoth_3823) identified in *P. thermoglucosidasius* C56YS93 (Lin *et al.*, 2014). Orthologues of both enzymes, sharing 95.7% and 97.3%, respectively, are encoded on the genomes of all 35 comparator strains (Figure 4.8; Appendix C Supp. Table S4.7), suggesting members of both genera could serve as targets for metabolic engineering for isobutanol production. *P. thermoglucosidasius* also produces 2,3-butanediol (2,3-BDO), which can be used as liquid fuel, fuel additive or chemically modified to produce high octane isomers for use in aviation fuels (Novik *et al.*, 2019; Sheng *et al.*, 2023). Orthologues of one key enzyme involved in 2,3-BDO synthesis identified in *P. thermoglucosidasius* NCIMB 11955 (Sheng *et al.*, 2023), namely acetolactate synthase (ALS), were observed in all compared *Parageobacillus* and *Saccharococcus* taxa (93.1% AAI; Figure 4.8; Supp. Table 4.7). The final enzyme in 2,3-BDO synthesis, butanediol dehydrogenase (BDH), was restricted to the genus *Parageobacillus* (97.9%), with a single copy encoded on the genome of 23/24 taxa, with the exception of *P. toebii* WCH70, where two copies exist (96.6% amino acid identity between copies). The middle enzyme involved in the conversion of acetoin to 2,3-BDO, acetolactate dehydrogenase (ALD), is rare in thermophiles (Sheng *et al.*, 2023) and no orthologues were identified in the taxa compared in this study. However, candidate heterologous ALD enzymes have been identified for metabolic engineering for 2,3-BDO synthesis (Sheng *et al.*, 2023).

Recent interest has focused on the production of hydrogen gas, an environmentally friendly and sustainable alternative energy carrier, from carbon monoxide-containing waste gases by *P. thermoglucosidasius* (Mohr *et al.*, 2018a). This biological water-gas shift reaction (WGS) involves an enzyme complex comprising a carbon-monoxide dehydrogenase (CODH) and hydrogen-evolving hydrogenase (Chapter 1 Section 1.6; Mohr *et al.*, 2018a; Mohr *et al.*, 2018b). Interest in *P. thermoglucosidasius* is further driven by the fact that the CODH-hydrogenase combination, is found almost exclusively in anaerobic bacterial taxa, while *P. thermoglucosidasius* is a facultative anaerobe, implying that biomass can be generated aerobically prior to a shift to anaerobic hydrogenogenesis via the WGS reaction (Mohr *et al.*, 2018a; Mohr *et al.*, 2018b). Previous analyses showed the CODH-hydrogenase locus to be restricted to *P. thermoglucosidasius*. Analysis of our annotated dataset showed homologous loci in all fourteen *P. thermoglucosidasius* taxa, with the CODH proteins CooCSF and hydrogenase proteins PhcABCDEFGHIJKL sharing 99.5% and 99.1% AAI, respectively among these taxa, while no orthologues were found in any other *Parageobacillus* or *Saccharococcus* taxa. However, a recent study identified the CODH-hydrogenase locus on the genome of *Parageobacillus* sp. G301 (Imaura *et al.*, 2023). ANI and dDDH values (97.24% and 76.3%, respectively with *P. toebii* DSM 14590^{T}) indicate that this strain belongs to the species *P. toebii*, and the CODH and hydrogenase proteins share an average AAI of 91.5% and 91.2% with those of the fourteen *P. thermoglucosidasius* taxa. As such, broader evaluation of the genera *Parageobacillus* and *Saccharococcus* for hydrogen-evolving systems of potential biotechnological value is warranted.

2.4 Conclusions

In contrast to previous studies, this phylogenomic analysis delineates *Parageobacillus* and *Saccharococcus* as two distinct genera, both of which present open pan-genomes. *P. toebii* in particular presents the greatest potential for novel gene accrual within *Parageobacillus*. Plasmids, bacteriophages and transposable elements are key drivers of genomic and functional, diversification of these genera. Both *Parageobacillus* and *Saccharococcus* harbour a wealth of biotechnological potential including potential novel antimicrobials and a range of thermostable enzymes. Functional and *in vivo* analyses of the novel antimicrobial peptides should serve to validate the potential of the studied taxa to contribute towards combatting antibiotic-resistant target bacteria. Similarly, the broad range of carbohydrate-, protein- and lipid-active enzymes, identified here and in previous studies, should be evaluated to expand the current repertoire of thermostable enzymes for a wide array of biotechnological applications. Our analyses have also further highlighted the potential for members of both *Parageobacillus* and *Saccharococcus* in a broad spectrum of whole-cell applications, including bioremediation of various xenobiotic compounds and environmental pollutants, the degradation of lignocellulosic biomass to generate various value-added products, as well as the use of these taxa to contribute towards the green energy market. Given the extensive genomic variability and the potential biotechnological pathways and enzyme complement, additional discovery and characterization, both genomic and functional, of novel *Parageobacillus* and *Saccharococcus* isolates will continue to expand the biotechnological toolkit of these intriguing genera.

SUMMARY

H_2 gas is an important industrial reagent and an increasingly prominent alternative energy carrier. The global shift towards utilising renewable energy sources, such as H_2 gas, has highlighted the need for environmentally friendly H_2 gas production processes. Presently, the majority of H_2 gas is produced almost exclusively through non-renewable and environmentally harmful practices. Consequently, ongoing research is focused on developing more sustainable H_2 gas production methods. One promising approach involves the use of thermophilic bacteria *P. thermoglucosidasius*, to produce H_2 gas from CO rich gases via the biologically mediated WGS reaction. The hydrogenogenic potential of *P. thermoglucosidasius* has undergone preliminary characterisation, although various facets concerning its broader H_2 gas metabolism and applicability towards industrial H_2 production remain unexplored.

Attempts to elucidate the influence of two putative group 1d and 2a [NiFe]-uptake hydrogenases on the H_2 metabolism of *P. thermoglucosidasius*, through knockout mutagenesis were undertaken. Generation of strains bearing stable knock-out mutants of the large catalytic subunits of these hydrogenases was not achieved. Colonies that harboured either chromosomally-integrated or plasmid-bound mutant copies of the target genes, in tandem with the wild type genes exhibited reduced growth as long as the mutant gene copies were detected. Following loss of the mutant genes, growth was comparable to the wild type, suggesting that disruption of the genes coding for these hydrogenases negatively impacts cell viability.

To feasibly apply the *P.* thermoglucosidasius hydrogenogenic system in an industrial context, necessitates the use of commercially/industrially relevant, highly variable waste/synthesis gases. The cultivation of *P. thermoglucosidasius* under anaerobic synthetic syngas atmospheres led to faster hydrogen production compared to aerobic conditions, however, at the expense of biomass accumulation. Supplementation of an initially aerobic fermentation with 5% O_2 at various stages hindered H_2 gas production, but supported growth. Supplementation with 1% O_2 during peak hydrogenogenesis minimally influenced H_2 gas production, while still supporting growth, although increasing volume O_2 additions demonstrated more

greater disruptions to H_2 production, greater degrees of H_2 consumption and poorer final H_2 yields. Overall, the system showed some tolerance to O_2, rapidly resuming H_2 production following O_2 depletion.

Beyond H_2 gas production, *P. thermoglucosidasius* and closely related thermophilic taxa represent appealing targets for further biotechnological developments. Phylogenomic and pan-genome analysis identified *Parageobacillus* and *Saccharococcus* as distinct genera with open pan-genomes. *P. toebii* in particular, shows the strongest potential for novel gene acquisition within *Parageobacillus*. Genomic and functional diversification in these genera is driven by plasmids, bacteriophages, and transposable elements. Both genera have distinct biotechnological potential, including potential novel antimicrobials and thermostable enzymes. Furthermore, both genera show potential for bioremediation, biomass degradation, and green energy applications.

APPENDIX A

Luria-Bertani medium (LB):

A 1 L volume of LB contained 10 g casein hydrolysate, 5 g yeast extract, 5 g NaCl, 1.25 mL 10% (w/v) NaOH, pH adjusted to 7.3, with 15 g agar bacteriological added before autoclaving when solid medium was required.

Modified Luria-Bertani medium (mLB):

A 1 L volume of mLB contained 10 g casein hydrolysate, 5 g yeast extract, 5 g NaCl, 10 g 1.25 mL NaOH, pH adjusted to pH 7.3 and 1 mL of each of filter-sterilized stock solutions of 1.05 M nitrilotriacetic acid, 0.59 M $MgSO_4 \cdot 7H_2O$, 0.91 M $CaCl_2 \cdot 2H_2O$ and 0.04 M $FeSO_4 \cdot 7H_2O$ with 15 g agar bacteriological when solid medium was required.

TGP medium:

A 1 L volume of TGP contained 17 g casein hydrolysate, 3 g soy peptone 2.5 g K_2HPO_4, 5 g NaCl, pH 7.3, with 4 g sodium pyruvate and 4 mL glycerol added after autoclaving as a filter-sterilized 10x concentrate and 15 g agar bacteriological when solid medium was required.

2TY medium:

A 1 L volume of 2TY contained 16 g casein hydrolysate, 10 g yeast extract, 5 g NaCl, pH 7.3 and 15 g agar bacteriological when solid medium was required.

***E. coli* SOC medium:**

A 1 L volume of medium contained 20 g casein hydrolysate, 5 g yeast extract, 0.58 g KCl, 4.8 g MgSO4 and 3.6 g D-glucose at, pH 7.

Phosphate buffered saline:

A 1 L volume contained 8 g NaCl, 0.2 g KCl, 1.42 g Na_2HPO_4, 0.24 g KH_2PO_4.

SPYNG medium:

A 1 L volume of medium contained 16 g soy peptone, 10 g yeast extract, 5 g NaCl, pH 7.3 and 15 g agar bacteriological when solid medium was required

APPENDIX B

Supplementary Table S3.1: Significant differences observed between variable initial gas headspace compositions.

	Timepoint (h)	Aerobic control: Anaerobic control	Aerobic control: Syngas (50%)	Aerobic control: Syngas (100%)	Anaerobic control: Syngas (50%)	Anaerobic control: Syngas (100%)	Syngas (50%): Syngas (100%)
pH	0		***	***	***	***	***
	14						*
	24	**	**	*			
	38	***	***	***	**	***	**
	48	***	***	***		**	**
	62	***	***	***		**	
	72	***	***	***		*	
	84	***	***	***	**	***	***
	96	***	***	***		***	**
	110	***	***	***		***	*
	120	***	***	***		**	*
	134	***	***	***	*	**	
OD_{600} absorbance	0						
	14	***	***	***	*		***
	24	***	***	***			*
	38	***	***	***			
	48	***	***	***			
	62	***	***	***			
	72	***	***	***			
	84	***	***	***			
	96	***	***	***			
	110	***	***	***			
	120	***	***	***			
	134	***	***	***			
H_2 [mmol]	0		***	***	***	***	***
	14		***	***	***	**	*
	24	***	***	***	*		*
	38	***	*	***			*
	48		***	*	**		**
	62	***	***	***	***	**	***
	72	***	***	***	***	***	***
	84	**	***	***	***	***	***
	96	***	***	***	***	***	***
	110	**	***	***	***	***	***
	120	***	***	***	***	***	***
	134	*	***	***	***	***	***
CO_2 [mmol]	0		***	***	***	***	***
	14	***	***	***	***	***	***
	24		***	***	***	***	***
	38		***	***	***	***	***
	48	***		***	***	***	***

	Timepoint (h)	Aerobic control: Anaerobic control	Aerobic control: Syngas (50%)	Aerobic control: Syngas (100%)	Anaerobic control: Syngas (50%)	Anaerobic control: Syngas (100%)	Syngas (50%): Syngas (100%)
	62	***	***	***	***	***	***
	72	***	***	***	***	***	***
	84	***	***	***	***	***	***
	96	***	***	***	***	***	***
	110	***	***	***	***	***	***
	120	***	***	***	***	***	***
	134	***	***	***	***	***	***
O_2 [mmol]	0	***	***	***			
	14	***	***	***			
	24	***	***	***			
	38						
	48			*			
	62						
	72						
	84		*				
	96		*		*		
	110	*	***	*	*		*
	120						
	134				*		
CH_4 [mmol]	0		***	***	***	***	***
	14		***	***	***	***	***
	24		***	***	***	***	***
	38		***	***	***	***	***
	48		***	***	***	***	***
	62		***	***	***	***	***
	72		***	***	***	***	***
	84		***	***	***	***	***
	96		***	***	***	***	***
	110		***	***	***	***	***
	120		***	***	***	***	***
	134		***	***	***	***	***
Ethane [mmol]	0						
	14						
	24						
	38						
	48						
	62						
	72						
	84						
	96						
	110						
	120						
	134						
Propane [mmol]	0						
	14						
	24						

	Timepoint (h)	Aerobic control: Anaerobic control	Aerobic control: Syngas (50%)	Aerobic control: Syngas (100%)	Anaerobic control: Syngas (50%)	Anaerobic control: Syngas (100%)	Syngas (50%): Syngas (100%)
	38						
	48						
	62						
	72						
	84						
	96						
	110						
	120						
	134						
Butane [mmol]	0						
	14						
	24						
	38						
	48						
	62						
	72						
	84						
	96						
	110						
	120						
	134						
CO [mmol]	0	***	***	***	***	***	***
	14	***	***	***	***	***	***
	24	***	***	***	***		***
	38	***	***	***	***	*	***
	48		***		***		***
	62		*	*	**		***
	72			**	*		**
	84			**		*	**
	96			**		**	**
	110			**		**	**
	120			**		**	**
	134			*		*	*
H_2 mmol/L.h	0						
	14		***		***		***
	24	***	***	*		*	
	38		**		*		**
	48	***	***	***	***	***	**
	62	***	***	***	***	***	***
	72		*		**		**
	84	***	*	*	***	**	***
	96						
	110						
	120			**		**	
	134	***	***	***	***	***	***
	0						

	Timepoint (h)	Aerobic control: Anaerobic control	Aerobic control: Syngas (50%)	Aerobic control: Syngas (100%)	Anaerobic control: Syngas (50%)	Anaerobic control: Syngas (100%)	Syngas (50%): Syngas (100%)
Formate [mmol]	14		***		***		***
	24	***	***	*		*	
	38		**		*		**
	48	***	***	***	***	***	**
	62	***	***	***	***	***	***
	72		*		**		**
	84	***	*	*	***	**	***
	96						
	110						
	120			**		**	
	134	***	***	***	***	***	***
Acetate [mmol]	0						
	14	***	***	***		***	***
	24	***	***	***			
	38	***	***	***			
	48	***	***	***			
	62	***	***	***			
	72	***	***	***			
	84	***	***	***			
	96	***	***	***			
	110	***	***	***			
	120	***	***	***			
	134	***	***	***			
Glucose [mmol]	0						
	14			***		*	**
	24						
	38		*				
	48	**	***	***	**	**	
	62		***	***	**	*	
	72	*	***	**			
	84	**	***	***	*		
	96	*		***			*
	110			*			
	120	*		**			
	134	***	***	***			
Lactate [mmol]	0						
	14	***	***	*		*	**
	24	***	***	***			
	38	***	***	***			
	48	***	***	***			
	62	***	***	***			
	72	***	***	***			
	84	***	***	***			
	96	***	***	***			
	110	***	***	***			
	120	***	***	***			

	Timepoint (h)	Aerobic control: Anaerobic control	Aerobic control: Syngas (50%)	Aerobic control: Syngas (100%)	Anaerobic control: Syngas (50%)	Anaerobic control: Syngas (100%)	Syngas (50%): Syngas (100%)
	134	***	***	***			
Butyrate [mmol]	0						
	14						
	24						
	38						
	48	***	***	***			
	62						
	72						
	84						
	96						
	110						
	120						
	134						
Propionate [mmol]	0	**		*			
	14						
	24						
	38						
	48	**	*	*			
	62	***	***	***			
	72	*	*	**			
	84						
	96	***	**	***			
	110	***	***	**			
	120	**	**	**			
	134	***	***	***			
Valerate [mmol]	0						
	14						
	24						
	38						
	48						
	62						
	72						
	84						
	96						
	110	**	**			**	**
	120						
	134						
Propanol [mmol]	0						
	14						
	24						
	38						
	48						
	62						
	72						
	84						
	96						

	Timepoint (h)	Aerobic control: Anaerobic control	Aerobic control: Syngas (50%)	Aerobic control: Syngas (100%)	Anaerobic control: Syngas (50%)	Anaerobic control: Syngas (100%)	Syngas (50%): Syngas (100%)
	110						
	120						
	134						
Isobutyrate [mmol]	0	*		*			
	14						
	24						
	38	*					
	48	**		*			
	62						
	72						
	84						
	96						
	110						
	120						
	134						
Glyoxylate [mmol]	0						
	14						
	24						
	38						
	48	***	***	***			
	62	***	***	***			
	72	***	***	***			
	84	***	***	***			
	96	**	**	***			
	110	**	**	***			
	120	**	**	***			
	134	**	***	***			
Isovalerate [mmol]	0						
	14						
	24						
	38						
	48						
	62						
	72						
	84						
	96						
	110						
	120						
	134						
Succinate [mmol]	0						
	14	***		**	*		
	24						
	38	***	***	***			*
	48	***	***	***	**		**
	62	***	***	***	***	*	*
	72	***	***	***	**		*

	Timepoint (h)	Aerobic control: Anaerobic control	Aerobic control: Syngas (50%)	Aerobic control: Syngas (100%)	Anaerobic control: Syngas (50%)	Anaerobic control: Syngas (100%)	Syngas (50%): Syngas (100%)
	84	***	***	***	***	**	**
	96	***	***	***	***	***	***
	110	***	***	***	***	***	***
	120	***	***	***	***	***	***
	134	***	***	***	***	***	*
Ethanol [mmol]	0						
	14	**	***				*
	24	***	***	**			
	38	***	***	***			
	48	**	*	**			
	62	**	**	***			
	72	***	***	***			
	84	**	**	**			
	96	***	***	***			
	110	***	***	***			
	120	***	***	***			
	134	***	***	**			
Acetate mmol/L.h	0						
	14	***	***	***		***	***
	24						
	38						
	48						
	62						
	72						
	84						
	96						
	110						
	120						
	134						
Succinate mmol/L.h	0						
	14	**		*			
	24						
	38		**		*		*
	48	***	***	***	**		*
	62	***	***	***	***	***	
	72	*	**	*			
	84						
	96	**	**	**			
	110						
	120						
	134						

* P<0.05, ** P<0.01, ***P<0.001

Supplementary Table S3.2: Significant differences observed between H_2/CO yields across all experiments.

Different synthetic syngas composition experiment		
	H_2/CO yield at first point where CO is fully consumed	H_2/CO yield at fermentation endpoint
Aerobic control:Anaerobic control	***	***
Aerobic control:Syngas (50%)	***	*
Aerobic control:Syngas (100%)	N/A	N/A
Anaerobic control:Syngas (50%)	***	***
Anaerobic control:Syngas (100%)	N/A	***
Syngas (50%):Syngas (100%)	N/A	**
Pre-hydrogenogenic phase oxygen addition experiment.		
	H_2/CO yield at first point where CO is fully consumed	H_2/CO yield at fermentation endpoint
Aerobic control:14h Oxygen addition	***	***
Aerobic control:24h Oxygen addition	***	***
14h Oxygen addition:24h Oxygen addition	***	***
Hydrogenogenic phase oxygen addition experiment.		
	H_2/CO yield at first point where CO is fully consumed	H_2/CO yield at fermentation endpoint
Aerobic control:38h Oxygen addition	N/A	**
Aerobic control:48h Oxygen addition	N/A	*
38h Oxygen addition:48h Oxygen addition	N/A	
Variable oxygen additions during peak hydrogenogenesis experiment		
	H_2/CO yield at first point where CO is fully consumed	H_2/CO yield at fermentation endpoint
Aerobic control:1% Oxygen addition	***	***
Aerobic control:3% Oxygen addition	**	***
Aerobic control:5% Oxygen addition	***	**
Aerobic control:10% Oxygen addition	***	***
1% Oxygen addition:3% Oxygen addition	***	
1% Oxygen addition:5% Oxygen addition	***	***
1% Oxygen addition:10% Oxygen addition	***	***
3% Oxygen addition:5% Oxygen addition	***	*

3% Oxygen addition:10% Oxygen addition	***	***
5% Oxygen addition:10% Oxygen addition		

* P<0.05, ** P<0.01, ***P<0.001

Supplementary Table S3.3: Significant differences observed between pre-hydrogenogenic phase oxygen addition conditions.

	Timepoint	Aerobic control:14h Oxygen addition	Aerobic control:24h Oxygen addition	14h Oxygen addition:24h Oxygen addition
pH	0			*
	10			
	14			
	24	***		***
	34	***	***	**
	38	***	***	**
	48	***	***	
	58	***	***	
	72	***	***	
	84	***	***	
	96	***	***	
	110	***	***	
	120	***	***	
OD_{600} absorbance	0			
	10			
	14			
	24	***		***
	34			
	38			*
	48		***	***
	58	***	***	**
	72	*	**	*
	84		***	**
	96		**	
	110	*	**	
	120	*	**	
H_2 [mmol]	0			
	10			
	14			
	24			
	34	***	***	
	38	***	***	
	48	***	***	***
	58	***	***	***
	72	**		***
	84	**	*	*
	96	***	***	
	110		*	*
	120	***	***	***

	Timepoint	Aerobic control:14h Oxygen addition	Aerobic control:24h Oxygen addition	14h Oxygen addition:24h Oxygen addition
CO_2 [mmol]	0	*	***	*
	10			
	14			
	24	***		***
	34	***	***	
	38	***	***	
	48	***	***	***
	58	***	**	***
	72			
	84	*	**	
	96	***	***	
	110	**	**	
	120	**	**	
O_2 [mmol]	0			
	10		*	
	14			
	24	***		***
	34	***		***
	38	***	***	***
	48	***	***	***
	58	***	***	***
	72	***	***	**
	84			
	96			
	110			
	120			
CO [mmol]	0			
	10			
	14			
	24	***		***
	34	***	***	
	38	***	***	
	48	***	***	***
	58	***	***	**
	72	***	***	**
	84	***	**	*
	96	***	**	*
	110	**		
	120			
H_2 mmol/L.h	0			
	10			
	14			

	Timepoint	Aerobic control:14h Oxygen addition	Aerobic control:24h Oxygen addition	14h Oxygen addition:24h Oxygen addition
	24			
	34	***	***	
	38	***	***	***
	48		***	***
	58	***	***	***
	72	***	***	***
	84	*	***	**
	96			
	110	***	*	*
	120	***		***
Formate [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72			
	84			
	120			
Acetate [mmol]	0			
	14			
	24	***		***
	34	***	***	
	38	***	***	
	48	*	*	
	58	***	***	
	72	***	***	
	84	***	***	
	120	***	***	
Glucose [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72			
	84			
	120			
Lactate [mmol]	0			
	14			

	Timepoint	Aerobic control:14h Oxygen addition	Aerobic control:24h Oxygen addition	14h Oxygen addition:24h Oxygen addition
	24			
	34			
	38			
	48			
	58			
	72			
	84			
	120			
Butyrate [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72			
	84		*	
	120			
Propionate [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72			
	84			
	120			
Valerate [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72			
	84			
	120			
Propanol [mmol]	0			
	14			
	24			
	34			

	Timepoint	Aerobic control:14h Oxygen addition	Aerobic control:24h Oxygen addition	14h Oxygen addition:24h Oxygen addition
	38			
	48			
	58			
	72			
	84			
	120			
Isobutyrate [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72			
	84			
	120			
Glyoxylate [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58	*	*	
	72			
	84			
	120			
Isovalerate [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72	*		
	84			
	120	*	***	*
Succinate [mmol]	0			
	14		**	**
	24		***	***
	34	***		***
	38	***	***	
	48	**	*	

	Timepoint	Aerobic control:14h Oxygen addition	Aerobic control:24h Oxygen addition	14h Oxygen addition:24h Oxygen addition
	58	***	**	
	72			
	84	**		*
	120	**		*
Ethanol [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72			
	84			
	120			
Acetate mmol/L.h	0			
	14			
	24	***		***
	34	**	***	***
	38		*	*
	48			
	58			
	72			
	84			
	120			
Succinate mmol/L.h	0			
	14		*	*
	24			**
	34	**	***	
	38	**	***	
	48	*		
	58			
	72	***	**	
	84			
	120			

* P<0.05, ** P<0.01, ***P<0.001

Supplementary Table S3.4: Significant differences observed between hydrogenogenic phase oxygen addition conditions.

