Exploiting acetogenic bacteria for sustainable production of chemicals

Margaux Poulalier Delavelle

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

August 2023

Declaration

Unless otherwise acknowledged, the work presented in this thesis is my own. No part has been submitted for any other degree in the University of Nottingham or other institute of learning.

Margaux Poulalier Delavelle, 2023

COVID-19 Impact Statement

The COVID-19 pandemic had a direct impact on this work because of the immediate closure of laboratory spaces for a total of 15 weeks between March and July 2020. While *in silico* work was carried out during this time, the disruption to *in vivo* work, central to this project, was substantial, with experiments having to be stopped at short notice and restarted in suboptimal conditions.

To accommodate social distancing upon return to the laboratory, bench spaces and offices were reallocated and shifts were put in place from July 2020 to late 2021, with alternating morning and afternoon patterns, with no out of hours work permitted. This resulted in considerable loss of available weekly working hours. At the time, autoclave runs, sequencing sample collection and consumable/primers collection were arranged only in the morning based on the regular work hours of the building staff and laboratory technicians. Therefore, when working on the afternoon shift, significant but unavoidable time was lost. These restricted working hours also constituted a hindrance because of the time-sensitive nature of microbiology work. The main bacteria studied in this work, *Acetobacterium woodii,* has a long doubling-time, exacerbating any delays encountered. Minor variations of growth or delays in experimental procedures resulted in the abandonment of experiments for lack of time.

Additionally, the systems in place for orders and deliveries, as well as the supply chains, were significantly disrupted. This led to severe delays in the shipping of reagents and consumables, resulting, at times, in consumables shortages. The turnaround of sequencing results was also severely impacted with sequencing results regularly taking a week instead of the expected 24 h.

To mitigate time loss, the thesis plan was regularly reviewed but additional time would have enabled the execution of repeats, the validation of results, and the completion of ongoing studies.

Publication

Poulalier-Delavelle M., Baker J.P., Millard J., Winzer K. and Minton N.P. (2023) Endogenous CRISPR/Cas systems for genome engineering in the acetogens *Acetobacterium woodii* and *Clostridium autoethanogenum*. *Front. Bioeng. Biotechnol.* 11:1213236. doi: 10.3389/fbioe.2023.1213236

Posters and presentations

Poulalier-Delavelle M., Baker J.P., Millard J., Winzer K. and Minton N.P.. Characterization of *Acetobacterium woodii* as an Industrial Chassis for Sustainable Chemical Production. Talk and poster presented at: Carbon Recycling Network Conference 3; September 11-13th, 2022; Nottingham, UK

Poulalier-Delavelle M., Baker J.P., Millard J., Winzer K. and Minton N.P.. Characterization of *Acetobacterium woodii* as an Industrial Chassis for Sustainable Chemical Production. Poster presented at: Clostridium XVI; September 14-17th, 2022; Toulouse, France

Poulalier-Delavelle M., Baker J.P., Millard J., Winzer K. and Minton N.P.. Use of the endogenous CRISPR/Cas system of the acetogen *Acetobacterium woodii* for genome engineering. Talk and poster presented at: 45th Symposium on Biomaterials, Fuels and Chemicals; April 30th-May 3rd, 2023; Portland, OR, USA

Abstract

Acetogens can play a role in achieving Net Zero through their ability to convert CO_2 into industrially relevant chemicals and fuels. Among acetogens, *Acetobacterium woodii* displays robust growth on CO_2 and adaptability to a variety of industrially relevant substrates but, as a homoacetogen, produces only acetate. Genomic engineering is required to alleviate potential metabolic bottlenecks and extension of its product range.

Reported here is the development and validation of an *in silico* tool to facilitate the harnessing of bacterial endogenous CRISPR/Cas systems through the automated elucidation of Protospacer Adjacent Motifs (PAM). It was exploited to characterise *A. woodii* endogenous Type I-B system which was subsequently employed to make gene deletions and insertions, followed by optimisation of parameters to improve editing efficiencies. The method was subsequently applied to *Clostridium autoethanogenum*.

To further assist in genome engineering of acetogens, genes essential to growth under both heterotrophic and autotrophic conditions were identified. This was achieved by high-throughput nucleotide sequencing of a pooled, *mariner*-based transposon library grown on fructose or H_2 + CO_2 in a continuous fermenter. Interestingly, the analysis additionally revealed genes which, when disrupted, led to increased fitness during autotrophic growth.

In *A. woodii*, genomic integration of the acetone pathway was shown to decrease acetone yields. To identify genomic loci that favoured high expression, a transposon was used to randomly deliver a fluorescent reporter to the genome and high expressing mutants identified by Fluorescence-Activated Cell Sorting (FACS). Future correlation of insertion loci with the fluorescence levels could identify candidate loci for exogenous gene integration.