	Timepoint (h)	Aerobic control:38h Oxygen addition	Aerobic control:48h Oxygen addition	38h Oxygen addition:48h Oxygen addition
pH	0			
	10			
	14		***	***
	24			*
	34			
	38		*	*
	48	***	***	***
	58	***	***	
	72	***	***	
	84	***	***	*
	96	***	***	
	110	***	***	*
	120	***	***	***
OD_{600} absorbance	0			
	10			
	14		*	
	24			
	34		**	
	38		**	
	48		*	**
	58	***	***	***
	72	***	***	*
	84	***	***	
	96	**	**	
	110	***	***	
	120	***	***	
H_2 [mmol]	0			
	10			
	14			
	24			
	34			
	38		*	
	48	***	*	***
	58	***	***	**
	72	***	***	
	84	***	***	*
	96	**	**	
	110	**	*	
	120	***	***	*

	Timepoint (h)	Aerobic control:38h Oxygen addition	Aerobic control:48h Oxygen addition	38h Oxygen addition:48h Oxygen addition
CO_2 [mmol]	0			
	10			
	14			
	24			
	34		*	
	38		*	
	48	***	*	***
	58	***	***	*
	72	*	*	
	84	*	*	
	96	**	**	
	110	**	**	
	120	**	**	
O_2 [mmol]	0			
	10			
	14			
	24			
	34			
	38			
	48	***		***
	58	*	***	***
	72			
	84			
	96			
	110			
	120			
CO [mmol]	0			
	10		*	
	14			
	24	**	**	
	34		**	
	38		*	
	48	***	*	***
	58	***	***	**
	72	***	***	
	84	***	***	
	96	**	**	
	110	*	**	
	120	**	***	
H_2 mmol/L.h	0			
	10			
	14			

	Timepoint (h)	Aerobic control:38h Oxygen addition	Aerobic control:48h Oxygen addition	38h Oxygen addition:48h Oxygen addition
	24			
	34			
	38		*	
	48	***		***
	58		***	***
	72	***	***	***
	84	***	***	
	96			
	110			
	120			
Formate [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72			
	84		*	
	120	*	**	
Acetate [mmol]	0			
	14			
	24			
	34			
	38			
	48	**		***
	58	***	***	
	72	***	***	
	84	***	***	
	120	**	**	
Glucose [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72			
	84			
	120			
Lactate [mmol]	0			
	14			

	Timepoint (h)	Aerobic control:38h Oxygen addition	Aerobic control:48h Oxygen addition	38h Oxygen addition:48h Oxygen addition
	24			
	34			
	38			
	48			
	58	*		
	72			
	84			
	120		**	
Butyrate [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72			
	84			
	120			
Propionate [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72			
	84			
	120			
Valerate [mmol]	0			
	14			
	24			
	34		*	
	38			
	48			
	58			
	72			
	84			
	120			
Propanol [mmol]	0			
	14			
	24			
	34			

	Timepoint (h)	Aerobic control:38h Oxygen addition	Aerobic control:48h Oxygen addition	38h Oxygen addition:48h Oxygen addition
	38			
	48			
	58			
	72			
	84			
	120			
Isobutyrate [mmol]	0			
	14			
	24			
	34			
	38		*	
	48			
	58			
	72			
	84			
	120	*		
Glyoxylate [mmol]	0			
	14			
	24			
	34			
	38			
	48			*
	58	***	***	
	72	***	***	**
	84	***	***	***
	120	**	***	
Isovalerate [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72			
	84	*		
	120			
Succinate [mmol]	0			
	14			
	24			
	34			
	38			
	48	***	*	***

	Timepoint (h)	Aerobic control:38h Oxygen addition	Aerobic control:48h Oxygen addition	38h Oxygen addition:48h Oxygen addition
	58	***	*	*
	72			
	84	*	*	
	120			
Ethanol [mmol]	0			
	14			
	24			
	34			
	38			
	48			
	58			
	72			
	84			
	120			
Acetate mmol/L.h	0			
	14			
	24			
	34			
	38			
	48	**		**
	58	*	***	**
	72	***	***	
	84	*		**
	120			
Succinate mmol/L.h	0			
	14			
	24	*		*
	34			
	38			
	48	***	**	***
	58	**	*	
	72	***	**	*
	84	*		
	120			

* P<0.05, ** P<0.01, ***P<0.001

Supplementary Table S3.5: Significant differences in H_2 levels observed between timepoints.

Pre-hydrogenogenic phase oxygen addition experiment.				
	Timepoints (h)	Aerobic control	14h Oxygen addition	24h Oxygen addition
H_2 [mmol]	0-10			
	10-14			
	14-24			
	24-34	***		
	34-38	***		***
	38-48	***	***	***
	48-62	***	***	***
	62-72	***	***	***
	72-84			***
	84-96			***
	96-110			***
	110-120		***	***
Hydrogenogenic phase oxygen addition experiment.				
	Timepoints (h)	Aerobic control	38h Oxygen addition	48h Oxygen addition
H_2 [mmol]	0-10			
	10-14			
	14-24			
	24-34	***		
	34-38	***	***	
	38-48	***		***
	48-62	***	***	
	62-72	***	***	***
	72-84		***	***
	84-96			
	96-110			
	110-120			

Variable oxygen additions during peak hydrogenogenesis experiment						
	Timepoints (h)	Aerobic control	1% Oxygen addition	3% Oxygen addition	5% Oxygen addition	10% Oxygen addition
H_2 [mmol]	0-10					
	10-14					
	14-24	***	***	***	***	***
	24-34	**				
	34-38	**	***			
	38-48	***	***			
	48-62	***	***	***	***	***
	62-72		*	***	**	***
	72-84					***
	84-96					

	96-110					
	110-120					
	120-134					

* P<0.05, ** P<0.01, ***P<0.001

Supplementary Table S3.6: Significant differences observed between variable oxygen additions during peak hydrogenogenesis.

	Timepoint (h)	Aerobic control:1 % Oxygen addition	Aerobic control:3 % Oxygen addition	Aerobic control:5 % Oxygen addition	Aerobic control: 10% Oxygen addition	1% Oxygen addition:3 % Oxygen addition	1% Oxygen addition:5 % Oxygen addition	1% Oxygen addition: 10% Oxygen addition	3% Oxygen addition:5 % Oxygen addition	3% Oxygen addition: 10% Oxygen addition	5% Oxygen addition: 10% Oxygen addition
pH	0										
	14										
	24										
	38										
	41										
	44	*	***	***	***	**	***	***	*	*	
	48		***	***	***	**	***	***	**	***	**
	62		***	***	***	**	***	***	*	***	***
	72		***	***	***	***	***	***	***	***	***
	84		**	***	***	**	***	***	**	***	***
	96		***	***	***	***	***	***	***	***	***
	110		**	***	***	**	***	***	**	***	***
	120		***	***	***	**	***	***	*	***	***
	134		***	***	***	**	***	***	**	***	***
OD_{600} absorbance	0										
	14										
	24										
	38										
	41										
	44			**			*		*		*
	48	**		***	***			*	**	***	
	62	***	***	***	***		***	***	***	***	
	72	**	***	***	***		***	***	**	***	***

	Timepoint (h)	Aerobic control:1 % Oxygen addition	Aerobic control:3 % Oxygen addition	Aerobic control:5 % Oxygen addition	Aerobic control: 10% Oxygen addition	1% Oxygen addition:3 % Oxygen addition	1% Oxygen addition:5 % Oxygen addition	1% Oxygen addition: 10% Oxygen addition	3% Oxygen addition:5 % Oxygen addition	3% Oxygen addition: 10% Oxygen addition	5% Oxygen addition: 10% Oxygen addition
	84	*	**	***	***		***	***	**	***	
	96	*	***	***	***		**	***		***	*
	110	**	***	***	***		**	**			
	120		**	***	***		**	*			
	134	*	***	***	***		***	**	**		
H_2 [mmol]	0										
	14										
	24										
	38										
	41										
	44		**	**	***			*			
	48	*	***	***	***	***	***	***			
	62	*	***	***	***	**	***	***		***	***
	72	**	***	***	***		**	***		***	***
	84	***	***	***	***		**	***		***	***
	96	***	***	***	***	*	**	***		***	***
	110	***	***	***	***	**	***	***		**	*
	120	***	***	***	***		*	***		***	**
	134	***	***	***	***		***	***	**	**	
CO_2 [mmol]	0		**	**	***			**			
	14										
	24										
	38										
	41										

	Timepoint (h)	Aerobic control:1 % Oxygen addition	Aerobic control:3 % Oxygen addition	Aerobic control:5 % Oxygen addition	Aerobic control: 10% Oxygen addition	1% Oxygen addition:3 % Oxygen addition	1% Oxygen addition:5 % Oxygen addition	1% Oxygen addition: 10% Oxygen addition	3% Oxygen addition:5 % Oxygen addition	3% Oxygen addition: 10% Oxygen addition	5% Oxygen addition: 10% Oxygen addition
	44		*		**			*			
	48		***	***	***	**	**	***			
	62				***			***		***	***
	72				**			*		*	**
	84			*	**			*			
	96		**	***	***	**	***	***	*	***	***
	110		***	***	***	*	***	***	**	***	***
	120			***	***		**	***	*	***	***
	134	*	**	***	***		***	***	**	***	***
O_2 [mmol]	0										
	14										
	24										
	38										
	41	*	***	***	***	***	***	***	***	***	***
	44		***	***	***	***	***	***	***	***	***
	48		***	***	***	**	***	***	***	***	***
	62				***			***		***	***
	72										
	84										
	96										
	110										
	120										
	134										
CO [mmol]	0										
	14										

	Timepoint (h)	Aerobic control:1 % Oxygen addition	Aerobic control:3 % Oxygen addition	Aerobic control:5 % Oxygen addition	Aerobic control: 10% Oxygen addition	1% Oxygen addition:3 % Oxygen addition	1% Oxygen addition:5 % Oxygen addition	1% Oxygen addition: 10% Oxygen addition	3% Oxygen addition:5 % Oxygen addition	3% Oxygen addition: 10% Oxygen addition	5% Oxygen addition: 10% Oxygen addition
	24		*	*						*	
	38										
	41										
	44		*		*						
	48		***	***	***	**	**	***			
	62		*	**	***		*	***		***	
	72				***			***		***	
	84				***			***		***	
	96				***			***		***	
	110				***			***		***	
	120				***			***		***	
	134										
H_2 mmol/L.h	0										
	14										
	24										
	38										
	41	*	***	***	***	***	***	***			
	44		***	***	***	***	***	***			
	48		***	***		**	**			***	***
	62				***	*	**	***		***	***
	72				***			***		***	***
	84				***			***		***	***
	96				**			**		***	***
	110									*	
	120										

	Timepoint (h)	Aerobic control:1 % Oxygen addition	Aerobic control:3 % Oxygen addition	Aerobic control:5 % Oxygen addition	Aerobic control: 10% Oxygen addition	1% Oxygen addition:3 % Oxygen addition	1% Oxygen addition:5 % Oxygen addition	1% Oxygen addition: 10% Oxygen addition	3% Oxygen addition:5 % Oxygen addition	3% Oxygen addition: 10% Oxygen addition	5% Oxygen addition: 10% Oxygen addition
	134										
Formate [mmol]	0										
	14				*						*
	24										
	38										
	41										
	44										
	48										
	62					*				*	
	72										
	86		*			*			**	*	
	96		*			**			***	**	
	110								*	**	
	120		*	**		**	***	**			
	134								*		
Acetate [mmol]	0										
	14										
	24										
	38										
	41										
	44			*			**				
	48		*	***	***		***	***	***	***	***
	62		***	***	***	**	***	***	***	***	***
	72		***	***	***	***	***	***	***	***	***
	86	**	***	***	***	***	***	***	***	***	***

	Timepoint (h)	Aerobic control:1 % Oxygen addition	Aerobic control:3 % Oxygen addition	Aerobic control:5 % Oxygen addition	Aerobic control: 10% Oxygen addition	1% Oxygen addition:3 % Oxygen addition	1% Oxygen addition:5 % Oxygen addition	1% Oxygen addition: 10% Oxygen addition	3% Oxygen addition:5 % Oxygen addition	3% Oxygen addition: 10% Oxygen addition	5% Oxygen addition: 10% Oxygen addition
	96		***	***	***	***	***	***	***	***	***
	110		***	***	***	**	***	***	***	***	***
	120		**	***	***	*	***	***	**	***	*
	134		**	***	***	*	***	***	**	***	**
Glucose [mmol]	0										
	14										
	24										
	38										
	41										
	44										
	48										
	62										
	72										
	86										
	96										
	110										
	120										
	134										
Lactate [mmol]	0										
	14				*						*
	24										
	38										
	41										
	44										
	48										

	Timepoint (h)	Aerobic control:1 % Oxygen addition	Aerobic control:3 % Oxygen addition	Aerobic control:5 % Oxygen addition	Aerobic control: 10% Oxygen addition	1% Oxygen addition:3 % Oxygen addition	1% Oxygen addition:5 % Oxygen addition	1% Oxygen addition: 10% Oxygen addition	3% Oxygen addition:5 % Oxygen addition	3% Oxygen addition: 10% Oxygen addition	5% Oxygen addition: 10% Oxygen addition
	62				**			**		**	*
	72										
	86			*	*						
	96			***	***		***	***	***	***	
	110	*		***	***		***	***	***	***	
	120				**						
	134				*			*			
Butyrate [mmol]	0										
	14										
	24										
	38										
	41										
	44										
	48										
	62										
	72										
	86										
	96										
	110										
	120										
	134										
Propionate [mmol]	0										
	14										
	24										
	38										

	Timepoint (h)	Aerobic control:1 % Oxygen addition	Aerobic control:3 % Oxygen addition	Aerobic control:5 % Oxygen addition	Aerobic control: 10% Oxygen addition	1% Oxygen addition:3 % Oxygen addition	1% Oxygen addition:5 % Oxygen addition	1% Oxygen addition: 10% Oxygen addition	3% Oxygen addition:5 % Oxygen addition	3% Oxygen addition: 10% Oxygen addition	5% Oxygen addition: 10% Oxygen addition
	41										
	44										
	48										
	62					*					
	72										
	86										
	96			**	**		**	**	*	**	
	110			*	*		*	*	**	**	
	120										
	134										
Valerate [mmol]	0										
	14	**	*	**	***						
	24	*	**	**	**						
	38	***		**	**	*			*	*	
	41			*	*		**	**			
	44			*							
	48										
	62										
	72										
	86										
	96										
	110										
	120			*							
	134										
	0										

	Timepoint (h)	Aerobic control:1 % Oxygen addition	Aerobic control:3 % Oxygen addition	Aerobic control:5 % Oxygen addition	Aerobic control: 10% Oxygen addition	1% Oxygen addition:3 % Oxygen addition	1% Oxygen addition:5 % Oxygen addition	1% Oxygen addition: 10% Oxygen addition	3% Oxygen addition:5 % Oxygen addition	3% Oxygen addition: 10% Oxygen addition	5% Oxygen addition: 10% Oxygen addition
Propanol [mmol]	14										
	24										
	38										
	41										
	44										
	48										
	62										
	72										
	86										
	96										
	110										
	120										
	134										
Isobutyrate [mmol]	0										
	14										
	24										
	38										
	41										
	44										
	48										
	62										
	72				***			***		***	***
	86				***			***		**	*
	96				*			*			
	110				**			*		*	

	Timepoint (h)	Aerobic control:1 % Oxygen addition	Aerobic control:3 % Oxygen addition	Aerobic control:5 % Oxygen addition	Aerobic control: 10% Oxygen addition	1% Oxygen addition:3 % Oxygen addition	1% Oxygen addition:5 % Oxygen addition	1% Oxygen addition: 10% Oxygen addition	3% Oxygen addition:5 % Oxygen addition	3% Oxygen addition: 10% Oxygen addition	5% Oxygen addition: 10% Oxygen addition
	120				*			**		*	*
	134				**			*			
Glyoxylate [mmol]	0										
	14			***	***		***	***	***	***	
	24			***	***		***	***	***	***	
	38			***	***		***	***	***	***	
	41			**	**		*	*	**	**	
	44										
	48			**	**						
	62	***	***	***	***	***	***	***	***	***	
	72	***	***	***	***		***	***	***	***	
	86			***			***		***		***
	96			***	**		***	**	***	**	**
	110			***	***		***	***	***	***	*
	120	*	*	***	***		***	***	***	***	
	134			***	***		***	***	***	***	
Isovalerate [mmol]	0										
	14										
	24										
	38							*		*	
	41										
	44										
	48										
	62										
	72		*		**			*			*

	Timepoint (h)	Aerobic control:1 % Oxygen addition	Aerobic control:3 % Oxygen addition	Aerobic control:5 % Oxygen addition	Aerobic control: 10% Oxygen addition	1% Oxygen addition:3 % Oxygen addition	1% Oxygen addition:5 % Oxygen addition	1% Oxygen addition: 10% Oxygen addition	3% Oxygen addition:5 % Oxygen addition	3% Oxygen addition: 10% Oxygen addition	5% Oxygen addition: 10% Oxygen addition
	86										
	96										
	110				*						
	120			*	**		**	**			
	134										
Succinate [mmol]	0										
	14										
	24										
	38										
	41										
	44				*						
	48				**			*		*	
	62				***			***		***	**
	72				***			***		***	***
	86				**			***		***	***
	96				**			***		***	***
	110			*				**		*	**
	120			*				*		**	**
	134			*							*
Ethanol [mmol]	0										
	14										
	24										
	38										
	41										
	44										

	Timepoint (h)	Aerobic control:1 % Oxygen addition	Aerobic control:3 % Oxygen addition	Aerobic control:5 % Oxygen addition	Aerobic control: 10% Oxygen addition	1% Oxygen addition:3 % Oxygen addition	1% Oxygen addition:5 % Oxygen addition	1% Oxygen addition: 10% Oxygen addition	3% Oxygen addition:5 % Oxygen addition	3% Oxygen addition: 10% Oxygen addition	5% Oxygen addition: 10% Oxygen addition
	48										
	62										
	72										
	86										
	96										
	110										
	120										
	134										
Acetate mmol/L.h	0										
	14										
	24				*						
	38										
	41										
	44										
	48			***	***	***	***	***	***	***	
	62		**	***	***	**	***		***	**	
	72										
	86			*	*						
	96										
	110										
	120										
	134										
Succinate mmol/L.h	0										
	14										
	24										

	Timepoint (h)	Aerobic control:1 % Oxygen addition	Aerobic control:3 % Oxygen addition	Aerobic control:5 % Oxygen addition	Aerobic control: 10% Oxygen addition	1% Oxygen addition:3 % Oxygen addition	1% Oxygen addition:5 % Oxygen addition	1% Oxygen addition: 10% Oxygen addition	3% Oxygen addition:5 % Oxygen addition	3% Oxygen addition: 10% Oxygen addition	5% Oxygen addition: 10% Oxygen addition
	38										
	41				*						
	44				*						
	48				***			**		*	*
	62				***		**	***		***	***
	72	**	***	***	**		**				**
	86				**			**		**	
	96										
	110									*	
	120										
	134										

* P<0.05, ** P<0.01, ***P<0.001

APPENDIX C

Supplementary Table S4.1: Genome metadata of the taxa analysed.

Organism Name	Strain	Assembly Accession	Bioproject	Biosample	Assembly Level	WGS project accession	Isolation source	Country	BUSCO completeness (%)	# contigs/ replicons	Genome size (Mb)	G+C content (%)	Protein #	Plasmids	% genome/ plasmid	%proteins/ plasmid	Phage elements (intact/ incomplete)	% phages/ genome	% phage proteins/ genome
Parageobacillus thermantarcticus	DSM 9572^T	GCA_900111865.1	PRJEB17059	SAMN05192569	Scaffold	FOJS01	Geothermal soil	Antarctica	99.1	9	3.445	43.67	3497	1	1.67	0.86	6 (2/4)	4.19	6.41
Parageobacillus thermoglucosidasius	23.6	GCA_024509915.1	PRJNA668107	SAMN16397577	Complete Genome	CP063414.1-CP063417.1	Soil	Spain	99.1	3	4.024	43.69	3944	2	3.81	4.31	4 (1/3)	2.75	3.35
Parageobacillus thermoglucosidasius	B4168	GCA_001587555.1	PRJNA270597	SAMN03267297	Contig	LQYU01	Dairy product	Netherlands	99.1	1	3.737	43.85	3686	-	0.00	0.00	1 (1/0)	1.91	1.79
Parageobacillus thermoglucosidasius	BH4-1	GCA_022846475.1	PRJDB12551	SAMD00442892	Complete Genome	AP025621.1-AP025622.1	Non-contaminated soil	Japan	99.1	2	3.892	43.88	3762	1	2.40	2.50	1 (0/1)	0.57	0.56
Parageobacillus thermoglucosidasius	C56-YS93	GCA_000178395.2	PRJNA40781	SAMN02232024	Complete Genome	CP002835.1-CP002837.1	Hot Spring	USA	99.1	3	3.994	43.93	3944	2	2.52	2.41	6 (1/5)	2.36	3.22
Parageobacillus thermoglucosidasius	DSM 2542^T	GCA_001295365.1	PRJNA296418	SAMN04099008	Complete Genome	CP012712.1	Soil	Japan	98.6	3	3.980	43.84	3863	2	3.07	3.44	3 (1/2)	2.30	3.55
Parageobacillus thermoglucosidasius	DSM 2543	GCA_014218625.1	PRJNA482718	SAMN09711376	Contig	QQOJ01	Soil	Japan	99.1	5	3.962	43.80	3863	2	3.20	3.47	6 (0/6)	1.38	1.68
Parageobacillus thermoglucosidasius	DSM 6285	GCA_014218645.1	PRJNA482719	SAMN09711377	Contig	QQOK01	River sediment	USA	99.1	9	3.968	43.58	3868	1	1.77	1.84	7 (0/7)	1.58	1.99
Parageobacillus thermoglucosidasius	DSM 21625	GCA_014218665.1	PRJNA482720	SAMN09711378	Contig	QQOL01	Flax plants	Germany	99.1	22	4.006	43.83	3946	2	2.32	2.20	1 (0/1)	0.28	0.41
Parageobacillus thermoglucosidasius	G08C001	SRX1619421	PRJNA311332	SAMN04549607	Contig	SRX1619421	Soil	USA	99.1	14	3.858	43.83	3779	2	2.69	2.75	2 (0/2)	1.20	1.35
Parageobacillus thermoglucosidasius	GT23	GCA_001651535.1	PRJNA314192	SAMN04532072	Scaffold	LUCT01	Casein pipeline	Netherlands	99.1	1	3.694	43.80	3661	-	0.00	0.00	1 (1/0)	1.93	1.80
Parageobacillus thermoglucosidasius	NCIMB 11955	GCA_001700985.1	PRJNA330787	SAMN05416582	Complete Genome	CP016622.1-CP016624.1	TMO Renewables	United Kingdom	98.9	3	3.989	43.83	3871	2	3.30	3.54	6 (0/6)	1.37	1.68
Parageobacillus thermoglucosidasius	TG4	GCA_003865195.2	PRJDB7652	SAMD00150856	Scaffold	BHZK01	Marine sediment	Japan	99.1	3	3.945	43.93	3878	1	2.40	2.37	2 (1/1)	1.71	2.27
Parageobacillus thermoglucosidasius	TNO-09.020	GCA_000258725.1	PRJNA81577	SAMN02471215	Chromosome	AJJN01	Dairy factory biofilm	Netherlands	99.1	1	3.773	43.96	3701	-	0.00	0.00	1 (1/0)	1.82	1.70
Parageobacillus thermoglucosidasius	Y4.1MC1	GCA_000166075.1	PRJNA33183	SAMN00002562	Complete Genome	CP002293.1-CP002294.1	Hot Spring	USA	99.3	2	3.912	44.01	3817	1	1.83	1.86	2 (1/1)	1.96	2.65
Parageobacillus toebii	44C	GCA_014679925.1	PRJNA354604	SAMN06347029	Complete Genome	CP061475.1	Gold mine shaft	USA	99.1	2	3.336	42.53	3308	1	1.43	1.42	3 (1/2)	1.81	2.09
Parageobacillus toebii	B4110	GCA_001587455.1	PRJNA270597	SAMN03267295	Scaffold	LQYW01	Pea soup	Netherlands	99.5	9	3.526	42.16	3548	1	1.68	1.66	3 (1/2)	2.33	2.54
Parageobacillus toebii	DSM 14590^T	GCA_003688615.2	PRJNA455457	SAMN09062732	Complete Genome	CP049703.1-CP049704.1	Hay compost	Korea	98	2	3.323	42.36	3281	1	1.59	1.92	2 (0/2)	0.80	0.73
Parageobacillus toebii	DSM 18751	GCA_002217735.1	PRJNA383662	SAMN06770004	Contig	NDYL01	Compost	Italy	99.6	4	3.796	41.58	3816	-	0.00	0.00	7 (0/7)	2.76	3.38
Parageobacillus toebii	E263	GCA_007197795.1	PRJNA553850	SAMN12252226	Complete Genome	CP041632.1	Deep-sea hydrothermal vent	China	99.3	1	3.478	42.61	3397	-	0.00	0.00	4 (0/4)	1.57	1.41
Parageobacillus toebii	NEB718	GCA_016939435.1	PRJNA622823	SAMN17843211	Complete Genome	CP070511.1-CP070513.1	-	-	99.1	3	3.456	42.22	3441	2	1.56	1.71	3 (0/3)	1.87	2.18
Parageobacillus toebii	WCH70	GCA_000023385.1	SAMN00000635	SAMN00000635	Complete Genome	CP001638.1-CP001640.1	Compost	USA	99.8	3	3.509	42.80	3433	2	1.26	1.54	4 (0/4)	1.02	0.87
***Parageobacillus* genomosp. nov. A**	W-2	GCA_001655645.1	PRJNA320062	SAMN04915186	Contig	LXMA01	Oil reservoir	China	99.8	11	3.899	43.15	3882	1	2.03	1.78	8 (3/5)	5.79	6.29
***Parageobacillus* genomosp. nov. A**	KH3-4	GCA_022846435.1	PRJDB12551	SAMD00442693	Complete Genome	AP025627.1	Non-contaminated soil	Japan	99.8	1	3.817	43.41	3725	-	0.00	0.00	4 (0/4)	2.39	2.74
Saccharococcus caldoxylosilyticus	44B	GCA_014680125.1	PRJNA662697	SAMN06347028	Complete Genome	CP061476.1	Gold mine shaft	USA	99.4	1	3.774	44.27	3698	-	0.00	0.00	4 (0/4)	0.73	0.81
Saccharococcus caldoxylosilyticus	B4119	GCA_001587505.1	PRJNA270597	SAMN03267290	Scaffold	LQYS01	Dairy product	Netherlands	100	18	3.943	44.02	3986	3	3.23	3.31	4 (2/2)	2.80	3.49

Organism Name	Strain	Assembly Accession	Bioproject	Biosample	Assembly Level	WGS project accession	Isolation source	Country	BUSCO completeness (%)	# contigs/ replicons	Genome size (Mb)	G+C content (%)	Protein #	Plasmids	% genome/ plasmid	%proteins/ plasmid	Phage elements (intact/ incomplete)	% phages/ genome	% phage proteins/ genome
Saccharococcus caldoxylosilyticus	CIC9	GCA_000313345.1	PRJNA175758	SAMN02470176	Contig	AMRO01	Hot Spring	Indonesia	100	4	3.824	44.18	3757	-	0.00	0.00	2 (1/1)	1.84	2.24
Saccharococcus caldoxylosilyticus	DSM 12041^T	GCA_014196025.1	PRJNA583512	SAMN13173495	Scaffold	JACICX01	Soil	Australia	99.3	22	3.854	44.12	3764	1	1.50	1.86	2 (0/2)	1.27	0.74
Saccharococcus caldoxylosilyticus	ER4B	GCA_019272935.1	PRJNA344835	SAMN06209283	Complete Genome	CP040553.1-CP040554.1	Compost	Malaysia	99.8	2	3.912	44.32	3863	1	1.39	1.50	2 (1/1)	1.89	2.43
Saccharococcus caldoxylosilyticus	KH1-5	GCA_022846395.1	PRJDB12551	SAMD00442691	Complete Genome	AP025623.1	Non-contaminated soil	Japan	100	1	3.851	44.32	3783	-	0.00	0.00	3 (0/3)	1.10	0.90
Saccharococcus caldoxylosilyticus	KH1-6	GCA_022846415.1	PRJDB12551	SAMD00442692	Complete Genome	AP025624.1	Non-contaminated soil	Japan	100	1	3.851	44.32	3782	-	0.00	0.00	3 (0/3)	1.10	0.90
Saccharococcus caldoxylosilyticus	KH3-5	GCA_022846455.1	PRJDB12551	SAMD00442694	Complete Genome	AP025625.1-AP025626.1	Non-contaminated soil	Japan	100	2	3.839	44.19	3753	1	0.18	0.27	1 (0/1)	0.31	0.27
Saccharococcus caldoxylosilyticus	VR-IP	GCA_013357975.1	PRJNA637488	SAMN15143704	Scaffold	JABVYQ01	Iron particles from steam vent	India	99.8	2	3.823	44.04	3809	1	1.15	1.50	4 (1/3)	4.28	4.54
***Saccharococcus* genomosp. nov. A**	NUB3621	GCA_000632515.1	PRJNA189971	SAMN02727286	Chromosome	AOTZ01	Soil	China	99.8	6	3.626	44.40	3637	-	0.00	0.00	1 (0/1)	0.93	0.97
Saccharococcus thermophilus	DSM 4749^T	GCA_011761475.1	PRJNA332068	SAMN05444710	Contig	JAASRS01	Sugar beet extraction	Sweden	99.6	3	3.135	44.90	3085	2	2.60	2.63	3 (1/2)	3.29	4.51
Geobacillus thermodenitrificans	DSM 465^T	GCA_002072065.1	PRJNA347632	SAMN05894115	Complete Genome	CP017694.1	Sugar beet juice	Austria	99.3	1	3.473	49.13	3384	-	0.00	0.00	2 (2/0)	2.29	3.34

Supplementary Table S4.2:ANI (upper quadrant) and dDDH (lower quadrant) values between the analysed taxa.

		ANI																																	
dDDH	*Parageobacillus thermantarcticus* DSM 9572^{T}		90.3	90.6	90.5	90.8	90.5	90.5	90.5	90.6	90.3	90.4	90.5	90.4	90.6	90.6	86.7	87.1	86.9	86.5	87.0	86.6	86.9	92.0	92.1	83.4	83.6	83.4	83.4	83.5	83.5	83.5	83.3	83.1	83.2	83.3	75.0
	Parageobacillus thermoglucosidasius 23.6	40.5		99.0	98.9	98.9	98.9	98.9	98.8	98.9	99.5	99.0	98.9	99.5	99.0	99.0	85.9	86.1	86.1	85.9	86.2	86.1	85.9	91.5	91.6	83.3	83.4	83.5	83.3	83.4	83.3	83.3	83.3	83.0	82.9	83.3	74.5
	Parageobacillus thermoglucosidasius B4168	41.0	90.7		99.4	99.2	99.3	99.3	99.2	99.2	99.0	100.0	99.3	99.0	100.0	99.2	85.9	86.0	86.0	85.9	86.2	86.1	85.8	91.6	91.6	83.2	83.2	83.2	83.2	83.3	83.2	83.2	83.0	82.9	82.8	83.2	74.4
	Parageobacillus thermoglucosidasius BH4-1	41.2	89.4	94.4		99.2	99.3	99.4	99.2	99.2	98.9	99.4	99.3	98.9	99.4	99.3	86.0	86.1	86.1	86.0	86.2	86.0	85.9	91.5	91.7	83.3	83.4	83.3	83.3	83.5	83.2	83.2	83.2	83.0	82.8	83.3	74.4
	Parageobacillus thermoglucosidasius C56-YS93	41.4	89.0	92.7	91.7		99.2	99.2	99.2	99.2	98.9	99.2	99.2	98.8	99.2	99.2	86.0	86.1	86.3	86.1	86.4	86.2	86.1	91.2	91.4	83.1	83.2	83.1	83.2	83.3	83.2	83.1	83.0	82.7	82.9	83.3	74.6
	Parageobacillus thermoglucosidasius DSM 2542^{T}	41.1	89.7	93.4	94.2	93.1		100.0	99.3	99.2	98.9	99.3	100.0	98.8	99.3	99.3	86.0	86.0	86.2	86.1	86.3	86.1	85.9	91.6	91.6	83.2	83.5	83.3	83.5	83.5	83.2	83.2	83.3	83.2	83.0	83.3	74.5
	Parageobacillus thermoglucosidasius DSM 2543	41.1	89.6	93.5	94.2	93.1	99.9		99.3	99.2	98.9	99.3	100.0	98.8	99.3	99.3	85.9	86.2	86.1	86.0	86.2	86.2	85.8	91.5	91.6	83.3	83.4	83.3	83.4	83.5	83.2	83.2	83.3	83.2	83.1	83.3	74.6
	Parageobacillus thermoglucosidasius DSM 6285	40.9	88.7	93.2	92.8	92.2	93.6	93.6		99.1	98.8	99.2	99.2	98.8	99.2	99.2	85.9	86.2	86.4	86.0	86.2	86.2	86.0	91.5	91.6	83.2	83.3	83.2	83.8	83.4	83.2	83.2	83.1	83.0	83.0	83.3	74.4
	Parageobacillus thermoglucosidasius DSM 21625	41.2	88.9	92.3	92.3	92.5	93.2	93.1	92.8		98.8	99.2	99.2	98.8	99.2	99.3	86.0	86.1	86.3	86.0	86.2	86.0	86.0	91.4	91.5	83.3	83.2	83.1	83.1	83.3	83.2	83.2	83.1	82.8	83.0	83.2	74.6
	Parageobacillus thermoglucosidasius G08C001	40.5	94.5	90.7	89.4	89.0	89.9	89.9	88.9	89.1		99.0	98.9	99.5	99.0	99.0	85.8	86.0	86.1	85.8	86.2	85.9	85.9	91.5	91.6	83.1	83.3	83.2	83.3	83.3	83.1	83.1	83.3	83.1	82.9	83.3	74.6
	Parageobacillus thermoglucosidasius GT23	40.9	91.1	99.8	94.8	93.2	93.7	93.8	93.7	92.9	91.2		99.3	99.0	100.0	99.3	85.9	86.0	85.9	85.9	86.2	85.9	85.8	91.7	91.5	83.2	83.2	83.1	83.0	83.3	83.1	83.1	83.0	82.9	82.7	83.1	74.4
	Parageobacillus thermoglucosidasius NCIMB 11955	41.1	89.6	93.4	94.1	93.1	100.0	99.9	93.5	93.1	89.8	93.7		98.8	99.3	99.3	86.0	86.0	86.2	86.1	86.3	86.1	85.9	91.6	91.6	83.2	83.5	83.3	83.5	83.5	83.2	83.2	83.3	83.2	83.0	83.3	74.5
	Parageobacillus thermoglucosidasius TG4	40.6	94.6	90.5	88.7	89.0	89.5	89.4	89.1	88.8	95.6	91.0	89.4		99.0	98.9	85.9	85.9	85.9	85.8	86.1	86.1	85.9	91.4	91.7	83.4	83.3	83.2	83.2	83.5	83.2	83.2	83.2	83.1	82.8	83.3	74.6
	Parageobacillus thermoglucosidasius TNO-09.020	41.0	90.6	99.9	94.4	92.6	93.4	93.4	93.2	92.3	90.7	99.8	93.3	90.5		99.2	86.0	86.0	86.0	86.0	86.3	86.0	85.9	91.5	91.7	83.2	83.3	83.2	83.2	83.4	83.1	83.1	83.2	83.1	82.9	83.3	74.6
	Parageobacillus thermoglucosidasius Y4.1MC1	41.0	90.1	93.0	93.1	92.9	93.6	93.5	92.4	93.1	89.6	93.3	93.5	89.8	92.9		86.0	86.0	86.0	85.9	86.3	86.0	85.9	91.6	91.6	83.3	83.1	83.0	83.2	83.3	83.2	83.2	83.1	82.9	82.8	83.2	74.5
	Parageobacillus toebii 44C	32.0	30.8	30.7	30.9	31.1	30.9	30.9	30.8	30.9	30.7	30.5	30.9	30.9	30.9	30.9		97.5	98.8	98.8	97.8	98.8	98.0	86.6	86.8	86.7	84.7	84.6	84.8	84.8	84.8	84.8	84.7	84.6	84.7	85.0	74.3
	Parageobacillus toebii B4110	33.1	30.9	30.8	31.0	31.5	31.1	31.1	31.0	31.0	30.7	30.7	31.2	30.7	30.9	31.2	78.5		97.5	97.6	98.2	97.7	97.8	86.6	86.7	84.4	84.7	84.4	84.5	84.5	84.4	84.4	84.5	84.2	84.7	84.8	74.2
	Parageobacillus toebii DSM 18751	32.7	31.2	31.1	31.4	32.2	31.5	31.5	31.8	31.6	31.0	30.9	31.5	31.0	31.3	31.5	90.9	78.3		98.6	97.7	99.2	98.0	86.7	86.8	84.6	84.8	84.5	84.7	84.7	84.5	84.5	84.4	84.5	84.9	85.1	74.3
	Parageobacillus toebii DSM 14590^{T}	31.9	30.8	30.6	30.9	31.3	31.3	31.2	30.7	30.9	30.7	30.5	31.3	30.7	30.8	31.0	89.9	79.5	88.0		97.5	98.5	98.0	86.5	86.6	84.4	84.5	84.3	84.8	84.6	84.6	84.6	84.4	84.1	84.7	85.1	74.4
	Parageobacillus toebii E263	33.1	31.1	31.2	31.2	31.9	31.4	31.4	31.2	31.2	31.0	31.1	31.4	31.0	31.4	31.4	79.6	83.1	79.4	79.4		97.9	98.0	86.7	86.8	84.5	84.5	84.4	84.4	84.5	84.3	84.3	84.4	84.2	84.7	84.8	74.4
	Parageobacillus toebii NEB718	31.9	31.2	30.8	31.0	31.4	31.4	31.3	31.1	31.0	30.8	30.7	31.4	30.8	30.9	31.0	89.6	79.4	87.1	87.1	80.9		97.9	86.6	86.8	84.8	84.8	84.6	84.9	84.7	84.5	84.5	84.6	84.3	84.7	85.1	74.2
	Parageobacillus toebii WCH70	33.1	31.0	31.0	31.0	31.9	31.2	31.1	31.0	31.1	30.9	30.9	31.2	31.0	31.2	31.3	82.6	80.9	82.2	82.2	82.4	82.4		86.5	86.6	84.6	84.8	84.3	84.6	84.8	84.5	84.5	84.5	84.5	84.8	84.9	74.4
	Parageobacillus genomosp. A W-2	47.0	43.1	43.2	43.4	42.8	43.3	43.3	43.2	43.1	43.1	43.2	43.3	43.2	43.2	43.1	31.8	31.8	31.8	31.8	31.9	31.8	31.9		99.4	84.0	83.9	84.3	84.1	84.1	84.1	84.1	84.1	83.9	83.1	83.7	74.3
	Parageobacillus genomosp. nov. A KH3-4	47.3	43.9	43.7	44.1	43.6	44.1	44.0	43.6	43.8	43.6	43.6	44.0	44.0	43.9	44.0	32.3	32.1	32.2	32.2	32.4	32.4	32.4	95.0		84.1	84.0	84.5	84.3	84.3	84.2	84.2	84.2	83.9	83.4	83.8	74.2
	Saccharococcus caldoxylosilyticus 44B	26.8	27.0	26.8	27.1	27.1	27.1	27.1	26.8	27.1	26.8	26.6	27.1	26.9	26.9	27.3	29.0	28.4	28.7	28.7	28.8	29.0	28.8	27.8	28.3		99.0	99.0	98.9	99.1	99.2	99.2	99.2	98.1	90.3	92.1	75.1
	Saccharococcus caldoxylosilyticus B4119	27.4	27.0	26.9	27.1	27.0	27.3	27.3	26.8	27.0	26.8	26.8	27.4	26.9	27.0	27.1	28.9	28.8	29.0	29.0	29.1	29.1	29.0	27.7	28.0	91.7		98.7	98.8	99.2	99.1	99.1	99.1	97.9	90.3	92.1	75.5
	Saccharococcus caldoxylosilyticus CIC9	26.7	27.1	26.6	26.9	26.8	27.0	27.0	26.7	26.8	26.8	26.5	27.0	26.8	26.8	26.8	28.7	28.1	28.5	28.5	28.4	28.7	28.5	28.4	28.7	90.6	88.6		98.8	98.8	98.8	98.8	98.9	97.9	90.2	92.2	75.2
	Saccharococcus caldoxylosilyticus DSM 12041^{T}	27.1	27.0	26.9	27.2	27.2	27.7	27.6	27.7	27.2	27.0	26.8	27.7	27.1	27.1	27.3	29.1	28.8	29.6	29.6	29.1	29.4	29.4	28.0	28.5	91.1	89.7	89.4		98.9	98.9	98.9	99.1	98.9	90.3	92.2	75.2
	Saccharococcus caldoxylosilyticus ER4B	27.0	27.1	27.0	27.3	27.4	27.6	27.5	27.0	27.3	27.0	26.8	27.6	27.2	27.1	27.6	29.1	28.6	28.9	28.9	28.9	29.3	29.2	27.8	28.3	92.1	92.8	89.1	90.0		99.1	99.1	99.2	97.9	90.3	92.1	75.1
	Saccharococcus caldoxylosilyticus KH1-5	26.9	27.0	26.8	27.1	27.1	27.1	27.1	26.9	27.1	26.8	26.7	27.1	27.0	27.0	27.2	29.3	28.5	29.0	29.0	28.9	29.0	28.9	27.7	28.3	93.2	92.2	90.2	91.0	92.9		100.0	99.4	98.0	90.2	92.0	75.1
	Saccharococcus caldoxylosilyticus KH1-6	26.9	27.0	26.8	27.1	27.1	27.1	27.1	26.9	27.1	26.8	26.7	27.1	27.0	27.0	27.2	29.3	28.5	29.0	29.0	28.9	29.0	28.9	27.7	28.3	93.2	92.2	90.2	91.0	92.9	100.0		99.4	98.0	90.2	92.0	75.1
	Saccharococcus caldoxylosilyticus KH3-5	26.8	27.0	26.9	27.2	27.2	27.3	27.3	26.9	27.1	26.9	26.7	27.3	27.0	27.0	27.2	29.2	28.5	28.9	28.9	28.9	29.0	29.0	27.9	28.5	93.0	91.9	90.2	91.9	93.0	96.3	96.3		98.1	90.2	92.1	75.2
	Saccharococcus caldoxylosilyticus VR-IP	26.4	26.5	26.3	26.5	26.4	26.8	26.7	26.2	26.5	26.4	26.1	26.8	26.5	26.4	26.5	28.4	27.8	27.9	27.9	28.2	28.4	28.3	27.4	27.8	82.0	80.7	81.0	81.4	80.9	82.2	82.2	82.7		90.0	91.7	75.1
	Saccharococcus thermophilus DSM 4749^{T}	27.0	26.3	26.2	26.4	26.8	26.9	26.8	26.4	26.6	26.1	26.0	26.9	26.3	26.4	26.6	29.3	29.5	29.4	29.4	29.7	29.3	30.0	26.4	26.9	40.7	40.8	40.4	40.9	41.2	40.7	40.7	40.7	39.7		90.3	75.4
	Saccharococcus genomosp. nov. A NUB3621	26.5	26.9	26.5	26.7	26.6	26.7	26.7	26.6	26.6	26.6	26.4	26.8	26.7	26.6	26.7	29.2	28.9	29.4	29.4	29.0	29.4	29.0	27.1	27.6	46.7	46.4	46.7	46.6	46.7	46.6	46.6	46.8	44.9	40.8		75.2
	Geobacillus thermodenitrificans DSM 465^{T}	19.9	20.9	20.3	21.1	21.3	20.8	20.7	20.4	20.9	20.4	20.0	20.9	20.7	20.8	21.1	21.3	20.9	20.8	20.8	22.4	21.7	22.0	19.0	20.6	21.0	20.6	20.0	20.7	20.9	21.1	21.1	21.0	20.6	20.7	21.4	

Supplementary Table S4.3: The number of glycosyl hydrolases, polysaccharide lyases, carboxylesterases/lipases and proteins encoded on the genomes of the analysed taxa. The number of CAZYme and MEROPS families to which they are classified are indicated in brackets.

Taxon	Glycosyl hydrolases (GH families)	Polysaccharide lyases (PL families)	Carboxyl-esterase/ lipases (True lipases)	Proteases/ peptidases (MEROPS families)
***Parageobacillus thermantarcticus* DSM 9572^T**	37 (19)	0	13 (2)	130 (55)
***Parageobacillus thermoglucosidasius* 23.6**	23 (14)	0	9 (2)	132 (56)
***Parageobacillus thermoglucosidasius* B4168**	28 (16)	0	13 (2)	143 (55)
***Parageobacillus thermoglucosidasius* BH4-1**	22 (14)	0	11 (2)	137 (55)
***Parageobacillus thermoglucosidasius* C56-YS93**	26 (15)	0	12 (2)	138 (55)
***Parageobacillus thermoglucosidasius* DSM 2542^T**	29 (19)	0	12 (2)	142 (55)
***Parageobacillus thermoglucosidasius* DSM 2543**	29 (18)	0	12 (2)	141 (56)
***Parageobacillus thermoglucosidasius* DSM 6285**	29 (18)	0	12 (2)	143 (55)
***Parageobacillus thermoglucosidasius* DSM 21625**	20 (14)	0	12 (2)	143 (55)
***Parageobacillus thermoglucosidasius* G08C001**	32 (22)	1 (1)	12 (2)	142 (55)
***Parageobacillus thermoglucosidasius* GT23**	26 (15)	0	11 (2)	140 (55)
***Parageobacillus thermoglucosidasius* NCIMB 11955**	22 (14)	0	11 (2)	138 (55)
***Parageobacillus thermoglucosidasius* TG4**	29 (18)	0	12 (2)	144 (55)
***Parageobacillus thermoglucosidasius* TNO-09.020**	31 (16)	0	11 (2)	140 (56)
***Parageobacillus thermoglucosidasius* Y4.1MC1**	22 (14)	0	11 (2)	138 (55)
***Parageobacillus toebii* 44C**	24 (15)	0	11 (2)	138 (55)
***Parageobacillus toebii* B4110**	17 (10)	0	9 (2)	125 (55)
***Parageobacillus toebii* DSM 18751**	21 (12)	0	8 (2)	132 (54)

Taxon	Glycosyl hydrolases (GH families)	Polysaccharide lyases (PL families)	Carboxyl-esterase/ lipases (True lipases)	Proteases/ peptidases (MEROPS families)
***Parageobacillus toebii* DSM 14590^T**	18 (10)	0	10 (2)	132 (55)
***Parageobacillus toebii* E263**	14 (8)	0	8 (2)	124 (53)
***Parageobacillus toebii* NEB718**	14 (9)	0	8 (2)	132 (55)
***Parageobacillus toebii* WCH70**	14 (7)	0	8 (2)	128 (52)
***Parageobacillus* genomosp. A W-2**	15 (10)	0	7 (2)	125 (54)
***Parageobacillus* genomosp. nov. A KH3-4**	32 (20)	0	12 (2)	137 (55)
***Saccharococcus caldoxylosilyticus* 44B**	33 (20)	0	12 (2)	132 (55)
***Saccharococcus caldoxylosilyticus* B4119**	31 (17)	0	10 (2)	134 (52)
***Saccharococcus caldoxylosilyticus* CIC9**	32 (17)	0	10 (2)	152 (56)
***Saccharococcus caldoxylosilyticus* DSM 12041^T**	40 (24)	0	11 (2)	136 (54)
***Saccharococcus caldoxylosilyticus* ER4B**	41 (26)	0	11 (2)	140 (54)
***Saccharococcus caldoxylosilyticus* KH1-5**	38 (21)	0	12 (2)	145 (56)
***Saccharococcus caldoxylosilyticus* KH1-6**	44 (25)	3 (2)	15 (2)	135 (53)
***Saccharococcus caldoxylosilyticus* KH3-5**	44 (25)	3 (2)	15 (2)	135 (53)
***Saccharococcus caldoxylosilyticus* VR-IP**	37 (26)	0	12 (2)	137 (54)
***Saccharococcus thermophilus* DSM 4749^T**	36 (20)	1 (1)	10 (2)	140 (53)
***Saccharococcus* genomosp. nov. A NUB3621**	12 (6)	0	11 (2)	111 (50)
***Geobacillus thermodenitrificans* DSM 465^T**	25 (14)	0	10 (2)	139 (60)

Supplementary Table S4.4: Presence of DNA polymerase genes on the genomes of the analysed taxa.