Acknowledgements

First, I would like to thank my supervisors, Prof. Nigel Minton and Dr. Klaus Winzer for their guidance and support throughout my PhD. Thanks to Dr. Jonathan Baker for teaching me how to work with *A. woodii*, to Dr. James Millard for showing me all the tips and tricks to fermentation and to always being there to give me his opinion.

I am forever grateful to Dr. Christopher Humphreys and Michelle Kelly for always being there to answer my endless questions and give their opinions. I would also like to acknowledge the instrumental support of Dr. Craig Woods for implementing bioinformatics necessary for the transposon library.

I would like to thank everyone in the lab who made the long hours enjoyable. In particular, I would like to thank everyone from the office, Temi, Louise, Lauren, Danielle, and Natalia, for always being there and creating a safe space to laugh about the good and the bad.

Beyond the lab, I would like to thank my friends and family for their support. I would like to thank my partner, Liam, for his infinite support and keeping me sane through my studies. I am so grateful that we went through this journey together, I could not have done it without you. Par dessus tout, je remercie mes parents, pour leur compréhension, leurs encouragements, et leur soutien inconditionnel. Merci de l'intérêt que vous avez toujours porté à mes études et d'avoir toujours cru en moi, si j'en suis là aujourd'hui c'est grâce à vous.

I am thankful to funding from the University of Nottingham, Synthetic Biology Research Centre (SBRC) Doctoral Training Program and from the European Commission HORIZON 2020 Grant 760994 ENGICOIN.

List of contents

Declaration ..... **i**
COVID-19 Impact Statement ..... **ii**
Publication ..... **iii**
Abstract ..... **iv**
Acknowledgements ..... **v**
List of contents ..... **vi**
Abbreviations ..... **xiii**
List of Figures ..... **xvi**
List of Tables ..... **xxi**
CHAPTER 1: General Introduction **1**
1.1 Climate change and the bioeconomy 2
1.2 Acetogenic bacteria 5
1.2.1 Economical interest 5
1.2.2 Wood-Ljungdahl Pathway 5
1.2.3 Energy conservation 7
1.3 Acetobacterium woodii 8
1.3.1 WLP and energy conservation 9
1.3.2 Feedstock diversity 12
1.3.3 Culture conditions to optimise yields on $H_2 + CO_2$ 14
1.3.4 Plasmid-based strain engineering 15
1.4 Genome engineering 18
1.4.1 Allelic exchange 19
1.4.2 CRISPR/Cas9 21
1.5 Alternatives to CRISPR/Cas9 21
1.5.1 Cas9 variants 21
1.5.2 CRISPR/Cas12a 22
1.5.3 Endogenous CRISPR/Cas systems 23
1.6 Forward genetics 24
1.6.1 Transposon mutagenesis 25
1.6.2 Transposon libraries and Fluorescent-Activated Cell Sorting (FACS) 26
1.6.3 *mariner*-based transposon system in *Clostridium* 27
1.6.4 Fluorescent reporter genes for anaerobic conditions 27
1.7 Project aims 28

CHAPTER 2: Materials and methods 30
2.1 Python script development 31
2.2 Chemicals 31
2.3 Bacterial strains and plasmids 32
2.4 Growth media and conditions 32
2.5 DNA manipulations 33
2.6 Plasmid construction 34
2.6.1 Chapter 3 35
2.6.1.1 Validation of R-M recognition site impact on *A. woodii* transformation 35
2.6.1.2 Validation of *cas9* nuclease toxicity 35
2.6.1.3 Interference assays plasmids 36
2.6.1.4 Endogenous leader characterization 37
2.6.1.5 *pyrE* KO in *A. woodii* 37
2.6.1.6 *pheA* KO in *A. woodii* with 500 bp, 1.0 kb and 1.5 kb HA 38
2.6.1.7 *hsdR1* KO in *A. woodii* 39
2.6.1.8 *pyrE* KO in *C. autoethanogenum* 40
2.6.2 Chapter 5 41
2.6.2.1 Isopropanol production with *C. beijerinckii* SADH 41
2.6.2.2 Overexpression of *A. woodii* gene *gatD* 41
2.6.2.3 *gatD* KO in *A. woodii* 42
2.6.2.4 Study of creiLOV, Bs2 and Pp1 fluorescent proteins 42
2.6.2.5 FAST gene KI at the *pheA* locus in *A. woodii* 43
2.6.2.6 Terminator characterisation 45
2.6.2.7 Transposon library plasmid with FAST reporter gene 46
2.6.2.8 *C. beijerinckii* SADH gene KI at the *purD* locus in *A. woodii* 46
2.6.2.9 Acetone operon KI at the *pheA* locus in *A.woodii* 47
2.7 Strain engineering 48
2.7.1 *Escherichia coli* 48
2.7.1.1 Chemically competent cells 48
2.7.1.2 Transformation and colony PCR 48
2.7.2 *Acetobacterium woodii* 49
2.7.2.1 Competent cells: standard and high density 49
2.7.2.2 Transformation and colony PCR 49
2.7.2.3 Plasmid loss 49
2.7.3 *Clostridium autoethanogenum* 50
2.7.3.1 Conjugation and colony PCR 50