Polymerase family	**A**	**B**	**C**					**Y**		**X**
Polymerase	**DNA PolI**	**DNA PolII**	**DNA PolIII**					**DNA PolIV**	**DNA PolV**	**DNA PolX**
Subunit			**α**	**τγ**	**δ**	**δ′**	**β**			
Gene name	***polA***	***polB***	***dnaE***	***dnaX***	***holA***	***holB***	***dnaN***	***dinB***	***umuC***	***polX***
***Parageobacillus thermantarcticus* DSM 9572^T**	1		1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* 23.6**	1	1	1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* B4168**	1		1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* BH4-1**	1		1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* C56-YS93**	1		1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* DSM 2542^T**	1		1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* DSM 2543**	1	1	1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* DSM 6285**	1		1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* DSM 21625**	1		1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* G08C001**	1		1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* GT23**	1		1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* NCIMB 11955**	1		1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* TG4**	1		1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* TNO-09.020**	1		1	1	1	1	1	1		1
***Parageobacillus thermoglucosidasius* Y4.1MC1**	1		1	1	1	1	1	1		1
***Parageobacillus toebii* 44C**	1		1	1	1	1	1	1		1
***Parageobacillus toebii* B4110**	1		1	1	1	1	1	1		1
***Parageobacillus toebii* DSM 18751**	1		1	1	1	1	1	1	1	1
***Parageobacillus toebii* DSM 14590^T**	1		1	1	1	1	1	1		1
***Parageobacillus toebii* E263**	1		1	1	1	1	1	1		1

Polymerase family	**A**	**B**	**C**					**Y**		**X**
Polymerase	**DNA PolI**	**DNA PolII**	**DNA PolIII**					**DNA PolIV**	**DNA PolV**	**DNA PolX**
Subunit			**α**	**τγ**	**δ**	**δ′**	**β**			
***Parageobacillus toebii* NEB718**	1		1	1	1	1	1	1		1
***Parageobacillus toebii* WCH70**	1		1	1	1	1	1	1		1
***Parageobacillus* genomosp. A W-2**	1		1	1	1	1	1	1	1	1
***Parageobacillus* genomosp. nov. A KH3-4**	1		1	1	1	1	1	1		1
***Saccharococcus caldoxylosilyticus* 44B**	1		1	1	1	1	1	1		1
***Saccharococcus caldoxylosilyticus* B4119**	1		1	1	1	1	1	1	1	1
***Saccharococcus caldoxylosilyticus* CIC9**	1		1	1	1	1	1	1	1	1
***Saccharococcus caldoxylosilyticus* DSM 12041^{T}**	1		1	1	1	1	1	1		1
***Saccharococcus caldoxylosilyticus* ER4B**	1		1	1	1	1	1	1	1	1
***Saccharococcus caldoxylosilyticus* KH1-5**	1		1	1	1	1	1	1	1	1
***Saccharococcus caldoxylosilyticus* KH1-6**	1		1	1	1	1	1	1	1	1
***Saccharococcus caldoxylosilyticus* KH3-5**	1		1	1	1	1	1	1	1	1
***Saccharococcus caldoxylosilyticus* VR-IP**	1		1	1	1	1	1	1	1	1
***Saccharococcus thermophilus* DSM 4749^{T}**	1		1	1	1	1	1	1	1	1
***Saccharococcus* genomosp. nov. A NUB3621**	1		1	1	1	1	1	1		1
***Geobacillus thermodenitrificans* DSM 465^{T}**	1		1	1	1	1	1	1		1

Supplementary figure S4.5: Restriction modification system components encoded on the genomes of the analysed taxa. Total number of components for each RM type are indicated outside of brackets. Numbers in brackets indicate the number of Restriction (R), Methylation (M), Specificity (S) or Restriction/Methylation (RM) components encoded on the genomes of the analysed taxa.

	Restriction Modification System Components			
Taxon	**Type I (R/M/S)**	**Type II (R/M/RM)**	**Type III (R/M)**	**Type IV**
***Parageobacillus thermantarcticus* DSM 9572^T**	8 (3/3/2)		2 (1/1)	
***Parageobacillus thermoglucosidasius* 23.6**	6 (3/2/1)		2 (2/0)	1
***Parageobacillus thermoglucosidasius* B4168**	6 (2/2/2)	1 (0/0/1)		1
***Parageobacillus thermoglucosidasius* BH4-1**	2 (1/0/1)			
***Parageobacillus thermoglucosidasius* C56-YS93**	10 (3/4/3)	2 (1/1/0)		
***Parageobacillus thermoglucosidasius* DSM 2542^T**	4 (1/2/1)		3 (2/1)	
***Parageobacillus thermoglucosidasius* DSM 2543**	1 (0/1/0)	1 (0/0/1)	3 (2/1)	
***Parageobacillus thermoglucosidasius* DSM 6285**	1 (0/1/0)	1 (0/0/1)	3 (2/1)	
***Parageobacillus thermoglucosidasius* DSM 21625**	9 (2/4/3)		3 (2/1)	
***Parageobacillus thermoglucosidasius* G08C001**	1 (0/1/0)		2 (2/0)	1
***Parageobacillus thermoglucosidasius* GT23**	6 (2/2/2)	1 (0/0/1)		1
***Parageobacillus thermoglucosidasius* NCIMB 11955**	1 (0/1/0)	1 (0/0/1)	3 (2/1)	
***Parageobacillus thermoglucosidasius* TG4**	1 (0/1/0)	3 (1/2/0)	4 (3/1)	
***Parageobacillus thermoglucosidasius* TNO-09.020**	6 (3/2/1)	1 (0/0/1)		1
***Parageobacillus thermoglucosidasius* Y4.1MC1**	4 (1/2/1)	4 (0/2/2)		
***Parageobacillus toebii* 44C**	1 (0/1/0)		2 (1/1)	
***Parageobacillus toebii* B4110**	7 (2/3/2)	1 (0/1/0)		1
***Parageobacillus toebii* DSM 18751**	3 (0/2/1)		5 (2/3)	2
***Parageobacillus toebii* DSM 14590^T**	2 (0/2/0)	1 (0/0/1)		1
***Parageobacillus toebii* E263**	9 (3/3/3)	1 (0/1/0)	2 (1/1)	
***Parageobacillus toebii* NEB718**	1 (0/1/0)	2 (1/1/0)		3
***Parageobacillus toebii* WCH70**	7 (1/4/2)	3 (1/1/1)	2 (1/1)	3
***Parageobacillus* genomosp. A W-2**				1
***Parageobacillus* genomosp. nov. A KH3-4**				1

	Restriction Modification System Components			
Taxon	**Type I (R/M/S)**	**Type II (R/M/RM)**	**Type III (R/M)**	**Type IV**
***Saccharococcus caldoxylosilyticus* 44B**		1 (0/0/1)		
***Saccharococcus caldoxylosilyticus* B4119**	4 (1/1/2)	4 (1/3/0)		
***Saccharococcus caldoxylosilyticus* CIC9**	4 (1/1/2)			1
***Saccharococcus caldoxylosilyticus* DSM 12041^{T}**	5 (1/2/2)	1 (0/1/0)		
***Saccharococcus caldoxylosilyticus* ER4B**	1 (0/0/1)	2 (0/2/0)		
***Saccharococcus caldoxylosilyticus* KH1-5**	1 (0/0/1)			
***Saccharococcus caldoxylosilyticus* KH1-6**	1 (0/0/1)			
***Saccharococcus caldoxylosilyticus* KH3-5**	1 (0/0/1)			
***Saccharococcus caldoxylosilyticus* VR-IP**				
Saccharococcus thermophilus DSM 4749^{T}	3 (1/1/1)	2 (1/1/0)	1 (1)	
***Saccharococcus* genomosp. nov. A NUB3621**				
Geobacillus thermodenitrificans DSM 465^{T}	1 (1/0/0)			

Supplementary figure S4.6: CRISPR-Cas genes on the genomes of the analysed taxa. Numbers outside of brackets indicate the total number of genes on the genomes of the analysed taxa, while numbers in brackets indicate the number of unique orthologs.

	Cas genes identified by CrisperCasFinder															
Taxon	***csx1u2_* Type IIIU**	***cmr1_* Type IIIB**	***cmr3_* Type IIIB**	***cmr4_* Type IIIB**	***cmr5_* Type IIIB**	***cmr6_* Type IIIB**	***cas1_* Type IB**	***cas2_* Type III-III**	***cas3_* Type I**	***cas4_* Type I-II**	***cas5b_* Type IB**	***cas5b2_* Type IB**	**cas6_ Type I-III**	***cas7b_* Type IB**	***cas7b2_* Type IB**	***cas8a1b2_* Type IB**
Parageobacillus thermantarcticus DSM 9572^{T}		1(1)	1(1)	2(1)	1(1)	2(2)	2(2)		3(3)	1(1)	2(2)		2(2)	2(2)		
Parageobacillus thermoglucosidasius 23.6		1(1)	1(1)	2(1)	1(1)	2(2)	2(2)	1(1)	3(3)	1(1)	2(2)		2(2)	2(2)		
Parageobacillus thermoglucosidasius B4168		1(1)	1(1)	1(1)	1(1)	2(2)	1(1)		2(2)		1(1)		1(1)	1(1)		
Parageobacillus thermoglucosidasius BH4-1	1(1)	3(3)	2(2)	3(2)	2(2)	3(3)	2(2)	1(1)	3(3)	1(1)	2(2)		2(2)	2(2)		
Parageobacillus thermoglucosidasius C56-YS93	1(1)	3(3)	2(2)	3(2)	2(2)	3(3)	2(2)	1(1)	3(3)	1(1)	2(2)		2(2)	2(2)		
Parageobacillus thermoglucosidasius DSM 2542^{T}	1(1)	3(3)	2(2)	4(3)	3(3)	4(4)	2(2)	1(1)	3(3)	1(1)	2(2)		2(2)	2(2)		
Parageobacillus thermoglucosidasius DSM 2543	1(1)	3(3)	2(2)	2(2)	2(2)	3(3)	1(1)		2(2)		1(1)		1(1)	1(1)		
Parageobacillus thermoglucosidasius DSM 6285	1(1)	3(3)	2(2)	2(2)	2(2)	3(3)	1(1)		2(2)		1(1)		1(1)	1(1)		
Parageobacillus thermoglucosidasius DSM 21625		2(2)	2(2)	3(2)	2(2)	3(3)	2(2)	1(1)	3(3)	1(1)	2(2)		2(2)	2(2)		
Parageobacillus thermoglucosidasius G08C001		2(2)	2(2)	2(2)	2(2)	3(3)	1(1)		2(2)		1(1)		1(1)	1(1)		

	Cas genes identified by CrisperCasFinder															
Taxon	***csx1u2_* Type IIIU**	***cmr1_* Type IIIB**	***cmr3_* Type IIIB**	***cmr4_* Type IIIB**	***cmr5_* Type IIIB**	***cmr6_* Type IIIB**	***cas1_* Type IB**	***cas2_* Type III-III**	***cas3_* Type I**	***cas4_* Type I-II**	***cas5b_* Type IB**	***cas5b2_* Type IB**	**cas6_ Type I-III**	***cas7b_* Type IB**	***cas7b2_* Type IB**	***cas8a1b2_* Type IB**
Parageobacillus thermoglucosidasius GT23		1(1)	1(1)	1(1)	1(1)	2(2)	1(1)		2(2)		1(1)		1(1)	1(1)		
Parageobacillus thermoglucosidasius NCIMB 11955	1(1)	3(3)	2(2)	2(2)	2(2)	3(3)	1(1)		2(2)		1(1)		1(1)	1(1)		
Parageobacillus thermoglucosidasius TG4		1(1)	1(1)	2(1)	1(1)	2(2)	2(2)	1(1)	3(3)	1(1)	2(2)		2(2)	2(2)		
Parageobacillus thermoglucosidasius TNO-09.020		1(1)	1(1)	1(1)	1(1)	2(2)	1(1)		2(2)		1(1)		1(1)	1(1)		
Parageobacillus thermoglucosidasius Y4.1MC1		1(1)	1(1)	2(1)	1(1)	1(1)	2(2)	1(1)	3(3)	1(1)	2(2)		2(2)	2(2)		
Parageobacillus toebii 44C						2(2)			2(2)							
Parageobacillus toebii B4110		2(2)	2(2)	3(2)	2(2)	2(2)	2(2)		2(2)		1(1)		1(1)	1(1)		
Parageobacillus toebii DSM 18751				1(1)		2(2)	1(1)	1(1)	3(3)	1(1)	1(1)		1(1)	1(1)		
Parageobacillus toebii DSM 14590^{T}						1(1)			2(2)							
Parageobacillus toebii E263		1(1)	1(1)	1(1)	1(1)	1(1)	1(1)		2(2)		1(1)		1(1)	1(1)		
Parageobacillus toebii NEB718						1(1)		1(1)	2(2)							
Parageobacillus toebii WCH70	1(1)	2(2)	2(2)	3(2)	2(2)	2(2)	2(2)	2(2)	4(4)	1(1)	1(1)	1(1)	1(1)	1(1)	1(1)	1(1)
Parageobacillus genomosp. A W-2	1(1)	2(2)	2(2)	2(2)	2(2)	2(2)	1(1)	1(1)	3(3)			1(1)	1(1)		1(1)	1(1)

	Cas genes identified by CrisperCasFinder															
Taxon	***csx1u2_* Type IIIU**	***cmr1_* Type IIIB**	***cmr3_* Type IIIB**	***cmr4_* Type IIIB**	***cmr5_* Type IIIB**	***cmr6_* Type IIIB**	***cas1_* Type IB**	***cas2_* Type III-III**	***cas3_* Type I**	***cas4_* Type I-II**	***cas5b_* Type IB**	***cas5b2_* Type IB**	**cas6_ Type I-III**	***cas7b_* Type IB**	***cas7b2_* Type IB**	***cas8a1b2_* Type IB**
Parageobacillus genomosp. nov. A KH3-4	1(1)	2(2)	2(2)	2(2)	2(2)	2(2)	1(1)	1(1)	3(3)			1(1)	1(1)		1(1)	1(1)
Saccharococcus caldoxylosilyticus 44B						1(1)		1(1)	2(2)			1(1)			1(1)	1(1)
Saccharococcus caldoxylosilyticus B4119						1(1)		1(1)	2(2)			1(1)			1(1)	1(1)
Saccharococcus caldoxylosilyticus CIC9						1(1)		1(1)	2(2)			1(1)			1(1)	1(1)
Saccharococcus caldoxylosilyticus DSM 12041^{T}						1(1)		1(1)	2(2)			1(1)			1(1)	1(1)
Saccharococcus caldoxylosilyticus ER4B						1(1)		1(1)	2(2)			1(1)			1(1)	1(1)
Saccharococcus caldoxylosilyticus KH1-5						1(1)		1(1)	2(2)			1(1)			1(1)	1(1)
Saccharococcus caldoxylosilyticus KH1-6						1(1)		1(1)	2(2)			1(1)			1(1)	1(1)
Saccharococcus caldoxylosilyticus KH3-5						1(1)		1(1)	2(2)			1(1)			1(1)	1(1)
Saccharococcus caldoxylosilyticus VR-IP						1(1)		1(1)	2(2)			1(1)			1(1)	1(1)
Saccharococcus thermophilus DSM 4749^{T}									2(2)							

	Cas genes identified by CrisperCasFinder															
Taxon	***csx1u2_* Type IIIU**	***cmr1_* Type IIIB**	***cmr3_* Type IIIB**	***cmr4_* Type IIIB**	***cmr5_* Type IIIB**	***cmr6_* Type IIIB**	***cas1_* Type IB**	***cas2_* Type III-III**	***cas3_* Type I**	***cas4_* Type I-II**	***cas5b_* Type IB**	***cas5b2_* Type IB**	**cas6_ Type I-III**	***cas7b_* Type IB**	***cas7b2_* Type IB**	***cas8a1b2_* Type IB**
Saccharococcus genomosp. nov. A NUB3621								1(1)	2(2)							
Geobacillus thermodenitrificans DSM 465^{T}						2(2)		1(1)	2(2)							

Supplementary Table S4.7: Bioremediation potential related genes on of the genomes of the analysed taxa.

	Xenobiotic degradation pathway components													
Taxon	**Amino-benzoate**	**Atra-zine**	**Ben-zoate**	**Capro-lactam**	**Chloroalkanes and chloroalkenes**	**Chlorocyclohexane and chlorobenzene**	**Dioxin**	**Ethyl-benzene**	**Fur-fural**	**Naph-thalene**	**Nitro-toluene**	**Sty-rene**	**Tolu-ene**	**Xyl-ene**
Parageobacillus thermantarcticus DSM 9572^{T}	6	3	11	2	2	3	3	2		1		1	1	5
Parageobacillus thermoglucosidasius 23.6	6	3	32	2	7	4	9	2	5	2		4	2	20
Parageobacillus thermoglucosidasius B4168	5	3	28	2	7	4	8	2		2		4	1	14
Parageobacillus thermoglucosidasius BH4-1	5	3	32	2	7	4	9	2		2		4	1	17
Parageobacillus thermoglucosidasius C56-YS93	5	3	28	2	7	4	8	2		2		4	1	14
Parageobacillus thermoglucosidasius DSM 2542^{T}	5	5	28	2	7	4	8	2		2		4	1	14
Parageobacillus thermoglucosidasius DSM 2543	5	3	32	2	7	4	9	2	5	2		4	2	18
Parageobacillus thermoglucosidasius DSM 6285	5	3	32	2	7	4	9	2	5	2		4	2	18
Parageobacillus thermoglucosidasius DSM 21625	5	3	31	2	7	4	9	2		2		4	1	17
Parageobacillus thermoglucosidasius G08C001	6	3	34	2	7	4	9	2	5	2		5	2	20
Parageobacillus thermoglucosidasius GT23	5	3	28	2	7	4	8	2		2		4	1	14
Parageobacillus thermoglucosidasius NCIMB 11955	5	3	32	2	7	4	9	2	5	2		4	2	18
Parageobacillus thermoglucosidasius TG4	7	3	34	2	7	4	8	2	5	2		6	2	20
Parageobacillus thermoglucosidasius TNO-09.020	5	3	28	2	7	4	8	2		2		4	1	14
Parageobacillus thermoglucosidasius Y4.1MC1	5	3	28	2	7	4	7	2	1	2		5	1	14
Parageobacillus toebii 44C	6		21	1	1	1	3	2		1		1	1	9
Parageobacillus toebii B4110	3		14	2	4	3	3	1		3			1	3
Parageobacillus toebii DSM 18751	4		20			1	3	2		2		1	1	8
Parageobacillus toebii DSM 14590^{T}	5		17	1	1	1	4	2		4		2	1	3
Parageobacillus toebii E263	3		14	1	1	1	3	1		2			1	1
Parageobacillus toebii NEB718	4		14	1	3	1		2	5	1		2	3	2
Parageobacillus toebii WCH70	2		12			1		1		2			1	1
Parageobacillus genomosp. A W-2	8	3	24	1	3	2	8	2	5	3		3	2	14
Parageobacillus genomosp. nov. A KH3-4	7	3	24	1	3	3	8	2	5	3	1	2	2	14
Saccharococcus caldoxylosilyticus 44B	6	3	13	2	5	3	3	2		1	1	2	1	4
Saccharococcus caldoxylosilyticus B4119	6	5	13	2	5	3	3	2	5	1	2	2	2	5
Saccharococcus caldoxylosilyticus CIC9	6	3	26	2	5	3	8	2		4	2	3	1	15
Saccharococcus caldoxylosilyticus DSM 12041^{T}	6	3	13	2	3	3	3	2		1	2	2	1	4
Saccharococcus caldoxylosilyticus ER4B	6	5	21	2	4	3	7	2	6	1	2	3	2	12
Saccharococcus caldoxylosilyticus KH1-5	7	3	21	2	5	3	7	2		1	2	2	1	11

	Xenobiotic degradation pathway components													
Taxon	**Amino-benzoate**	**Atra-zine**	**Ben-zoate**	**Capro-lactam**	**Chloroalkanes and chloroalkenes**	**Chlorocyclohexane and chlorobenzene**	**Dioxin**	**Ethyl-benzene**	**Fur-fural**	**Naph-thalene**	**Nitro-toluene**	**Sty-rene**	**Tolu-ene**	**Xyl-ene**
Saccharococcus caldoxylosilyticus KH1-6	8	3	21	2	5	3	7	2		1	2	2	1	11
Saccharococcus caldoxylosilyticus KH3-5	6		21	2	5	3	7	2		1	2	2	1	11
Saccharococcus caldoxylosilyticus VR-IP	6		17	2	5	2	3	2		3	2	2	1	5
Saccharococcus thermophilus DSM 4749^{T}	1		12	1	1	2		1		1	2		2	2
Saccharococcus genomosp. nov. A NUB3621	6	3	14	2	3		3	2	5	2		4	3	5
Geobacillus thermodenitrificans DSM 465^{T}	4		12	2	3	2	3	1		1	2		1	3

Supplementary Table S4.7 (continued)

	Chromosomally located phenol degradation associated genes											
Taxon	***pheA1/hpaB***	***pheA2/sgcE6***	***pheB***	***pheC***	***pheD***	***pheE***	***pheF***	***pheG***	***pheH***	***pheI***	***fdx***	***pheR***
Parageobacillus thermantarcticus DSM 9572^T												
Parageobacillus thermoglucosidasius 23.6	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius B4168	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius BH4-1	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius C56-YS93	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius DSM 2542^T	1	2	1	1	1	1	2	1	1	1	1	1
Parageobacillus thermoglucosidasius DSM 2543	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius DSM 6285	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius DSM 21625	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius G08C001	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius GT23	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius NCIMB 11955	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius TG4	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius TNO-09.020	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius Y4.1MC1	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus toebii 44C	1	2	1	1	1	1	1	1	1	1	1	1
Parageobacillus toebii B4110		1										
Parageobacillus toebii DSM 18751	1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus toebii DSM 14590^T		1										
Parageobacillus toebii E263		1										
Parageobacillus toebii NEB718		1										
Parageobacillus toebii WCH70												
Parageobacillus genomosp. A W-2	1	2	1	1	3	1	1	1	1	2	1	1
Parageobacillus genomosp. nov. A KH3-4	1	2	1	1	3	1	1	1	1	2	1	1
Saccharococcus caldoxylosilyticus 44B		1										
Saccharococcus caldoxylosilyticus B4119		1										
Saccharococcus caldoxylosilyticus CIC9	1	2	1	1	3	1	1	1	1	2	1	1
Saccharococcus caldoxylosilyticus DSM 12041^T		1										
Saccharococcus caldoxylosilyticus ER4B	1	2	1	1	1	1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus KH1-5	1	2	1	1	1	1	1	1	1	1	1	1

	Chromosomally located phenol degradation associated genes											
Taxon	***pheA1/hpaB***	***pheA2/sgcE6***	***pheB***	***pheC***	***pheD***	***pheE***	***pheF***	***pheG***	***pheH***	***pheI***	***fdx***	***pheR***
Saccharococcus caldoxylosilyticus KH1-6	1	2	1	1	1	1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus KH3-5	1	2	1	1	1	1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus VR-IP		1										
Saccharococcus thermophilus DSM 4749^{T}		1										
Saccharococcus genomosp. nov. A NUB3621		1										
Geobacillus thermodenitrificans DSM 465^{T}		2										

Supplementary Table S4.7 (continued)

	Plasmid located phenol degradation									
Taxon	***pheB***	***pheH***	***pheA1***	***pheA2***	***fdx***	***pheR***	***pheD***	***pheE***	***pheF***	***pheG***
Parageobacillus thermantarcticus DSM 9572^{T}										
Parageobacillus thermoglucosidasius 23.6	1	1	1	1	1	1	2	1	1	2
Parageobacillus thermoglucosidasius B4168										
Parageobacillus thermoglucosidasius BH4-1	1	1	1	1	1	1	2	1	1	2
Parageobacillus thermoglucosidasius C56-YS93										
Parageobacillus thermoglucosidasius DSM 2542^{T}										
Parageobacillus thermoglucosidasius DSM 2543	1	1	1	1	1	1	2	1	1	2
Parageobacillus thermoglucosidasius DSM 6285	1	1	1	1	1	1	2	1	1	2
Parageobacillus thermoglucosidasius DSM 21625	1	1	1	1	1	1	2	1	1	2
Parageobacillus thermoglucosidasius G08C001	1	1	1	1	1	1	2	1	1	2
Parageobacillus thermoglucosidasius GT23										
Parageobacillus thermoglucosidasius NCIMB 11955	1	1	1	1	1	1	2	1	1	2
Parageobacillus thermoglucosidasius TG4	1	1	1	1	1	1	2	1	1	2
Parageobacillus thermoglucosidasius TNO-09.020										
Parageobacillus thermoglucosidasius Y4.1MC1							1			1
Parageobacillus toebii 44C										
Parageobacillus toebii B4110										
Parageobacillus toebii DSM 18751										
Parageobacillus toebii DSM 14590^{T}										
Parageobacillus toebii E263										
Parageobacillus toebii NEB718										
Parageobacillus toebii WCH70										
Parageobacillus genomosp. A W-2								1	1	3
Parageobacillus genomosp. nov. A KH3-4								1	1	3
Saccharococcus caldoxylosilyticus 44B										
Saccharococcus caldoxylosilyticus B4119										
Saccharococcus caldoxylosilyticus CIC9								1	1	3
Saccharococcus caldoxylosilyticus DSM 12041^{T}										
Saccharococcus caldoxylosilyticus ER4B										
Saccharococcus caldoxylosilyticus KH1-5										

	Plasmid located phenol degradation									
Taxon	***pheB***	***pheH***	***pheA1***	***pheA2***	***fdx***	***pheR***	***pheD***	***pheE***	***pheF***	***pheG***
Saccharococcus caldoxylosilyticus KH1-6										
Saccharococcus caldoxylosilyticus KH3-5										
Saccharococcus caldoxylosilyticus VR-IP										
Saccharococcus thermophilus DSM 4749^{T}										
Saccharococcus genomosp. nov. A NUB3621										
Geobacillus thermodenitrificans DSM 465^{T}										

Supplementary Table S4.7 (continued)

	Sulphonate transporters								
Taxon	**A7K69_04610**	**A7K69_04615**	**A7K69_00870**	**A7K69_03840**	**A7K69_03835**	**A7K69_00875**	**A7K69_00880**	**A7K69_04620**	**A7K69_03845**
***Parageobacillus thermantarcticus* DSM 9572^T**									
***Parageobacillus thermoglucosidasius* 23.6**									
***Parageobacillus thermoglucosidasius* B4168**		1	1	1	1	1	1	1	1
***Parageobacillus thermoglucosidasius* BH4-1**		1	1	1	1	1	1	1	1
***Parageobacillus thermoglucosidasius* C56-YS93**		1	1	1	1	1	1	1	1
***Parageobacillus thermoglucosidasius* DSM 2542^T**		1	1			1	1	1	
***Parageobacillus thermoglucosidasius* DSM 2543**		1	1	1	1	1	1	1	1
***Parageobacillus thermoglucosidasius* DSM 6285**	1	1	1	1	1	1	1	1	1
***Parageobacillus thermoglucosidasius* DSM 21625**	1	1	1	1	1	1	1	1	1
***Parageobacillus thermoglucosidasius* G08C001**		1	1	1	1	1	1	1	1
***Parageobacillus thermoglucosidasius* GT23**		1	1	1	1	1	1	1	1
***Parageobacillus thermoglucosidasius* NCIMB 11955**		1	1	1	1	1	1	1	1
***Parageobacillus thermoglucosidasius* TG4**	1	1	1	1	1	1	1	1	1
***Parageobacillus thermoglucosidasius* TNO-09.020**		1	1	1	1	1	1	1	1
***Parageobacillus thermoglucosidasius* Y4.1MC1**		1	1	1	1	1	1	1	1
***Parageobacillus toebii* 44C**		1	1	1	1	1	1	1	1
***Parageobacillus toebii* B4110**									
***Parageobacillus toebii* DSM 18751**									
***Parageobacillus toebii* DSM 14590^T**									
***Parageobacillus toebii* E263**									
***Parageobacillus toebii* NEB718**									
***Parageobacillus toebii* WCH70**									
***Parageobacillus* genomosp. A W-2**		1	1			1	1	1	
***Parageobacillus* genomosp. nov. A KH3-4**		1	1	1	1	1	1	1	1
***Saccharococcus caldoxylosilyticus* 44B**		1	1	1	1	1	1	1	1
***Saccharococcus caldoxylosilyticus* B4119**		1	1			1	1	1	
***Saccharococcus caldoxylosilyticus* CIC9**		1	1			1	1	1	
***Saccharococcus caldoxylosilyticus* DSM 12041^T**		1	1			1	1	1	
***Saccharococcus caldoxylosilyticus* ER4B**		1	1			1	1	1	
***Saccharococcus caldoxylosilyticus* KH1-5**		1	1			1	1	1	

	Sulphonate transporters								
Taxon	**A7K69_04610**	**A7K69_04615**	**A7K69_00870**	**A7K69_03840**	**A7K69_03835**	**A7K69_00875**	**A7K69_00880**	**A7K69_04620**	**A7K69_03845**
***Saccharococcus caldoxylosilyticus* KH1-6**		1	1			1	1	1	
***Saccharococcus caldoxylosilyticus* KH3-5**		1	1			1	1	1	
***Saccharococcus caldoxylosilyticus* VR-IP**		1	1			1	1	1	
***Saccharococcus thermophilus* DSM 4749^{T}**		1	1			1	1	1	
***Saccharococcus* genomosp. nov. A NUB3621**									
***Geobacillus thermodenitrificans* DSM 465^{T}**									

Supplementary Table S4.7 (continued)

	Desulphurisation gene cluster 1														
Taxon	***gmuE***	***ssuE***	***ytlI 3***	**Hypothetical protein**	***ssuB***	**AEH463 06.1**	**WP_042 384758.1**	***ssuC***	**EID4233 5.1**	**WP_042 384758.1**	***metN***	***metP_2***	***sfbA***	***sfbA***	**EID4233 0.1**
***Parageobacillus thermantarcticus* DSM 9572^T**	1	1			1		1	1			1	1			
***Parageobacillus thermoglucosidasius* 23.6**	1														
***Parageobacillus thermoglucosidasius* B4168**	1	1	1	1	1	1	1	1	1	1	2	2		1	1
***Parageobacillus thermoglucosidasius* BH4-1**	1	1	1	1	1	1	1	1	1	1	2	2		1	1
***Parageobacillus thermoglucosidasius* C56-YS93**	1	1	1	1	1	1	1	1	1	1	2	2		1	1
***Parageobacillus thermoglucosidasius* DSM 2542^T**	1	1	1	1	1	1	1	1	1	1	2	2		1	1
***Parageobacillus thermoglucosidasius* DSM 2543**	1	1	1	1	1	2	1	1	1	1	2	2		1	1
***Parageobacillus thermoglucosidasius* DSM 6285**	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1
***Parageobacillus thermoglucosidasius* DSM 21625**	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1
***Parageobacillus thermoglucosidasius* G08C001**	1	1	1	1	1	1	1	1	1	1	2	2		1	1
***Parageobacillus thermoglucosidasius* GT23**	1	1	1	1	1	1	1	1	1	1	2	2		1	1
***Parageobacillus thermoglucosidasius* NCIMB 11955**	1	1	1	1	1	1	1	1	1	1	2	2		1	1
***Parageobacillus thermoglucosidasius* TG4**	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1
***Parageobacillus thermoglucosidasius* TNO-09.020**	1	1	1	1	1	1	1	1	1	1	2	2		1	1
***Parageobacillus thermoglucosidasius* Y4.1MC1**	1	1	1	1	1	1	1	1	1	1	2	2		1	1
***Parageobacillus toebii* 44C**	1	1	1	1	1	1	1	1	1	1	2	2		1	1
***Parageobacillus toebii* B4110**	1														
***Parageobacillus toebii* DSM 18751**	2														
***Parageobacillus toebii* DSM 14590^T**	1														
***Parageobacillus toebii* E263**	1														
***Parageobacillus toebii* NEB718**	1														
***Parageobacillus toebii* WCH70**	1														
***Parageobacillus* genomosp. A W-2**	1														
***Parageobacillus* genomosp. nov. A KH3-4**	1									1	1	1		1	1
***Saccharococcus caldoxylosilyticus* 44B**	1									1	1	1		1	1
***Saccharococcus caldoxylosilyticus* B4119**	1										1	1			
***Saccharococcus caldoxylosilyticus* CIC9**	1										1	1			
***Saccharococcus caldoxylosilyticus* DSM 12041^T**	1														
***Saccharococcus caldoxylosilyticus* ER4B**	1										1	1			
***Saccharococcus caldoxylosilyticus* KH1-5**	1	1	1	1	1	1	1	1	1	1	2	2		1	1

	Desulphurisation gene cluster 1														
Taxon	***gmuE***	***ssuE***	***ytlI 3***	**Hypothetical protein**	***ssuB***	**AEH463 06.1**	**WP_042 384758.1**	***ssuC***	**EID4233 5.1**	**WP_042 384758.1**	***metN***	***metP_2***	***sfbA***	***sfbA***	**EID4233 0.1**
***Saccharococcus caldoxylosilyticus* KH1-6**	1										1	1			
***Saccharococcus caldoxylosilyticus* KH3-5**	1										1	1			
***Saccharococcus caldoxylosilyticus* VR-IP**	1	1	1	1	1	1	1	1	1	1	2	2		1	1
***Saccharococcus thermophilus* DSM 4749^{T}**	1										1	1			
***Saccharococcus* genomosp. nov. A NUB3621**											1	1			
***Geobacillus thermodenitrificans* DSM 465^{T}**	1										1	1			

Supplementary Table S4.7 (continued)

	Desulphurisation gene cluster 2							Desulphurisation gene cluster 3									
Taxon	***ssuE***	**EID43840.1**	**WP_042383252.1**	***ssuB***	***ssuA***	***ssuA***	***ssuC***	***trxA***	***ssuB***	***ssuA***	***ssuC***	***ssuE***	**EID45214.1**	**WP_042384492.1**	***mprF***	**WP_042384499.1**	**WP_013400279.1**
Parageobacillus thermantarcticus DSM 9572^T								1	1	1			1				
Parageobacillus thermoglucosidasius 23.6								1									
Parageobacillus thermoglucosidasius B4168	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius BH4-1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius C56-YS93	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius DSM 2542^T	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius DSM 2543	1	1	1	1		2	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius DSM 6285	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius DSM 21625	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius G08C001	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius GT23	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius NCIMB 11955	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius TG4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius TNO-09.020	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius Y4.1MC1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus toebii 44C	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Parageobacillus toebii B4110								1	1					1	1	1	1
Parageobacillus toebii DSM 18751								1									
Parageobacillus toebii DSM 14590^T								1	1					1	1	1	1
Parageobacillus toebii E263								1	1					1	1	1	1
Parageobacillus toebii NEB718								1									
Parageobacillus toebii WCH70								1	1					1	1	1	1
Parageobacillus genomosp. A W-2								1	1	1	1	1	1	1	1	1	1
Parageobacillus genomosp. nov. A KH3-4	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus 44B	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus B4119	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus CIC9	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus DSM 12041^T	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus ER4B	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus KH1-5	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1

	Desulphurisation gene cluster 2							Desulphurisation gene cluster 3									
Taxon	***ssuE***	**EID43840.1**	**WP_042383252.1**	***ssuB***	***ssuA***	***ssuA***	***ssuC***	***trxA***	***ssuB***	***ssuA***	***ssuC***	***ssuE***	**EID45214.1**	**WP_042384492.1**	***mprF***	**WP_042384499.1**	**WP_013400279.1**
Saccharococcus caldoxylosilyticus KH1-6	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus KH3-5	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus VR-IP	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Saccharococcus thermophilus DSM 4749^{T}								1	1	1	1	1	1	1	1	1	1
Saccharococcus genomosp. nov. A NUB3621								1									
Geobacillus thermodenitrificans DSM 465^{T}								1						1	1	1	1

Supplementary Table S4.7 (continued)

	Long-chain alkane degradation							Nitroalkane degradation		
Taxon	***ladAα***	***ladAβ***	***ladB***	**ABO68462.1**	**ABO66657.1**	**ABO67118.1**	**ABO68223.1**	**WP_064553126**	**WP_064551563.1**	**WP_064551165**
Parageobacillus thermantarcticus DSM 9572^{T}				1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius 23.6				1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius B4168	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius BH4-1	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius C56-YS93	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius DSM 2542^{T}	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius DSM 2543	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius DSM 6285	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius DSM 21625	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius G08C001	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius GT23	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius NCIMB 11955	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius TG4	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius TNO-09.020	1	1	1	1	1	1	1	1	1	1
Parageobacillus thermoglucosidasius Y4.1MC1	1	1	1	1	1	1	1	1	1	1
Parageobacillus toebii 44C	1	1	1	1	1	1	1	1	1	1
Parageobacillus toebii B4110				2	1	1	1	1	1	1
Parageobacillus toebii DSM 18751				2	1	1	1	1	1	1
Parageobacillus toebii DSM 14590^{T}				1	1	2	1	1	1	1
Parageobacillus toebii E263				2	1	1	1	1		1
Parageobacillus toebii NEB718				2	1	1	1	1	1	1
Parageobacillus toebii WCH70				1	1	2	1	1	1	1
Parageobacillus genomosp. A W-2				1	1	1	1			1
Parageobacillus genomosp. nov. A KH3-4	1	1		1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus 44B	1	1		1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus B4119				2	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus CIC9				2	1	1	1	2	1	1
Saccharococcus caldoxylosilyticus DSM 12041^{T}				2	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus ER4B				1	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus KH1-5	1	1		1	1	1	1	1	1	1

	Long-chain alkane degradation							Nitroalkane degradation		
Taxon	***ladAα***	***ladAβ***	***ladB***	**ABO68462.1**	**ABO66657.1**	**ABO67118.1**	**ABO68223.1**	**WP_064553126**	**WP_064551563.1**	**WP_064551165**
Saccharococcus caldoxylosilyticus KH1-6				2	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus KH3-5				2	1	1	1	1	1	1
Saccharococcus caldoxylosilyticus VR-IP	1	1		2	1	1	1	2	1	1
Saccharococcus thermophilus DSM 4749^{T}				1	1	1	1	1	1	1
Saccharococcus genomosp. nov. A NUB3621					1		1			1
Geobacillus thermodenitrificans DSM 465^{T}				2	1	1	1	1	1	1

REFERENCES

Abby, S.S., Néron, B., Ménager, H., Touchon, M. and Rocha, E.P., 2014. MacSyFinder: A program to mine genomes for molecular systems with an application to CRISPR-Cas systems, *PLoS one*, *9*(10), p.e110726.

Abdalla, A.M., Hossain, S., Nisfindy, O.B., Azad, A.T., Dawood, M. and Azad, A.K., 2018. Hydrogen production, storage, transportation and key challenges with applications: A review. *Energy Conversion and Management*, *165*, pp.602–627.

Abdelsalam, E.M., Samer, M., Moselhy, M.A., Arisha, A.H., Abdelqader, A.A. and Attia, Y.A., 2023. Effects of He–Ne red and green laser irradiation on purple non-sulfur bacteria for biohydrogen production from food wastes. *Biomass Conversion and Biorefinery*, *13*(15), pp.13883–13895.

Abe, J.O., Popoola, A.P.I., Ajenifuja, E. and Popoola, O.M., 2019. Hydrogen energy, economy and storage: Review and recommendation. *International Journal of Hydrogen Energy*, *44*(29), pp.15072–15086.

Abubackar, H.N., Veiga, M.C. and Kennes, C. In: Taherzadeh, M.J., Bolton, K., Wong, J. and Pandey, A. Syngas Fermentation for Bioethanol and Bioproducts, in: Taherzadeh, M.J., Bolton, K., Wong, J., Pandey, A. (Eds.). Sustainable Resource Recovery and Zero Waste Approaches. Elsevier, 2019, pp.207–221.

Acar, C. and Dincer, I., 2014. Comparative assessment of hydrogen production methods from renewable and non-renewable sources. *International Journal of Hydrogen Energy*, *39*(1), pp.1–12.

Achour, Y., Berrada, A., Arechkik, A. and El Mrabet, R., 2023. Techno-economic assessment of hydrogen production from three different solar photovoltaic technologies. *International Journal of Hydrogen Energy*, *48*(83), pp.32261-32276.

Adachi, Y., Inoue, M., Yoshida, T. and Sako, Y., 2020. Short communication: Genetic engineering of carbon monoxide-dependent hydrogen-producing machinery in P*arageobacillus thermoglucosidasius*. *Microbes and Environments*, *35*(4), pp.1–5.

Adlan, N.A., Sabri, S., Masomian, M., Ali, M.S.M. and Rahman, R.N.Z.R.A., 2020. Microbial biodegradation of paraffin wax in Malaysian crude oil mediated by degradative enzymes. *Frontiers in Microbiology*, 11, p.565608.

Agrawal, P., Amir, S., Barua, D. and Mohanty, D., 2021. RiPPMiner-Genome: a web resource for automated prediction of crosslinked chemical structures of RiPPs by genome mining. *Journal of Molecular Biology*, *433*(11), p.166887.

Ahmad, S., Scopes, R.K., Rees, G.N. and Patel, B.K., 2000. *Saccharococcus caldoxylosilyticus* sp. nov., an obligately thermophilic, xylose-utilizing, endospore-forming bacterium. *International Journal of Systematic and Evolutionary Microbiology*, *50*(2), pp.517-523.

Akhlaghi, N. and Najafpour-Darzi, G., 2020. A comprehensive review on biological hydrogen production. *International Journal of Hydrogen Energy*, *45*(43), pp.22492–22512.

Al Mamun, A.A.M. (2007). Elevated expression of DNA polymerase II increases spontaneous mutagenesis in *Escherichia coli*. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *625*(1-2), pp.29–39.

Alfano, M. and Cavazza, C., 2018. The biologically mediated water-gas shift reaction: Structure, function and biosynthesis of monofunctional [NiFe]-carbon monoxide dehydrogenases. *Sustainable Energy & Fuels*, *2*(8), pp.1653–1670.

Aliotta, J.M., Pelletier, J.J., Ware, J.L., Moran, L.S., Benner, J.S. and Kong, H., 1996. Thermostable Bst DNA polymerase I lacks a 3′→ 5′ proofreading exonuclease activity. *Genetic analysis: biomolecular engineering*, *12*(5-6), pp.185-195.

Aliyu, H., De Maayer, P. and Neumann, A. (2021). Not all that glitters is gold: The paradox of CO-dependent hydrogenogenesis in *Parageobacillus thermoglucosidasius*. *Frontiers in Microbiology*, *12*, p.784652.

Aliyu, H., Kastner, R., De Maayer, P. and Neumann, A., 2021. Carbon monoxide induced metabolic shift in the carboxydotrophic *Parageobacillus thermoglucosidasius* DSM 6285. *Microorganisms*, *9*(5), p.1090.

Aliyu, H., Lebre, P., Blom, J., Cowan, D. and De Maayer, P., 2016. Phylogenomic re-assessment of the thermophilic genus *Geobacillus*. *Systematic and Applied Microbiology*, *39*(8), pp.527-533.

Aliyu, H., Lebre, P., Blom, J., Cowan, D. and De Maayer, P., 2018. Corrigendum to "Phylogenomic re-assessment of the thermophilic genus Geobacillus" [*Systematic and Applied Microbiology,* (2016), 39, pp.527-533]. *Systematic and Applied Microbiology*, *41*(5), pp.529-530.

Aliyu, H., Mohr, T., Cowan, D., De Maayer, P. and Neumann, A., 2020. Time-course transcriptome of *Parageobacillus thermoglucosidasius* DSM 6285 grown in the presence of carbon monoxide and air. *International Journal of Molecular Sciences*, *21*(11), pp.1–16.

Allaart, M.T., Diender, M., Sousa, D.Z. and Kleerebezem, R., 2023. Overflow metabolism at the thermodynamic limit of life: How carboxydotrophic acetogens mitigate carbon monoxide toxicity. *Microbial Biotechnology*, *16*(4), pp.697–705.

Argun, H. and Kargi, F., 2011. Bio-hydrogen production by different operational modes of dark and photo-fermentation: An overview. *International Journal of Hydrogen Energy*, *36*(13), pp.7443–7459.

Arimi, M.M., Knodel, J., Kiprop, A., Namango, S.S., Zhang, Y. and Geißen, S.U., 2015. Strategies for improvement of biohydrogen production from organic-rich wastewater: A review. *Biomass and Bioenergy*, *75*, pp.101–118.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y. and Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, *44*(W1), pp.W16-W21.

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*(4), pp.202-206.

Atalah, J., Cáceres-Moreno, P., Espina, G. and Blamey, J.M., 2019. Thermophiles and the applications of their enzymes as new biocatalysts. *Bioresource Technology*, *280*, pp.478-488.

Athia, N., Pandey, M., Sen, M. and Saxena, S., 2024. Factors affecting the production cost of green hydrogen and its challenge for sustainable development. *Environment, Development and Sustainability*, pp.1-45.

Ausfelder, F. and Bazzanella, A. Hydrogen in the chemical industry. In Stolten, D. and Emonts, B. (Eds). Hydrogen science and engineering: materials, processes, systems and technology. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA, 2016, pp.19–40.

Ayub, H.M.U., Alnouri, S.Y., Stijepovic, M., Stijepovic, V. and Hussein, I.A., 2024. A cost comparison study for hydrogen production between conventional and renewable methods. *Process Safety and Environmental Protection*, *186*, pp.921-932.

Bacon, L.F., Hamley-Bennett, C., Danson, M.J. and Leak, D.J., 2017. Development of an efficient technique for gene deletion and allelic exchange in *Geobacillus* spp. *Microbial Cell Factories*, *16*(1),pp.1-8.

Balat, M. 2008. Potential importance of hydrogen as a future solution to environmental and transportation problems. *International Journal of Hydrogen Energy*, *33*(15), pp.4013–4029.

Balitskii, A., Krohmalny, O. and Ripey, I., 2000. Hydrogen cooling of turbogenerators and the problem of rotor retaining ring materials degradation. *International Journal of Hydrogen Energy*, *25*(2), pp.167–171.

Banat, I.M. and Marchant, R. *Geobacillus* activities in soil and oil contamination remediation. In Logan, N.A. and Vos, P. (Eds). *Endospore-forming Soil Bacteria*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp.259-270.

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*(5819), pp.1709-1712.

Bartels, J.R., Pate, M.B. and Olson, N.K., 2010. An economic survey of hydrogen production from conventional and alternative energy sources. *International Journal of Hydrogen Energy*, *35*(16), pp.8371–8384.

Ben Fekih, I., Zhang, C., Li, Y.P., Zhao, Y., Alwathnani, H.A., Saquib, Q., Rensing, C. and Cervantes, C., 2018. Distribution of arsenic resistance genes in prokaryotes. *Frontiers in Microbiology*, *9*, p.2473.

Benevenuti, C., Amaral, P., Ferreira, T. and Seidl, P., 2021. Impacts of syngas composition on anaerobic fermentation. *Reactions*, *2*(4), pp.391-407.

Benoit, S.L., Maier, R.J., Sawers, R.G. and Greening, C., 2020. Molecular hydrogen metabolism: a widespread trait of pathogenic bacteria and protists. *Microbiology and Molecular Biology Reviews*, *84*(1), pp.10-1128.

Bergonzi, C., Schwab, M. and Elias, M., 2016. The quorum-quenching lactonase from *Geobacillus caldoxylosilyticus*: purification, characterization, crystallization and crystallographic analysis. *Acta Crystallographica Section F: Structural Biology Communications*, *72*(9), pp.681-686.

Berney, M. and Cook, G.M., 2010. Unique flexibility in energy metabolism allows mycobacteria to combat starvation and hypoxia. *PloS one*, *5*(1), p.e8614.

Berney, M., Greening, C., Conrad, R., Jacobs Jr, W.R. and Cook, G.M., 2014. An obligately aerobic soil bacterium activates fermentative hydrogen production to survive reductive stress during hypoxia. *Proceedings of the National Academy of Sciences*, *111*(31), pp.11479–11484.

Bernhard, M., Schwartz, E., Rietdorf, J. and Friedrich, B., 1996. The *Alcaligenes eutrophus* membrane-bound hydrogenase gene locus encodes functions involved in maturation and electron transport coupling. *Journal of Bacteriology*, *178*(15), pp.4522–4529.

Bernheim, A. and Sorek, R., 2020. The pan-immune system of bacteria: antiviral defence as a community resource. *Nature Reviews Microbiology*, *18*(2), pp.113–119.

Bezalel-Buch, R., Cheun, Y.K., Roy, U., Schärer, O.D. and Burgers, P.M., 2020. Bypass of DNA interstrand crosslinks by a Rev1-DNA polymerase ζ complex. *Nucleic Acids Research*, *48*(15), pp.8461–8473.

Bhalla, A., Arce, J., Ubanwa, B., Singh, G., Sani, R.K. and Balan, V., 2022. Thermophilic *Geobacillus* WSUCF1 secretome for saccharification of ammonia fiber expansion and extractive ammonia pretreated corn stover. *Frontiers in Microbiology*, *13*, p.844287.

Bhandari, R., Trudewind, C.A. and Zapp, P., 2014. Life cycle assessment of hydrogen production via electrolysis: A review. *Journal of Cleaner Production*, *85*, pp.151–163.

Bhatia, L., Sarangi, P.K., Shadangi, K.P., Srivastava, R.K., Sahoo, U.K., Singh, A.K., Rene, E.R. and Kumar, B., 2024. A systematic review on photocatalytic biohydrogen production from waste biomass. *BioEnergy Research*, *17*(2), pp.932-955.

Blin, K., Shaw, S., Augustijn, H.E., Reitz, Z.L., Biermann, F., Alanjary, M., Fetter, A., Terlouw, B.R., Metcalf, W.W., Helfrich, E.J. and van Wezel, G.P., 2023. antiSMASH 7.0: New and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Research*, *51*(W1), pp.W46–W50.

Boonmak, C., Takahashi, Y. and Morikawa, M., 2014. Cloning and expression of three *ladA*-type alkane monooxygenase genes from an extremely thermophilic alkane-degrading bacterium *Geobacillus thermoleovorans* B23. *Extremophiles*, *18*, pp.515-523.

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.F., Lió, P., Crescenzi, P., Fani, R. and Fondi, M., 2015. MeDuSa: a multi-draft based scaffolder. *Bioinformatics*, *31*(15), pp.2443-2451.

BP, 2022. Statistical review of world energy 2022. London, UK: Energy Institute.

Broczkowski, M.E., Zagidulin, D. and Shoesmith, D.W. The role of dissolved hydrogen on the corrosion/dissolution of spent nuclear fuel. In Wai, C.M. and

Mincher, B.J. (Eds). ACS Symposium Series. Washington, DC: American Chemical Society, 2010, pp.349–380.

Brugna-Guiral, M., Tron, P., Nitschke, W., Stetter, K.O., Burlat, B., Guigliarelli, B., Bruschi, M. and Giudici-Orticoni, M.T., 2003. [NiFe] hydrogenases from the hyperthermophilic bacterium *Aquifex aeolicus*: properties, function, and phylogenetics. *Extremophiles*, *7*, pp.145–157.

Brugna-Guiral, M., Tron, P., Nitschke, W., Stetter, K.O., Burlat, B., Guigliarelli, B., Bruschi, M. and Giudici-Orticoni, M.T., 2003. [NiFe] hydrogenases from the hyperthermophilic bacterium *Aquifex aeolicus*: properties, function, and phylogenetics. *Extremophiles*, *7*, pp.145–157.

Brumm, P.J., Land, M.L. and Mead, D.A., 2015. Complete genome sequence of *Geobacillus thermoglucosidasius* C56-YS93, a novel biomass degrader isolated from obsidian hot spring in Yellowstone National Park. *Standards in Genomic Sciences*, *10*(1), pp.1-10.

Brumm, P.J., Land, M.L. and Mead, D.A., 2015. Complete genome sequence of *Geobacillus thermoglucosidasius* C56-YS93, a novel biomass degrader isolated from obsidian hot spring in Yellowstone National Park. *Standards in Genomic Sciences*, *10*(1), p.73.

Buchfink, B., Reuter, K. and Drost, H.G., 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), pp.366-368.

Bumgardner, E.A., Kittichotirat, W., Bumgarner, R.E. and Lawrence, P.K., 2015. Comparative genomic analysis of seven *Mycoplasma hyosynoviae* strains. *Microbiologyopen*, *4*(2), pp.343-359.

Burgdorf, T., Lenz, O., Buhrke, T., Van Der Linden, E., Jones, A.K., Albracht, S.P. and Friedrich, B., 2005. [NiFe]-Hydrogenases of *Ralstonia eutropha* H16: Modular enzymes for oxygen-tolerant biological hydrogen oxidation. *Microbial Physiology*, *10*(2–4), pp.181–196.

Byun, B., Mahasenan, K.V., Dik, D.A., Marous, D.R., Speri, E., Kumarasiri, M., Fisher, J.F., Hermoso, J.A. and Mobashery, S., 2018. Mechanism of the *Escherichia*

coli MltE lytic transglycosylase, the cell-wall-penetrating enzyme for Type VI secretion system assembly. *Scientific Reports*, *8*(1), p.4110.

Cai, J., Zhao, Y., Fan, J., Li, F., Feng, C., Guan, Y., Wang, R. and Tang, N., 2019. Photosynthetic bacteria improved hydrogen yield of combined dark- and photo-fermentation. *Journal of Biotechnology*, *302*, pp.18–25.

Campbell, K., Herrera-Dominguez, L., Correia-Melo, C., Zelezniak, A. and Ralser, M., 2018. Biochemical principles enabling metabolic cooperativity and phenotypic heterogeneity at the single cell level. *Current Opinion in Systems Biology*, *8*, pp.97-108.

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P. and Huerta-Cepas, J., 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution*, *38*(12), pp.5825-5829.

Carroll, A.C. and Wong, A., 2018. Plasmid persistence: costs, benefits, and the plasmid paradox. *Canadian Journal of Microbiology*, *64*(5), pp.293–304.

Cecere, D., Giacomazzi, E. and Ingenito, A., 2014. A review on hydrogen industrial aerospace applications. *International Journal of Hydrogen Energy*, *39*(20), pp.10731–10747.

Chai, Y., Kolter and R., Losick, R., 2009. A widely conserved gene cluster required for lactate utilization in *Bacillus subtilis* and its involvement in biofilm formation. *Journal of Bacteriology*, *191*(8), pp.2423–2430.

Chan, K.H., Lee, K.-M. and Wong, K.B., 2012. Interaction between hydrogenase maturation factors HypA and HypB is required for [NiFe]-hydrogenase maturation. *PloS one*, *7*(2), p.e32592.

Chang, C., Gong, S., Liu, Z., Yan, Q. and Jiang, Z., 2021. High level expression and biochemical characterization of an alkaline serine protease from *Geobacillus stearothermophilus* to prepare antihypertensive whey protein hydrolysate. *BMC Biotechnology*, *21*(1), pp.1-13.

Charbonneau, D.M., Meddeb-Mouelhi, F. and Beauregard, M., 2010. A novel thermostable carboxylesterase from *Geobacillus thermodenitrificans*: evidence for a new carboxylesterase family. *The Journal of Biochemistry*, *148*(3), pp.299-308.

Chaubey, R., Sahu, S., James, O.O. and Maity, S., 2013. A review on development of industrial processes and emerging techniques for production of hydrogen from renewable and sustainable sources. *Renewable and Sustainable Energy Reviews*, *23*, pp.443–462.

Chaudhari, N.M., Gupta, V.K. and Dutta, C., 2016. BPGA-an ultra-fast pan-genome analysis pipeline. *Scientific Reports*, *6*(1), p.24373.

Chen, W.H. and Chen, C.Y., 2020. Water gas shift reaction for hydrogen production and carbon dioxide capture: A review. *Applied Energy*, *258*, p.114078.

Chettri, D., Verma, A.K. and Verma, A.K., 2020. Innovations in CAZyme gene diversity and its modification for biorefinery applications. *Biotechnology Reports*, *28*, p.e00525.

Chiche, D., Diverchy, C., Lucquin, A.C., Porcheron, F. and Defoort, F., 2013. Synthesis Gas Purification. *Oil & Gas Science and Technology–Revue d'IFP Energies nouvelles*, *68*(4), pp.707-723.

Chopra, L., Singh, G., Choudhary, V. and Sahoo, D.K., 2014. Sonorensin: an antimicrobial peptide, belonging to the heterocycloanthracin subfamily of bacteriocins, from a new marine isolate, *Bacillus sonorensis* MT93. *Applied and Environmental Microbiology*, *80*(10), pp.2981-2990.

Chopra, L., Singh, G., Kumar Jena, K. and Sahoo, D.K., 2015. Sonorensin: a new bacteriocin with potential of an anti-biofilm agent and a food biopreservative. *Scientific Reports*, *5*(1), p.13412.

Chow, J.Y., Xue, B., Lee, K.H., Tung, A., Wu, L., Robinson, R.C. and Yew, W.S., 2010. Directed evolution of a thermostable quorum-quenching lactonase from the amidohydrolase superfamily. *Journal of Biological Chemistry*, *285*(52), pp.40911-40920.

Christopher, L.P., Zambare, V.P., Zambare, A., Kumar, H. and Malek, L., 2015. A thermo-alkaline lipase from a new thermophile *Geobacillus thermodenitrificans* AV-5 with potential application in biodiesel production. *Journal of Chemical Technology & Biotechnology*, *90*(11), pp.2007-2016.

Contesini, F.J., Frandsen, R.J.N. and Damasio, A., 2021. CAZymes in Biorefinery: from genes to application. *Frontiers in Bioengineering and Biotechnology*, *9*, p.622817.

Cordero, P.R., Grinter, R., Hards, K., Cryle, M.J., Warr, C.G., Cook, G.M. and Greening, C., 2019. Two uptake hydrogenases differentially interact with the aerobic respiratory chain during mycobacterial growth and persistence. *Journal of Biological Chemistry*, *294*(50), pp.18980–18991.

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E.P., Vergnaud, G., Gautheret, D. and Pourcel, C., 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research*, *46*(W1), pp.W246-W251.

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*(6), pp.398–408.

Cruz, F.D.L. and Brouns, S.J., 2013. CRISPR-Cas systems preferentially target the leading regions of MOBF conjugative plasmids. *RNA Biology*, 10(5), pp.749–761.

Cuebas, M., Sannino, D. and Bini, E., 2011. Isolation and characterization of arsenic resistant Geobacillus kaustophilus strain from geothermal soils. *Journal of Basic Microbiology*, *51*(4), pp.364-371.

Dalena, F., Senatore, A., Marino, A., Gordano, A., Basile, M. and Basile, A. Methanol production and applications: An overview. In Basile, A. and Dalena, F. (Eds). Methanol, Elsevier, 2018, pp. 3–28.

Daniels, J.A., Krishnamurthi, R. and Rizvi, S.S., 1985. A review of effects of carbon dioxide on microbial growth and food quality. *Journal of food protection*, *48*(6), pp.532-537.

Das, D. and Veziroğlu, T.N., 2001. Hydrogen production by biological processes: a survey of literature. *International Journal of Hydrogen Energy*, *26*, pp.13–28.

Das, D. and Veziroğlu, T.N., 2008. Advances in biological hydrogen production processes. *International Journal of Hydrogen Energy*, *33*(21), pp.6046–6057.

De Maayer, P., Brumm, P.J., Mead, D.A. and Cowan, D.A., 2014b. Comparative analysis of the *Geobacillus* hemicellulose utilization locus reveals a highly variable target for improved hemicellulolysis. *BMC Genomics*, *15*(1), pp.1-17.

De Maayer, P., Chan, W.Y., Rubagotti, E., Venter, S.N., Toth, I.K., Birch, P.R. and Coutinho, T.A., 2014a. Analysis of the *Pantoea* ananatis pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts. *BMC Genomics*, *15*, pp.1-14.

Demirbas, A., 2016. Comparison of thermochemical conversion processes of biomass to hydrogen-rich gas mixtures. *Energy sources, Part A: Recovery, utilization, and environmental effects*, *30*(20), pp.2971-2976.

Dey, A. and Neogi, S., 2019. Oxygen scavengers for food packaging applications: A review. *Trends in Food Science & Technology*, *90*, pp.26–34.

Díaz, D.B., Neumann, A. and Aliyu, H. (2022). Thermophilic water gas shift reaction at high carbon monoxide and hydrogen partial pressures in *Parageobacillus thermoglucosidasius* KP1013. *Fermentation*, *8*(11), p.596.

Dincer, I. and Acar, C., 2014. Review and evaluation of hydrogen production methods for better sustainability. *International Journal of Hydrogen Energy*, *40*(34), pp.11094–11111.

Dincer, I., 2012. Green methods for hydrogen production. *International Journal of Hydrogen Energy*, *37*(2), pp.1954–1971.

Dinesh, G.K., Chauhan, R. and Chakma, S., 2018. Influence and strategies for enhanced biohydrogen production from food waste. *Renewable and Sustainable Energy Reviews*, *92*, pp.807–822.

Ding, W., Zhang, Y. and Shi, S., 2020. Development and application of CRISPR/Cas in microbial biotechnology. *Frontiers in Bioengineering and Biotechnology*, *8*, p.711.

Drula, E., Garron, M.L., Dogan, S., Lombard, V., Henrissat, B. and Terrapon, N., 2022. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research*, *50*(D1), pp.D571-D577.

Dürre, P. and Eikmanns, B.J., 2015. C1-carbon sources for chemical and fuel production by microbial gas fermentation. *Current Opinion in Biotechnology*, *35*, pp.63–72.

Egan, K., Field, D., Ross, R.P., Cotter, P.D. and Hill, C., 2018. *In silico* prediction and exploration of potential bacteriocin gene clusters within the bacterial genus *Geobacillus*. *Frontiers in Microbiology*, *9*, p.2116.

Elbeshbishy, E., Dhar, B.R., Nakhla, G. and Lee, H.S., 2017. A critical review on inhibition of dark biohydrogen fermentation. *Renewable and Sustainable Energy Reviews*, *79*, pp.656-668.

Elumalai, P., Parthipan, P., Narenkumar, J., Anandakumar, B., Madhavan, J., Oh, B.T. and Rajasekar, A., 2019. Role of thermophilic bacteria (*Bacillus* and *Geobacillus*) on crude oil degradation and biocorrosion in oil reservoir environment. *3 Biotech*, *9*, pp.1-11.

Emms, D.M. and Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(238), pp.1-14.

Ezraty, B., Gennaris, A., Barras, F. and Collet, J.F., 2017. Oxidative stress, protein damage and repair in bacteria. *Nature Reviews Microbiology*, *15*(7), pp.385-396.

Falkenberg, F., Voß, L., Bott, M., Bongaerts, J. and Siegert, P., 2023. New robust subtilisins from halotolerant and halophilic Bacillaceae. *Applied Microbiology and Biotechnology*, *107*(12), pp.3939-3954.

Fayaz, A.M., Girilal, M., Rahman, M., Venkatesan, R. and Kalaichelvan, P.T., 2011. Biosynthesis of silver and gold nanoparticles using thermophilic bacterium *Geobacillus stearothermophilus*. *Process Biochemistry*, *46*(10), pp.1958-1962.

Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., Tang, Y., Liu, X., Han, W., Peng, X. and Liu, R., 2007. Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proceedings of the National Academy of Sciences*, *104*(13), pp.5602-5607.

Fernandez, V.M., Hatchikian, E.C. and Cammack, R., 1985. Properties and reactivation of two different deactivated forms of *Desulfovibrio gigas* hydrogenase. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, *832*(1), pp.69-79.

Fernandez, V.M., Hatchikian, E.C., Patil, D.S. and Cammack, R., 1986. ESR-detectable nickel and iron-sulphur centres in relation to the reversible activation of *Desulfovibrio gigas* hydrogenase. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *883*(1), pp.145-154.

Fiala, E.S., Conaway, C.C. and Mathis, J.E., 1989. Oxidative DNA and RNA damage in the livers of Sprague-Dawley rats treated with the hepatocarcinogen 2-nitropropane. *Cancer Research*, *49*(20), pp.5518-5522.

Finney, A.J. and Sargent, F. Formate hydrogenlyase: a group 4 [NiFe]-hydrogenase in tandem with a formate dehydrogenase. In Poole, R.K. (Ed). Advances in Microbial Physiology. Academic Press. 2019, *74*, pp.465-486.

Finore, I., Lama, L., Di Donato, P., Romano, I., Tramice, A., Leone, L., Nicolaus, B. and Poli, A., 2019. *Parageobacillus thermantarcticus*, an Antarctic cell factory: From crop residue valorization by green chemistry to astrobiology studies. *Diversity*, *11*(8), p.128.

Fischer, M. and Pleiss, J., 2003. The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Research*, *31*(1), pp.319-321.

Fortina, M.G., Mora, D., Schumann, P., Parini, C., Manachini, P.L. and Stackebrandt, E., 2001. Reclassification of *Saccharococcus caldoxylosilyticus* as *Geobacillus caldoxylosilyticus* (Ahmad *et al.* 2000) comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, *51*(6), pp.2063-2071.

Fox, J.D., Kerby, R.L., Roberts, G.P. and Ludden, P.W., 1996. Characterization of the CO-induced, CO-tolerant hydrogenase from *Rhodospirillum rubrum* and the gene encoding the large subunit of the enzyme. *Journal of Bacteriology*, *178*(6), pp.1515–1524.

Fritsch, J., Scheerer, P., Frielingsdorf, S., Kroschinsky, S., Friedrich, B., Lenz, O. and Spahn, C.M., 2011. The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre. *Nature*, *479*(7372), pp.249–252.

Fukuyama, Y., Inoue, M., Omae, K., Yoshida, T. and Sako, Y., 2020. Anaerobic and hydrogenogenic carbon monoxide-oxidizing prokaryotes: Versatile microbial conversion of a toxic gas into an available energy. *Advances in Applied Microbiology*, *110*, pp.99-148.

Gama, J.A., Zilhão, R. and Dionisio, F., 2018. Impact of plasmid interactions with the chromosome and other plasmids on the spread of antibiotic resistance. *Plasmid*, *99*, pp.82–88.

Garg, N., Tang, W., Goto, Y., Nair, S.K. and Van Der Donk, W.A., 2012. Lantibiotics from *Geobacillus thermodenitrificans*. *Proceedings of the National Academy of Sciences*, *109*(14), pp.5241-5246.

Ghiasian, M. Biophotolysis-based hydrogen production by Cyanobacteria. In Rastegari, A., Yadav, A., Gupta, A. (Eds). Prospects of renewable bioprocessing in future energy systems. Springer, Cham, 2019, pp.161–184.

Gielen, D., Boshell, F., Saygin, D., Bazilian, M.D., Wagner, N. and Gorini, R., 2019. The role of renewable energy in the global energy transformation. *Energy Strategy Reviews*, *24*, pp.38–50.

Giovanella, P., Vieira, G.A., Otero, I.V.R., Pellizzer, E.P., de Jesus Fontes, B. and Sette, L.D., 2020. Metal and organic pollutants bioremediation by extremophile microorganisms. *Journal of Hazardous Materials*, *382*, p.121024.

Greening, C. and Grinter, R., 2022. Microbial oxidation of atmospheric trace gases. *Nature Reviews Microbiology*, *20*(9), pp.513–528.

Greening, C., Berney, M., Hards, K., Cook, G.M. and Conrad, R., 2014a. A soil actinobacterium scavenges atmospheric H_2 using two membrane-associated, oxygen-dependent [NiFe] hydrogenases. *Proceedings of the National Academy of Sciences*, *111*(11), pp.4257–4261.

Greening, C., Biswas, A., Carere, C.R., Jackson, C.J., Taylor, M.C., Stott, M.B., Cook, G.M. and Morales, S.E., 2016. Genomic and metagenomic surveys of hydrogenase distribution indicate H_2 is a widely utilised energy source for microbial growth and survival. *ISME Journal*, *10*(3), pp.761–777.

Greening, C., Villas-Bôas, S.G., Robson, J.R., Berney, M. and Cook, G.M., 2014b. The growth and survival of *Mycobacterium smegmatis* is enhanced by co-metabolism of atmospheric H_2. *PLoS one*, *9*(7), p.e103034.

Grissa, I., Vergnaud, G. and Pourcel, C., 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research*, *35*(suppl_2), pp.W52-W57.

Haddad, M., Cimpoia, R. and Guiot, S.R., 2014. Performance of *Carboxydothermus hydrogenoformans* in a gas-lift reactor for syngas upgrading into hydrogen. I*nternational Journal of Hydrogen Energy*, *39*(6), pp.2543-2548.

Haft, D.H., 2009. A strain-variable bacteriocin in *Bacillus anthracis* and *Bacillus cereus* with repeated Cys-Xaa-Xaa motifs. *Biology Direct*, *4*, pp.1-5.

Hall, T.A., 1999. BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, pp. 95–98.

Hallenbeck, P.C. Bioenergy from microorganisms: An overview. In Zannoni, D. and De Philippis R. (Eds). Microbial BioEnergy: Hydrogen production. Dordrecht: Springer Netherlands, 2014, pp.3–21.

Hallenbeck, P.C., 2011. Microbial paths to renewable hydrogen production. *Biofuels*, *2*(3), pp.285–302.

Hamilton, S.C., Farchaus, J.W. and Davis, M.C., 2001. DNA polymerases as engines for biotechnology. *Biotechniques*, *31*(2), pp.370-383.

Han, H., Ling, Z., Khan, A., Virk, A.K., Kulshrestha, S. and Li, X., 2019. Improvements of thermophilic enzymes: From genetic modifications to applications. *Bioresource Technology*, *279*, pp.350-361.

Harrington, L.B., Paez-Espino, D., Staahl, B.T., Chen, J.S., Ma, E., Kyrpides, N.C. and Doudna, J.A., 2017. A thermostable Cas9 with increased lifetime in human plasma. *Nature Communications*, *8*(1), p.1424.

Harrison, D.E.F. and Loveless, J.E., 1971. Transient responses of facultatively anaerobic bacteria growing in chemostat culture to a change from anaerobic to aerobic conditions. *Microbiology*, *68*(1), pp.45-52.

Hay, J.X.W., Wu, T.Y., Juan, J.C. and Md. Jahim, J., 2013. Biohydrogen production through photo fermentation or dark fermentation using waste as a substrate: Overview, economics, and future prospects of hydrogen usage. *Biofuels, Bioproducts and Biorefining*, *7*(3), pp.334–352.

Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q. and Vinh, L.S., 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), pp.518-522.

Hogrefe, C., Römermann, D. and Friedrich, B., 1984. *Alcaligenes eutrophus* hydrogenase genes (Hox). *Journal of Bacteriology*, *158*(1), pp.43–48.

Holladay, J.D., Hu, J., King, D.L. and Wang, Y., 2009. An overview of hydrogen production technologies. *Catalysis Today*, *139*(4), pp.244–260.

Hosseini, S.E. and Wahid, M.A., 2016. Hydrogen production from renewable and sustainable energy resources: Promising green energy carrier for clean development. Renew. Sustain. *Renewable and Sustainable Energy Reviews*, *57*, pp.850-866.

Hosseini, S.E. and Wahid, M.A., 2020. Hydrogen from solar energy, a clean energy carrier from a sustainable source of energy. *International Journal of Energy Research*, *44*(6), pp.4110–4131.

Hosseini, S.E., Abdul Wahid, M., Jamil, M.M., Azli, A.A. and Misbah, M.F., 2015. A review on biomass-based hydrogen production for renewable energy supply. *International Journal of Energy Research*, *39*(12), pp.1597–1615.

Hosseini, S.S., Aghbashlo, M., Tabatabaei, M., Younesi, H. and Najafpour, G., 2015. Exergy analysis of biohydrogen production from various carbon sources via anaerobic photosynthetic bacteria (*Rhodospirillum rubrum*). *Energy, 93*, pp.730–739.

Houchins, J.P. and Burris, R.H., 1981. Occurrence and localization of two distinct hydrogenases in the heterocystous cyanobacterium *Anabaena* sp. strain 7120. *Journal of Bacteriology*, *146*(1), pp.209–214.

Hudson, G.A., Burkhart, B.J., DiCaprio, A.J., Schwalen, C.J., Kille, B., Pogorelov, T.V. and Mitchell, D.A., 2019. Bioinformatic mapping of radical S-adenosylmethionine-dependent ribosomally synthesized and post-translationally modified peptides identifies new Cα, Cβ, and Cγ-linked thioether-containing peptides. *Journal of the American Chemical Society*, *141*(20), pp.8228-8238.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J. and von Mering, C., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), pp.D309-D314.

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

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W. and Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(119), pp.1-11.

IEA, 2019a. The role of CO_2 storage. IEA, Paris. https://www.iea.org/reports/the-role-of-co2-storage.

IEA, 2019b. The future of hydrogen. IEA, Paris. https://www.iea.org/reports/the-future-of-hydrogen.

Imaura, Y., Okamoto, S., Hino, T., Ogami, Y., Katayama, Y.A., Tanimura, A., Inoue, M., Kamikawa, R., Yoshida, T. and Sako, Y., 2023. Isolation, genomic sequence and physiological characterization of *Parageobacillus* sp. G301, an isolate capable of both hydrogenogenic and aerobic carbon monoxide oxidation. *Applied and Environmental Microbiology*, 89, p.e00185-23.

Imlay, J.A., 2019. Where in the world do bacteria experience oxidative stress? *Environmental microbiology*, *21*(2), pp.521-530.

Inoue, M., Tanimura, A., Ogami, Y., Hino, T., Okunishi, S., Maeda, H., Yoshida, T. and Sako, Y., 2019. Draft genome sequence of *Parageobacillus thermoglucosidasius* strain TG4, a hydrogenogenic carboxydotrophic bacterium isolated from a marine sediment. *Microbiology Resource Announcements*, *8*(5), pp.10-1128.

IRENA, 2019. GLOBAL energy transformation: A roadmap to 2050 (2019 edition). International Renewable Energy Agency, Abu Dhabi.

Islam, Z.F., Cordero, P.R.F. and Greening, C., 2019. Putative iron-sulfur proteins are required for hydrogen consumption and enhance survival of Mycobacteria. *Frontiers in Microbiology*, *10*, p.2749.

Jang, J.S., Kang, D.O., Chun, M.J. and Byun, S.M., 1992. Molecular cloning of a subtilisin J gene from *Bacillus stearothermophilus* and its expression in *Bacillus subtilis*. *Biochemical and Biophysical Research Communications*, *184*(1), pp.277-282.

Ji, C., Wang, S. and Zhang, B., 2012. Performance of a hybrid hydrogen-gasoline engine under various operating conditions. *Applied Energy*, *97*, pp.584–589.

Kaczmarska, A., Pieczywek, P.M., Cybulska, J. and Zdunek, A., 2022. Structure and functionality of Rhamnogalacturonan I in the cell wall and in solution: A review. *Carbohydrate Polymers*, *278*, p.118909.

Kalita, M., Chutia, M., Jha, D.K. and Subrahmanyam, G., 2022. Mechanistic understanding of *Gordonia* sp. in biodesulfurization of organosulfur compounds. *Current Microbiology*, *79*(3), p.82.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., Von Haeseler, A. and Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), pp.587-589.

Kapdan, I.K. and Kargi, F., 2006. Bio-hydrogen production from waste materials. *Enzyme and Microbial Technology*, *38*(5), pp.569–582.

Kazlauskas, D., Krupovic, M., Guglielmini, J., Forterre, P. and Venclovas, Č., 2020. Diversity and evolution of B-family DNA polymerases. *Nucleic Acids Research*, *48*(18), pp.10142-10156.

Ke, Q., Chen, A., Minoda, M. and Yoshida, H., 2013. Safety evaluation of a thermolysin enzyme produced from *Geobacillus stearothermophilus*. *Food and Chemical Toxicology*, *59*, pp.541-548.

Kerby, R.L., Ludden, P.W. and Roberts, G.P., 1995. Carbon monoxide-dependent growth of *Rhodospirillum rubrum*. *Journal of Bacteriology*, *177*(8), pp. 2241–2244.

Khaswal, A., Chaturvedi, N., Mishra, S.K., Kumar, P.R. and Paul, P.K., 2022. Current status and applications of genus *Geobacillus* in the production of industrially important products—A review. *Folia Microbiologica*, *67*(3), pp.389-404.

Kim, S.M., Lee, J., Kang, S.H., Heo, Y., Yoon, H.J., Hahn, J.S., Lee, H.H. and Kim, Y.H., 2022. O_2-tolerant CO dehydrogenase via tunnel redesign for the removal of CO from industrial flue gas. *Nature Catalysis*, *5*(9), pp.807-817.

King, P.W. and Przybyla, A.E., 1999. Response of *hya* expression to external pH in *Escherichia coli*. *Journal of Bacteriology*, *181*(17), pp.5250–5256.

Kitts, P.A., Church, D.M., Thibaud-Nissen, F., Choi, J., Hem, V., Sapojnikov, V., Smith, R.G., Tatusova, T., Xiang, C., Zherikov, A. and DiCuccio, M., 2016. Assembly: a resource for assembled genomes at NCBI. *Nucleic Acids Research*, *44*(D1), pp.D73-D80.

Koch, H., Galushko, A., Albertsen, M., Schintlmeister, A., Gruber-Dorninger, C., Lücker, S., Pelletier, E., Le Paslier, D., Spieck, E., Richter, A. and Nielsen, P.H., 2014. Growth of nitrite-oxidizing bacteria by aerobic hydrogen oxidation. *Science*, *345*(6200), pp.1052–1054.

Koch, H., Galushko, A., Albertsen, M., Schintlmeister, A., Gruber-Dorninger, C., Lücker, S., Pelletier, E., Le Paslier, D., Spieck, E., Richter, A. and Nielsen, P.H., 2014. Growth of nitrite-oxidizing bacteria by aerobic hydrogen oxidation. *Science*, *345*(6200), pp.1052-1054.

Kodama, T., Takamatsu, H., Asai, K., Kobayashi, K., Ogasawara, N. and Watabe, K., 1999. The *Bacillus subtilis yaaH* gene is transcribed by SigE RNA polymerase during sporulation, and its product is involved in germination of spores. *Journal of Bacteriology*, *181*(15), pp.4584-4591.

Koku, H., Eroğlu, İ., Gündüz, U., Yücel, M. and Türker, L., 2002. Aspects of the metabolism of hydrogen production by *Rhodobacter sphaeroides*. *International Journal of Hydrogen Energy*, *27*, pp.1315–1329.

Kumar, A.N. and Mohan, S.V., 2018. Acidogenic valorization of vegetable waste for short chain carboxylic acids and biohydrogen production: Influence of pretreatment and pH. *Journal of Cleaner Production*, *203*, pp.1055–1066.

Kumar, R., Singh, L., Wahid, Z.A. and Din, M.F.M., 2015. Exoelectrogens in microbial fuel cells toward bioelectricity generation: A review. *International Journal of Energy Research*, *39*(8), pp.1048–1067.

Kumar, S., Dangi, A.K., Shukla, P., Baishya, D. and Khare, S.K., 2019. Thermozymes: adaptive strategies and tools for their biotechnological applications. *Bioresource Technology*, *278*, pp.372-382.

Labrou, N.E., 2010. Random mutagenesis methods for *in vitro* directed enzyme evolution. *Current Protein and Peptide Science*, *11*(1), pp.91-100.

Lanza, V.F., Tedim, A.P., Martínez, J.L., Baquero, F. and Coque, T.M., 2015. The plasmidome of Firmicutes: Impact on the emergence and the spread of resistance to antimicrobials. *Microbiology Spectrum*, *3*(2), pp.379-419.

Latif, H., Zeidan, A.A., Nielsen, A.T. and Zengler, K., 2014. Trash to treasure: Production of biofuels and commodity chemicals via syngas fermenting microorganisms. *Current Opinion in Biotechnology*, *27*, pp.79–87.

Lee, I., Ouk Kim, Y., Park, S.C. and Chun, J., 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *International Journal of Systematic and Evolutionary Microbiology*, *66*(2), pp.1100-1103.

Levin, B.R. and Bergstrom, C.T., 2000. Bacteria are different: Observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. *Proceedings of the National Academy of Sciences*, *97*(13), pp.6981–6985.

Levin, D.B., Pitt, L. and Love, M., 2004. Biohydrogen production: Prospects and limitations to practical application. *International Journal of Hydrogen Energy*, *29*(2), pp.173–185.

Li, Y., Gao, Z., Hou, H., Li, L., Zhang, J., Yang, H., Dong, Y. and Tan, H., 2011. Crystal structure and site-directed mutagenesis of a nitroalkane oxidase from *Streptomyces ansochromogenes*. *Biochemical and Biophysical Research Communications*, *405*(3), pp.344-348.

Liang, J., Roberts, A., van Kranenburg, R., Bolhuis, A. and Leak, D.J., 2022. Relaxed control of sugar utilization in *Parageobacillus thermoglucosidasius* DSM 2542. *Microbiological Research*, *256*, p.126957.

Liao, H., McKenzie, T. and Hageman, R., 1986. Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. *Proceedings of the National Academy of Sciences*, *83*(3), pp.576–580.

Lin, J.H., Zhang, K.C., Tao, W.Y., Wang, D. and Li, S., 2019. *Geobacillus* strains that have potential value in microbial enhanced oil recovery. *Applied Microbiology and Biotechnology*, *103*, pp.8339-8350.

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*, pp.1-8.

Liu, J., Han, X., Tao, F. and Xu, P., 2023. Reprogramming a high robust *Geobacillus thermoglucosidasius* for efficient synthesis of polymer-grade lactic acid under extremely high temperature (60°C). *bioRxiv*, p.2023-04.

Liu, W., Zuo, H., Wang, J., Xue, Q., Ren, B. and Yang, F., 2021. The production and application of hydrogen in steel industry. *International Journal of Hydrogen Energy*, *46*(17), pp.10548–10569.

Liu, Y., Lin, R., Man, Y. and Ren, J., 2019. Recent developments of hydrogen production from sewage sludge by biological and thermochemical process. *International Journal of Hydrogen Energy*, *44*(36), pp.19676–19697.

Loenen, W.A., Dryden, D.T., Raleigh, E.A., Wilson, G.G. and Murray, N.E., 2014. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Research*, *42*(1), pp.3-19.

Lu, Y. and Koo, J., 2019. O_2 sensitivity and H_2 production activity of hydrogenases—A review. *Biotechnology and Bioengineering*, *116*(11), pp.3124-3135.

Lucena-Aguilar, G., Sánchez-López, A.M., Barberán-Aceituno, C., Carrillo-Avila, J.A., López-Guerrero, J.A. and Aguilar-Quesada, R., 2016. DNA source selection for downstream applications based on DNA quality indicators analysis. *Biopreservation and Biobanking*, *14*(4), pp.264–270.

Luis Felipe Valdez-Nuñez and Rivera-Jacinto, M.A. (2022). Thermophilic bacteria from Peruvian hot springs with high potential application in environmental biotechnology. *Environmental Technology*, pp.1–16.

Ma, Y., Zhang, B., Wang, M., Ou, Y., Wang, J. and Li, S., 2016. Enhancement of polymerase activity of the large fragment in DNA polymerase I from *Geobacillus stearothermophilus* by site-directed mutagenesis at the active site. *BioMed Research International*, *2016*, p.2906484 .

Maeda, T., Sanchez-Torres, V. and Wood, T.K., 2007. Enhanced hydrogen production from glucose by metabolically engineered *Escherichia coli*. *Applied Microbiology and Biotechnology*, *77*(4), pp.879–890.

Maness, P.C., Huang, J., Smolinski, S., Tek, V. and Vanzin, G., 2005. Energy generation from the CO oxidation-hydrogen production pathway in *Rubrivivax gelatinosus*. *Applied and Environmental Microbiology*, *71*(6), pp.2870-2874.

Maness, P.C., Smolinski, S., Dillon, A.C., Heben, M.J. and Weaver, P.F., 2002. Characterization of the oxygen tolerance of a hydrogenase linked to a carbon monoxide oxidation pathway in *Rubrivivax gelatinosus*. *Applied and environmental microbiology*, *68*(6), pp.2633-2636.

Manish, S. and Banerjee, R., 2008. Comparison of biohydrogen production processes. *International Journal of Hydrogen Energy*, *33*(1), pp.279–286.

Manoharan, Y., Hosseini, S.E., Butler, B., Alzhahrani, H., Senior, B.T.F., Ashuri, T. and Krohn, J., 2019. Hydrogen fuel cell vehicles; Current status and future prospect. *Applied Sciences*, *9*(11), p.2296.

Marchant, R., Banat, I.M., Rahman, T.J. and Berzano, M., 2002. The frequency and characteristics of highly thermophilic bacteria in cool soil environments. *Environmental Microbiology*, *4*(10), pp.595-602.

Markets and Markets, 2022. Enzymes market by product type (Industrial enzymes and Specialty Enzymes), Source (Microorganism, Plant, and Animal), Type, Industrial Enzyme Application, Specialty Enzymes Application and Region - Global Forecast to 2027. Market research report, FB 7583.

Martínez, B., Fernández, M., Suárez, J.E. and Rodríguez, A., 1999. Synthesis of lactococcin 972, a bacteriocin produced by *Lactococcus lactis* IPLA 972, depends on the expression of a plasmid-encoded bicistronic operon. *Microbiology*, *145*(11), pp.3155-3161.

Mathur, H., C Rea, M., D Cotter, P., Hill, C. and Paul Ross, R., 2015. The sactibiotic subclass of bacteriocins: an update. *Current Protein and Peptide Science*, *16*(6), pp.549-558.

Matsumura, M.A., Katakura, Y.O., Imanaka, T.A. and Aiba, S.H., 1984. Enzymatic and nucleotide sequence studies of a kanamycin-inactivating enzyme encoded by a plasmid from thermophilic Bacilli in comparison with that encoded by plasmid pUB110. *Journal of Bacteriology*, 160(1), pp.413–420.

McBirney, S.E., Trinh, K., Wong-Beringer, A. and Armani, A.M., 2016. Wavelength-normalized spectroscopic analysis of *Staphylococcus aureus* and *Pseudomonas aeruginosa* growth rates. *Biomedical optics express*, *7*(10), pp.4034-4042.

McKenzie, T., Hoshino, T., Tanaka, T. and Sueoka, N., 1986. The nucleotide sequence of pUB110: Some salient features in relation to replication and its regulation. *Plasmid*, *15*(2), pp.93–103.

Meier-Kolthoff, J.P., Carbasse, J.S., Peinado-Olarte, R.L. and Göker, M., 2022. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Research*, *50*(D1), pp.D801-D807.

Menon, S. and Ragsdale, S.W., 1996. Unleashing hydrogenase activity in carbon monoxide dehydrogenase/acetyl-CoA synthase and pyruvate: ferredoxin oxidoreductase. *Biochemistry*, *35*(49), pp.15814-15821.

Merrouch, M., Hadj-Saïd, J., Domnik, L., Dobbek, H., Léger, C., Dementin, S. and Fourmond, V., 2015. O_2 Inhibition of Ni-containing CO dehydrogenase is partly reversible. *Chemistry–A European Journal*, *21*(52), pp.18934-18938.

Metcalfe, G.D., Sargent, F. and Hippler, M., 2022. Hydrogen production in the presence of oxygen by *Escherichia coli* K-12. *Microbiology*, *168*(3), p.001167.

Mohr, T., Aliyu, H., Biebinger, L., Gödert, R., Hornberger, A., Cowan, D., De Maayer, P. and Neumann, A., 2019. Effects of different operating parameters on hydrogen production by *Parageobacillus thermoglucosidasius* DSM 6285. *AMB Express*, *9*(1), pp.1-10.

Mohr, T., Aliyu, H., Küchlin, R., Polliack, S., Zwick, M., Neumann, A., Cowan, D. and De Maayer, P., 2018a. CO-dependent hydrogen production by the facultative anaerobe *Parageobacillus thermoglucosidasius*. *Microbial Cell Factories*, *17*(1), pp.1-12.

Mohr, T., Aliyu, H., Küchlin, R., Zwick, M., Cowan, D., Neumann, A. and De Maayer, P., 2018b. Comparative genomic analysis of *Parageobacillus thermoglucosidasius* strains with distinct hydrogenogenic capacities. *BMC Genomics*, *19*(1), pp.1-10.

Mol, V., Bennett, M., Sánchez, B.J., Lisowska, B.K., Herrgård, M.J., Nielsen, A.T., Leak, D.J. and Sonnenschein, N., 2021. Genome-scale metabolic modeling of *P. thermoglucosidasius* NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism. *Metabolic Engineering*, *65*, pp.123-134.

Montoro-García, S., Martínez-Martínez, I., Navarro-Fernández, J., Takami, H., García-Carmona, F. and Sánchez-Ferrer, A., 2009. Characterization of a novel thermostable carboxylesterase from *Geobacillus kaustophilus* HTA426 shows the existence of a new carboxylesterase family. *Journal of Bacteriology*, *191*(9), pp.3076-3085.

Morya, R., Andrianantenaina, F.H., Singh, S., Pandey, A.K., Kim, G.B., Verma, J.P., Kumar, G., Raj, T. and Kim, S.H., 2023. Exploring rice straw as substrate for hydrogen production: Critical challenges and opportunities. *Environmental Technology & Innovation*, p.103153.

Mohsin, M., Rasheed, A.K. and Saidur, R., 2018. Economic viability and production capacity of wind generated renewable hydrogen. *International Journal of Hydrogen Energy*, *43*(5), pp.2621-2630.

Mougiakos, I., Mohanraju, P., Bosma, E.F., Vrouwe, V., Finger Bou, M., Naduthodi, M.I., Gussak, A., Brinkman, R.B., Van Kranenburg, R. and Van Der Oost, J., 2017. Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nature Communications*, *8*(1), p.1647.

Muin, N.A.A., Isah, A.N., Asli, U.A., Sadikin, A.N., Norazahar, N., Kamaruddin, M.J., Hassim, M.H., Shin, H.W. and Azman, N.R., 2020. A short review on various purification techniques suitable for biohydrogen-mixed gases. *Journal of Energy and Safety Technology*, *3*(2).

Najafpour, G., Younesi, H. and Mohamed, A.R., 2006. A survey on various carbon sources for biological hydrogen production via the water-gas reaction using a photosynthetic bacterium (*Rhodospirillum rubrum*). *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, *28*(11), pp.1013–1026.

Najafpour, G.D. and Younesi, H., 2007. Bioconversion of synthesis gas to hydrogen using a light-dependent photosynthetic bacterium, *Rhodospirillum rubrum*. *World Journal of Microbiology and Biotechnology*, *23*(2), pp.275–284.

Najar, I.N. and Thakur, N., 2020. A systematic review of the genera *Geobacillus* and *Parageobacillus*: Their evolution, current taxonomic status and major applications. *Microbiology*, *166*(9), pp.800-816.

Nakashima, N. and Miyazaki, K., 2014. Bacterial cellular engineering by genome editing and gene silencing. *International Journal of Molecular Sciences*, *15*(2), pp.2773–2793.

Nanda, S., Li, K., Abatzoglou, N., Dalai, A.K. and Kozinski, J.A. Advancements and confinements in hydrogen production technologies. In Dalena, F., Basile, A. and Rossi, C. (Eds.) *Bioenergy systems for the future*. Elsevier, 2017, pp.373-418.

Nazina, T.N., Sokolova, D.S., Grigoryan, A.A., Shestakova, N.M., Mikhailova, E.M., Poltaraus, A.B., Tourova, T.P., Lysenko, A.M., Osipov, G.A. and Belyaev, S.S., 2005. *Geobacillus jurassicus* sp. nov., a new thermophilic bacterium isolated from a high-temperature petroleum reservoir, and the validation of the *Geobacillus* species. *Systematic and Applied Microbiology*, *28*(1), pp.43-53.

NEBuilder HiFi DNA Assembly tool v. 2.8.1 . Available at: https://nebuilder.neb.com (Accessed: 10 February 2024).

Netz, D.J.A., Pohl, R., Beck-Sickinger, A.G., Selmer, T., Pierik, A.J., de Freire Bastos, M.D.C. and Sahl, H.G., 2002. Biochemical characterisation and genetic analysis of aureocin A53, a new, atypical bacteriocin from *Staphylococcus aureus*. *Journal of Molecular Biology*, *319*(3), pp.745-756.

Ni, M., Leung, D.Y.C., Leung, M.K.H. and Sumathy, K., 2006. An overview of hydrogen production from biomass. *Fuel processing technology*, *87*(5), pp.461-472.

Nikolaidis, P. and Poullikkas, A., 2017. A comparative overview of hydrogen production processes. *Renewable and Sustainable Energy Reviews*, *67*, pp.597–611.

Novik, G., Savich, V. and Meerovskaya, O. *Geobacillus* bacteria: potential commercial applications in industry, bioremediation and bioenergy production. In Mishra, M. (Ed) Growing and Handling of Bacterial Cultures. IntechOpen, 2018, pp.1-36.

Olivier, J.G., Schure, K.M. and Peters, J.A.H.W., 2017. Trends in global CO_2 and total greenhouse gas emissions 2017. *PBL Netherlands Environmental Assessment Agency*, *5*, pp. 1-11.

Omae, K., Fukuyama, Y., Yasuda, H., Mise, K., Yoshida, T. and Sako, Y., 2019. Diversity and distribution of thermophilic hydrogenogenic carboxydotrophs revealed by microbial community analysis in sediments from multiple hydrothermal environments in Japan. *Archives of Microbiology*, *201*, pp.969-982.

Oni, A.O., Anaya, K., Giwa, T., Di Lullo, G. and Kumar, A., 2022. Comparative assessment of blue hydrogen from steam methane reforming, autothermal reforming, and natural gas decomposition technologies for natural gas-producing regions. *Energy Conversion and Management*, *254*, p.115245.

Orellana, R., Macaya, C., Bravo, G., Dorochesi, F., Cumsille, A., Valencia, R., Rojas, C. and Seeger, M., 2018. Living at the frontiers of life: extremophiles in Chile and their potential for bioremediation. *Frontiers in Microbiology*, *9*, p.2309.

Oscorbin, I. and Filipenko, M., 2023. Bst polymerase—a humble relative of Taq polymerase. *Computational and Structural Biotechnology Journal*, 21, pp.4519–4535.

Oscorbin, I.P., Boyarskikh, U.A. and Filipenko, M.L. (2015). Large fragment of DNA polymerase I from *Geobacillus* sp. 777: Cloning and comparison with DNA polymerases I in practical applications. *Molecular Biotechnology*, 57(10), pp.947–959.

Osman, A.I., Deka, T.J., Baruah, D.C. and Rooney, D.W., 2020. Critical challenges in biohydrogen production processes from the organic feedstocks. *Biomass Conversion and Biorefinery*, *13*, pp.8383–8401.

Otoupal, P.B. and Chatterjee, A., 2018. CRISPR gene perturbations provide insights for improving bacterial biofuel tolerance. *Frontiers in Bioengineering and Biotechnology*, *6*, p.122.

Pachapur, V.L., Jyoti Sarma, S., Kaur Brar, S., Le Bihan, Y., Ricardo Soccol, C., Buelna, G. and Verma, M., 2015. Co-culture strategies for increased biohydrogen production. *International Journal of Energy Research*, *39*(11), pp.1479-1504.

Padmanabhan, S., Banerjee, S. and Mandi, N. Screening of bacterial recombinants: Strategies and preventing false positives. In Brown, G.G. (Ed). Molecular Cloning - Selected Applications in Medicine and Biology. IntechOpen, 2011, pp.1-19.

Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E. and Larsson, D.J., 2014. BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Research*, *42*(D1), pp.D737-D743.

Pandelia, M.E., Lubitz, W. and Nitschke, W., 2012. Evolution and diversification of Group 1 [NiFe] hydrogenases. Is there a phylogenetic marker for O_2-tolerance? *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1817*(9), pp.1565–1575.

Pareek, A., Dom, R., Gupta, J., Chandran, J., Adepu, V. and Borse, P.H., 2020. Insights into renewable hydrogen energy: Recent advances and prospects. *Materials Science for Energy Technologies*, *3*, pp.319–327.

Parte, A.C., Sardà Carbasse, J., Meier-Kolthoff, J.P., Reimer, L.C. and Göker, M., 2020. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology*, *70*(11), pp.5607-5612.

Parte, A.C., Sardà Carbasse, J., Meier-Kolthoff, J.P., Reimer, L.C. and Göker, M., 2020. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology*, *70*(11), pp.5607–5612.

Parthasarathy, P. and Narayanan, K.S., 2014. Hydrogen production from steam gasification of biomass: Influence of process parameters on hydrogen yield – A review. *Renewable Energy*, *66*, pp.570–579.

Partridge, J.D., Scott, C., Tang, Y., Poole, R.K. and Green, J., 2006. *Escherichia coli* transcriptome dynamics during the transition from anaerobic to aerobic conditions. *Journal of Biological Chemistry*, *281*(38), pp.27806-27815.

Peng, C., Huang, D., Shi, Y., Zhang, B., Sun, L., Li, M., Deng, X. and Wang, W., 2019. Comparative transcriptomic analysis revealed the key pathways responsible for organic sulfur removal by thermophilic bacterium *Geobacillus thermoglucosidasius* W-2. *Science of the Total Environment*, *676*, pp.639-650.

Peng, C., Shi, Y., Wang, S., Zhang, J., Wan, X., Yin, Y., Wang, D. and Wang, W., 2023. Genetic and functional characterization of multiple thermophilic organosulfur-removal systems reveals desulfurization potentials for waste residue oil cleaning. *Journal of Hazardous Materials*, *446*, p.130706.

Pinhal, S., Ropers, D., Geiselmann, J. and de Jong, H., 2019. Acetate metabolism and the inhibition of bacterial growth by acetate. *Journal of Bacteriology*, *201*(13), pp.10-1128.

Pisa, K.Y., Huber, H., Thomm, M. and Müller, V., 2007. A sodium ion-dependent A_1A_O ATP synthase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *The FEBS Journal*, *274*(15), pp.3928-3938.

Placier, G., Watzlawick, H., Rabiller, C. and Mattes, R., 2009. Evolved *β*-galactosidases from *Geobacillus stearothermophilus* with improved transgalactosylation yield for galacto-oligosaccharide production. *Applied and Environmental Microbiology*, *75*(19), pp.6312-6321.

Poli, A., Romano, I., Caliendo, G., Nicolaus, G., Orlando, P., De Falco, A., Lama, L., Gambacorta, A. and Nicolaus, B., 2006. *Geobacillus toebii* subsp. *decanicus* subsp. nov., a hydrocarbon-degrading, heavy metal resistant bacterium from hot compost. *The Journal of General and Applied Microbiology*, *52*(4), pp.223-234.

Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L. and Severinov, K., 2010. Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Molecular Microbiology*, *77*(6), pp.1367-1379.

Puopolo, R., Gallo, G., Mormone, A., Limauro, D., Contursi, P., Piochi, M., Bartolucci, S. and Fiorentino, G., 2020. Identification of a new heavy-metal-resistant strain of *Geobacillus stearothermophilus* isolated from a hydrothermally active volcanic area in southern Italy. *International Journal of Environmental Research and Public Health*, *17*(8), p.2678.

Qureshi, F., Yusuf, M., Tahir, M., Haq, M., Mohamed, M.M.I., Kamyab, H., Nguyen, H.H.T., Vo, D.V.N. and Ibrahim, H., 2023. Renewable hydrogen production via biological and thermochemical routes: nanomaterials, economic analysis and challenges. *Process Safety and Environmental Protection*, *179*, pp.68-88.

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Ragsdale, S.W., 2004. Life with carbon monoxide. *Critical Reviews in Biochemistry and Molecular Biology*, *39*(3), pp.165-195.

Rai, R., Bibra, M., Chadha, B.S. and Sani, R.K., 2019. Enhanced hydrolysis of lignocellulosic biomass with doping of a highly thermostable recombinant laccase. *International Journal of Biological Macromolecules*, *137*, pp.232-237.

Rajer, F. and Sandegren, L., 2022. The role of antibiotic resistance genes in the fitness cost of multiresistance plasmids. *MBio*, *13*(1), pp.e03552-21.

Ramachandran, R. and Menon, R.K., 1998. An overview of industrial uses of hydrogen. *International journal of hydrogen energy*, *23*(7), pp.593-598.

Ramachandriya, K., Kundiyana, D., Sharma, A., Kumar, A., Atiyeh, H., Huhnke, R. and Wilkins, M., 2016. Critical factors affecting the integration of biomass gasification and syngas fermentation technology. *AIMS Bioengineering*, *3*(2): pp.188-210.

Rasul, M.G., Hazrat, M.A., Sattar, M.A., Jahirul, M.I. and Shearer, M.J., 2022. The future of hydrogen: Challenges on production, storage and applications. *Energy Conversion and Management*, *272*, p.116326.

Rathore, D., Singh, A., Dahiya, D. and Nigam, P.S.N., 2019. Sustainability of biohydrogen as fuel: Present scenario and future perspective. *AIMS Energy*, *7*(1), pp.1–19.

Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A. and Finn, R.D., 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*, *46*(D1), pp.D624-D632.

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*(12), pp.1342–1347.

REN21, 2019. Global status report. Paris. France. https://www.ren21.net/gsr-2019.

Robb, F.T. and Techtmann, S.M., 2018. Life on the fringe: microbial adaptation to growth on carbon monoxide. F1000Research, *7*.

Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D., 2015. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Research*, *43*(D1), pp.D298-D299.

Rodríguez-Beltrán, J., DelaFuente, J., Leon-Sampedro, R., MacLean, R.C. and San Millan, A., 2021. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nature Reviews Microbiology*, *19*(6), pp.347–359.

Rossiter, S.E., Fletcher, M.H. and Wuest, W.M., 2017. Natural products as platforms to overcome antibiotic resistance. *Chemical Reviews*, *117*(19), pp.12415-12474.

Russell, J.B., 1992. Another explanation for the toxicity of fermentation acids at low pH: anion accumulation versus uncoupling. *Journal of Applied Bacteriology*, *73*(5), pp.363-370.

Russell, J.B. and Cook, G.M., 1995. Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiological reviews*, *59*(1), pp.48-62.

Sahu, R., Song, B.J., Im, J.S., Jeon, Y.P. and Lee, C.W., 2015. A review of recent advances in catalytic hydrocracking of heavy residues. *Journal of Industrial and Engineering Chemistry*, *27*, pp.12–24.

Sandalli, C., Singh, K., Modak, M.J., Ketkar, A., Canakci, S., Demir, İ. and Belduz, A.O., 2009. A new DNA polymerase I from *Geobacillus caldoxylosilyticus* TK4: cloning, characterization, and mutational analysis of two aromatic residues. *Applied Microbiology and Biotechnology*, *84*, pp.105-117.

Sander, J.D. and Joung, J.K., 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, *32*(4), pp.347-355.

Sarangi, P.K. and Nanda, S., 2020. Biohydrogen production through dark fermentation. *Chemical Engineering and Technology*, *43*(4), pp.601–612.

Sargent, F. The Model [NiFe]-Hydrogenases of *Escherichia coli*. In Poole, R.K. and Kelly, D.L. (Eds). Advances in Microbial Physiology. Academic Press, 2016, *68*, pp.433–507.

Sarkar, O., Butti, S.K. and Venkata Mohan, S., 2017. Acidogenesis driven by hydrogen partial pressure towards bioethanol production through fatty acids reduction. *Energy*, *118*, pp.425–434.

Sarkar, S., Banerjee, R., Chanda, S., Das, P., Ganguly, S. and Pal, S., 2010. Effectiveness of inoculation with isolated *Geobacillus* strains in the thermophilic stage of vegetable waste composting. *Bioresource Technology*, *101*(8), pp.2892-2895.

Sawers, G., Böck, A., 1988. Anaerobic regulation of pyruvate formate-lyase from *Escherichia coli* K-12. *Journal of Bacteriology*, *170*(11), pp.5330-5336.

Sawers, R.G., Jamieson, D.J., Higgins, C.F. and Boxer, D.H., 1986. Characterization and physiological roles of membrane-bound hydrogenase isoenzymes from *Salmonella typhimurium*. *Journal of Bacteriology*, *168*(1), pp.398–404.

Schneider, J., Yepes, A., Garcia-Betancur, J.C., Westedt, I., Mielich, B. and López, D., 2012. Streptomycin-induced expression in *Bacillus subtilis* of YtnP, a lactonase-homologous protein that inhibits development and streptomycin production in *Streptomyces griseus*. *Applied and Environmental Microbiology*, *78*(2), pp.599-603.

Schoelmerich, M.C., Müller, V., 2019. Energy conservation by a hydrogenase-dependent chemiosmotic mechanism in an ancient metabolic pathway. *Proceedings of the National Academy of Sciences*, *116*(13), pp.6329-6334.

Schwartz, E., Fritsch, J. and Friedrich, B. H_2-metabolizing prokaryotes. In Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E. and Thompson, F. (Eds). The Prokaryotes: Prokaryotic Physiology and Biochemistry. Springer-Verlag Berlin Heidelberg, 2013, pp.119–199.

Sharma, P., Kumar, R. and Capalash, N., 2013. Restriction enzymes from thermophiles. In Satyanarayana, T., Littlechild, J., and Kawarabayasi, Y. (Eds). *Thermophilic Microbes in Environmental and Industrial Biotechnology: Biotechnology of Thermophiles*. Springer, Dordrecht, 2013, pp.611-647.

Sharma, R.C. and Schimke, R.T., 1996. Preparation of electro-competent *E. coli* using salt-free growth medium. *BioTechniques*, *20*(1), pp.42–44.

Sharma, S., Basu, S., Shetti, N.P. and Aminabhavi, T.M., 2020. Waste-to-energy nexus for circular economy and environmental protection: Recent trends in hydrogen energy. *Science of the Total Environment*, *713*, p.136633.

Shen, Q., Fu, C., Wang, J., Yao, W., Wu, T., Ding, S. and Xu, P., 2024. Exergy-cost-carbon nexus of power-to-X system from carbon dioxide/water co-electrolysis driven by solar full-spectrum energy. *Energy Conversion and Management*, *308*, p.118382.

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*(1), pp.1-18.

Sheng, L., Madika, A., Lau, M.S., Zhang, Y. and Minton, N.P., 2023. Metabolic engineering for the production of acetoin and 2, 3-butanediol at elevated temperature in *Parageobacillus thermoglucosidasius* NCIMB 11955. *Frontiers in Bioengineering and Biotechnology*, *11*, p.1191079.

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), p.10-1128.

Shi, Q. and Wu, J., 2021. Review on sulfur compounds in petroleum and its products: State-of-the-art and perspectives. *Energy & Fuels*, *35*(18), pp.14445-14461.

Shivakumar, A.G. and Dubnau, D., 1978. Plasmid replication in DNA Ts mutants of *Bacillus subtilis*. *Plasmid*, *1*(3), pp.405–416.

Shokrollahi, M., Teymouri, N., Ashrafi, O., Navarri, P. and Khojasteh-Salkuyeh, Y., 2024. Methane pyrolysis as a potential game changer for hydrogen economy: Techno-economic assessment and GHG emissions. *International Journal of Hydrogen Energy*, *66*, pp.337-353.

Shomura, Y., 2019. Structural insights into the protective mechanisms against O_2 in the [NiFe]-hydrogenases, In Hirao, T. and Moriuchi, T. (Eds). Advances in Bioorganometallic Chemistry. Elsevier. 2019, pp.365-377.

Show, K.Y., Yan, Y., Ling, M., Ye, G., Li, T. and Lee, D.J., 2018. Hydrogen production from algal biomass – Advances, challenges and prospects. *Bioresource Technology*, *257*, pp.290–300.

Simon, K.M.R., Lee, H.S., Lim, J.K., Kim, T.W., Lee, J.H. and Kang, S.G., 2015. One-carbon substrate-based biohydrogen production: Microbes, mechanism, and productivity. *Biotechnology Advances*, *33*(1), pp.165–177.

Singh, L. and Wahid, Z.A., 2015. Methods for enhancing bio-hydrogen production from biological process: A review. *Journal of Industrial and Engineering Chemistry*, *21*, pp.70–80.

Singh, R. and Dutta, S., 2018. A review on H_2 production through photocatalytic reactions using TiO_2/TiO_2-assisted catalysts. *Fuel*, *220*, pp.607–620.

Singh, R., Kumar, M., Mittal, A. and Mehta, P.K., 2016. Microbial enzymes: industrial progress in 21st century. *3 Biotech*, *6*, pp.1-15.

Soboh, B., Linder, D. and Hedderich, R., 2002. Purification and catalytic properties of a CO-oxidizing:H_2-evolving enzyme complex from *Carboxydothermus hydrogenoformans*. *European Journal of Biochemistry*, *269*(22), pp.5712-5721.

Søndergaard, D., Pedersen, C.N.S. and Greening, C., 2016. HydDB: A web tool for hydrogenase classification and analysis. *Scientific Reports*, *6*(1), p.34212.

Sorgulu, F. and Dincer, I., 2018. Cost evaluation of two potential nuclear power plants for hydrogen production. *International Journal of Hydrogen Energy*, *43*(23), pp.10522-10529.

Staffell, I., Scamman, D., Abad, A.V., Balcombe, P., Dodds, P.E., Ekins, P., Shah, N. and Ward, K.R., 2019. The role of hydrogen and fuel cells in the global energy system. *Energy and Environmental Science*, *12*(2), pp.463–491.

Stam, M.R., Danchin, E.G., Rancurel, C., Coutinho, P.M. and Henrissat, B., 2006. Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins. *Protein Engineering, Design and Selection*, *19*(12), pp.555-562.

Stiebritz, M.T. and Reiher, M., 2012. Hydrogenases and oxygen. *Chemical Science*, *3*(6), pp.1739-1751.

Sun, L., Huang, D., Zhu, L., Zhang, B., Peng, C., Ma, T., Deng, X., Wu, J. and Wang, W., 2019. Novel thermostable enzymes from *Geobacillus thermoglucosidasius* W-2 for high-efficient nitroalkane removal under aerobic and anaerobic conditions. *Bioresource Technology*, *278*, pp.73-81.

Sun, S., Wan, X., Bian, Y., Wang, S., Zhang, J. and Wang, W., 2023. Biologically evolved dual-pathway catalytic pattern indicating an efficient bioremediation strategy for phenol removal. *Chemical Engineering Journal*, *454*, p.140195.

Sun, Y., Ning, Z., Yang, F. and Li, X., 2015. Characteristics of newly isolated *Geobacillus* sp. ZY-10 degrading hydrocarbons in crude oil. *Polish Journal of Microbiology*, *64*(3), pp.253-263.

Sung, M.H., Kim, H., Bae, J.W., Rhee, S.K., Jeon, C.O., Kim, K., Kim, J.J., Hong, S.P., Lee, S.G., Yoon, J.H. and Park, Y.H., 2002. *Geobacillus toebii* sp. nov., a novel thermophilic bacterium isolated from hay compost. *International Journal of Systematic and Evolutionary Microbiology*, *52*(6), pp.2251-2255.

Svetlichny, V.A., Sokolova, T.G., Gerhardt, M., Ringpfeil, M., Kostrikina, N.A. and Zavarzin, G.A., 1991. *Carboxydothermus hydrogenoformans* gen. nov., sp. nov., a CO-utilizing thermophilic anaerobic bacterium from hydrothermal environments of Kunashir Island. *Systematic and Applied Microbiology*, *14*(3), pp.254–260.

Talavera, G. and Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, *56*(4), pp.564-577.

Tang, Y.J., Sapra, R., Joyner, D., Hazen, T.C., Myers, S., Reichmuth, D., Blanch, H. and Keasling, J.D., 2008. Analysis of metabolic pathways and fluxes in a newly

discovered thermophilic and ethanol-tolerant *Geobacillus* strain. *Biotechnology and. Bioengineering*, 102(5), pp.1377–1386.

Teixeira, M., Moura, I., Xavier, A.V., Moura, J.J., LeGall, J., DerVartanian, D.V., Peck, H.D. and Huynh, B.H., 1989. Redox intermediates of *Desulfovibrio gigas* [NiFe] Hydrogenase generated under hydrogen: Mössbauer and EPR characterization of the metal centers. *Journal of Biological Chemistry*, *264*(28), pp.16435-16450.

Terpe, K., 2013. Overview of thermostable DNA polymerases for classical PCR applications: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, *97*(24), pp.10243-10254.

Thomas, C.M. and Nielsen, K.M., 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology*, *3*(9), pp.711–721.

Thompson, J.D., Higgins, D.G. and Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), pp.4673–4680.

Timinskas, K. and Venclovas, Č., 2019. New insights into the structures and interactions of bacterial Y-family DNA polymerases. *Nucleic Acids Research*, *47*(9), pp.4393-4405.

Tourova, T.P., Sokolova, D.S., Semenova, E.M., Poltaraus, A.B. and Nazina, T.N., 2018. Diversity of the *alkB* genes of n-alkane biodegradation in thermophilic hydrocarbon-oxidizing bacteria of the genera *Geobacillus*, *Parageobacillus*, and *Aeribacillus*. *Microbiology*, *87*, pp.301-307.

Tourova, T.P., Sokolova, D.S., Semenova, E.M., Shumkova, E.S., Korshunova, A.V., Babich, T.L., Poltaraus, A.B. and Nazina, T.N., 2016. Detection of n-alkane biodegradation genes *alkB* and *ladA* in thermophilic hydrocarbon-oxidizing bacteria of the genera *Aeribacillus* and *Geobacillus*. *Microbiology*, *85*, pp.693-707.

Trasanidou, D., Potocnik, A., Barendse, P., Mohanraju, P., Bouzetos, E., Karpouzis, E., Desmet, A., van Kranenburg, R., van der Oost, J., Staals, R.H. and Mougiakos,

I., 2023. Characterization of the AcrIIC1 anti–CRISPR protein for Cas9–based genome engineering in *E. coli*. *Communications Biology*, *6*(1), p.1042.

Unden, G., 1998. Transcriptional regulation and energetics of alternative respiratory pathways in facultatively anaerobic bacteria. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, *1365*(1-2), pp.220-224.

UNFCCC, 2015. United Nations Framework Convention on Climate Change. Paris Agreement. Available at: https://unfccc.int/files/meetings/paris_nov_2015/application/pdf/paris_agreemen

Valladares Juarez, A.G., Dreyer, J., Göpel, P.K., Koschke, N., Frank, D., Märkl, H. and Müller, R., 2009. Characterisation of a new thermoalkaliphilic bacterium for the production of high-quality hemp fibres, *Geobacillus thermoglucosidasius* strain PB94A. *Applied Microbiology and Biotechnology*, *83*(3), pp.521–527.

van Heel, A.J., de Jong, A., Song, C., Viel, J.H., Kok, J. and Kuipers, O.P., 2018. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Research*, *46*(W1), pp.W278-W281.

van Niel, E.W.J. Biological processes for hydrogen production. In Biological Processes for Hydrogen Production. In Hatti-Kaul, R., Mamo, G. and Mattiasson, B. (Eds). Anaerobes in Biotechnology. Advances in Biochemical Engineering/Biotechnology. Springer, Cham, 2016, 156, pp.155-193.

van Zyl, L.J., Taylor, M.P., Eley, K., Tuffin, M. and Cowan, D.A., 2014. Engineering pyruvate decarboxylase-mediated ethanol production in the thermophilic host *Geobacillus thermoglucosidasius*. *Applied Microbiology and Biotechnology*, *98*, pp.1247-1259.

Veziroğlu, T.N. and Şahin, S., 2008. 21st Century's energy: Hydrogen energy system. *Energy Conversion and Management*, *49*(7), pp.1820–1831.

Vignais, P.M. and Billoud, B., 2007. Occurrence, classification, and biological function of hydrogenases: An overview. *Chemical Reviews*, *107*(10), pp.4206-4272.

Volbeda, A., Amara, P., Darnault, C., Mouesca, J.M., Parkin, A., Roessler, M.M., Armstrong, F.A. and Fontecilla-Camps, J.C., 2012. X-ray crystallographic and computational studies of the O_2-tolerant [NiFe]-hydrogenase 1 from *Escherichia coli*. *Proceedings of the National Academy of Sciences*, *109*(14), pp.5305–5310.

von Haeseler, A., Schmidt, H.A., Bui, M.Q. and Nguyen, L.T., 2014. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution, 32*(1), pp.268-274.

Wallace, I.M., O'sullivan, O., Higgins, D.G. and Notredame, C., 2006. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Research*, *34*(6), pp.1692-1699.

Wang, F. and Yang, W., 2009. Structural insight into translesion synthesis by DNA Pol II. *Cell*, *139*(7), pp.1279-1289.

Wang, M., Zhu, H., Kong, Z., Li, T., Ma, L., Liu, D. and Shen, Q., 2020. Pan-genome analyses of *Geobacillus* spp. reveal genetic characteristics and composting potential. *International Journal of Molecular Sciences*, *21*(9), p.3393.

Wang, Y., Huang, C. and Zhao, W., 2022. Recent advances of the biological and biomedical applications of CRISPR/Cas systems. *Molecular Biology Reports*, *49*(7), pp.7087-7100.

Weiss, B., Jacquemin-Sablon, A., Live, T.R., Fareed, G.C. and Richardson, C.C., (1968). Enzymatic Breakage and Joining of Deoxyribonucleic Acid: VI. Further purification and properties of polynucleotide ligase from *Escherichia coli* infected with bacteriophage T4. *Journal of Biological Chemistry*, *243*(17), pp.4543–4555.

Welte, C., Krätzer, C. and Deppenmeier, U., 2010. Involvement of Ech hydrogenase in energy conservation of *Methanosarcina mazei*. *The FEBS Journal*, *277*(16), pp.3396-3403.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L. and Hester, J., 2019. Welcome to the Tidyverse *Journal of Open Source Software*, *4*(43), p.1686.

Winter, C.J. (2009). Hydrogen energy - Abundant, efficient, clean: A debate over the energy-system-of-change. *International Journal of Hydrogen Energy*, *34*(14), pp.S1-S52.

Wolfe, A.J., 2005. The acetate switch. *Microbiology and Molecular Biology Reviews*, *69*(1), pp.12-50.

Yanisch-Perron, C., Vieira, J. and Messing, J., 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene*, *33*(1), pp.103–119.

Yasin, M., Cha, M., Chang, I.S., Atiyeh, H.K., Munasinghe, P. and Khanal, S.K.. Syngas fermentation into biofuels and biochemicals. In Pandey, A. Larroche, C., Dussap, C., Gnansounou, E., Khanal, S. and Ricke, S. (Eds). Biomass, Biofuels, Biochemicals: Biofuels: Alternative Feedstocks and Conversion Processes for the Production of Liquid and Gaseous Biofuels (Second edition). Elsevier, 2019, pp.301–327.

Yokoi, H., Mori, S., Hirose, J., Hayashi, S. and Takasaki, Y., 1998. H_2 production from starch by a mixed culture of *Clostridium butyricum* and *Rhodobacter* sp. M[h]19. *Biotechnology Letters*, *20*(9), pp.895–899.

Younesi, H., Najafpour, G., Ismail, K.S.K., Mohamed, A.R. and Kamaruddin, A.H., 2008. Biohydrogen production in a continuous stirred tank bioreactor from synthesis gas by anaerobic photosynthetic bacterium: *Rhodopirillum rubrum*. *Bioresource Technology*, *99*(7), pp.2612–2619.

Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J. and Brinkman, F.S., 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, *26*(13), pp.1608-1615.

Zagrodnik, R. and Duber, A., 2024. Continuous dark-photo fermentative H_2 production from synthetic lignocellulose hydrolysate with different photoheterotrophic cultures: Sequential vs. co-culture processes. *Energy*, *290*, p.130105.

Zagrodnik, R. and Łaniecki, M., 2017. Hydrogen production from starch by co-culture of *Clostridium acetobutylicum* and *Rhodobacter sphaeroides* in one step hybrid dark- and photofermentation in repeated fed-batch reactor. *Bioresource Technology*, *224*, pp.298–306.

Zebrowska, J., Witkowska, M., Struck, A., Laszuk, P.E., Raczuk, E., Ponikowska, M., Skowron, P.M. and Zylicz-Stachula, A., 2022. Antimicrobial potential of the genera *Geobacillus* and *Parageobacillus*, as well as endolysins biosynthesized by their bacteriophages. *Antibiotics*, *11*(2), p.242.

Zeigler, D.R., 2014. The *Geobacillus* paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? *Microbiology*, *160*(1), pp.1-11.

Zhang, F., Zhao, P., Niu, M. and Maddy, J., 2016. The survey of key technologies in hydrogen energy storage. *International Journal of Hydrogen Energy*, *41*(33), pp.14535–14552.

Zhang, X., Huo, L., Yao, Z., Xie, T., Sun, Y., Liu, Z., Zhao, Y. and Zhao, L., 2024. A comparative review on the progress of thermochemical hydrogen production technologies from agricultural waste. *Fuel*, *366*, p.131284.

Zhang, Y. and Angelidaki, I., 2014. Microbial electrolysis cells turning to be versatile technology: Recent advances and future challenges. *Water Research*, *56*, pp.11–25.

Zhao, Y., Haddad, M., Cimpoia, R., Liu, Z. and Guiot, S.R., 2013. Performance of a *Carboxydothermus hydrogenoformans* immobilizing membrane reactor for syngas upgrading into hydrogen. *International Journal of Hydrogen Energy*, *38*(5), pp.2167–2175.

Zhao, Y., Jia, X., Yang, J., Ling, Y., Zhang, Z., Yu, J., Wu, J. and Xiao, J., 2014. PanGP: a tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics*, *30*(9), pp.1297-1299.

Zheng, J., Ge, Q., Yan, Y., Zhang, X., Huang, L. and Yin, Y., 2023. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Research*, 51(W1), pp.W115–W121.

Zhou, J., Wu, K. and Rao, C.V., 2016. Evolutionary engineering of *Geobacillus thermoglucosidasius* for improved ethanol production. *Biotechnology and Bioengineering*, *113*(10), pp.2156-2167.

Zieliński, M., Park, J., Sleno, B. and Berghuis, A.M., 2021. Structural and functional insights into esterase-mediated macrolide resistance. *Nature Communications*, *12*(1), p.1732.

Zuo, P. and Rabie, A.B.M., 2010. One-step DNA fragment assembly and circularization for gene cloning. *Current Issues in Molecular Biology*, *12*(1), pp.11–16.

Züttel, A., 2004. Hydrogen storage methods. *Naturwissenschaften*, *91*(4), pp.157–172.

PLAGIARISM REPORT

This thesis was submitted to turnitin® software. The software evaluates the originality of the writing by comparing it to a comprehensive database including > 50 billion web pages, > 500 million student papers, and > 140 000 journals, periodicals and books. The cumulative similarity index of this thesis was 10%, with < 2% similarity per source.