2.7.3.2 Plasmid loss 51
2.8 Transcript levels 51
2.8.1 RNA extraction 51
2.8.2 Quantitative reverse transcription real-time PCR (RT-qPCR) 52
2.9 Fluorescence measurements 52
2.9.1 Plate reader for creiLOV, Pp1, and Bs2 in *Escherichia coli* 52
2.9.2 Plate reader for Pp1 protein emission and excitation scans in A. *woodii* 53
2.9.3 Flow cytometry for FAST protein 53
2.10 Inverse PCR 54
2.11 Heterotrophic batch fermentation in serum bottles 55
2.12 TraDIS library 55
2.12.1 Continuous fermentation in stirred tank bioreactor 55
2.12.2 Library preparation for sequencing 56
2.12.3 Sequencing 57
2.12.4 Bioinformatics 58
CHAPTER 3: CRISPR/Cas systems for genome engineering of *A. woodii* 59
3.1 Introduction 60
3.1.1 Genetic engineering of *Acetobacterium woodii* 60
3.1.1.1 Plasmid-based engineering 60
3.1.1.2 Allelic Exchange 61
3.1.2 Barriers to the utilization of *Streptococcus pyogenes* Type II CRISPR/Cas9 62
3.1.2.1 Toxicity of the Cas9 nuclease 63
3.1.2.1.1 DNA Nickase 64
3.1.2.1.2 Inactive Cas9 65
3.1.2.1.3 Regulation of *cas9* expression 66
3.1.2.2 Restriction-Modification systems 67
3.1.3 Endogenous CRISPR/Cas systems 69
3.1.3.1 Distribution and description of endogenous CRISPR/Cas systems 69
3.1.3.2 Harnessing of endogenous CRISPR/Cas systems for genome engineering 72
3.1.3.2.1 Functional PAM identification 73
3.1.3.2.2 *In silico* strategies 73
3.1.3.2.3 *In vivo* strategies 74
3.1.3.2.4 Need for improved PAM prediction 74
3.1.3.3 Editing vector construction 75
3.1.4 Aims 77
3.2 Results 78

3.2.1 *S. pyogenes* CRISPR/Cas9 system in *A. woodii* 78
3.2.1.1 Restriction-Modification (R-M) system in *A. woodii* 78
3.2.1.1.1 *In silico* identification of R-M system specific to *cas9*-containing plasmid transformation 78
3.2.1.1.2 *In vivo* validation of the R-M system identification 81
3.2.1.2 Cas9 nuclease toxicity in *A. woodii* 82
3.2.2 Optimization of *in silico* steps for harnessing of endogenous CRISPR systems 84
3.2.2.1 Python script development 84
3.2.2.2 Python script performance evaluation against published literature 86
3.2.2.3 Analyses of other *Clostridium* spp. 87
3.2.2.4 Analyses of bacteria harbouring CRISPR Type I systems 89
3.2.3 Harnessing *A. woodii* Type I-B CRISPR/Cas system 91
3.2.3.1 *In silico* analysis of *A. woodii* Type I-B CRISPR system 91
3.2.3.2 PAM validation by interference assay 92
3.2.3.3 Characterisation of the leader sequence 95
3.2.3.4 Genome engineering capacity validation by *pyrE* knock-out generation 97
3.2.3.5 Impact of homology arms lengths on editing efficiency for *pheA* knockout 102
3.2.3.6 Validation of the system with 3.2 kb *hsdR1* knock-out and impact of cell densities and DNA concentration 105
3.2.3.6.1 Knock-out generation 105
3.2.3.6.2 Phenotypic analysis of the knock-out strains 107
3.2.4 Harnessing *C. autoethanogenum* Type I-B system 109
3.2.4.1 Description of CRISPR locus in *C. autoethanogenum* and PAM prediction with Python script 109
3.2.4.2 Genome engineering capacity validation by *pyrE* knockout in *C. autoethanogenum* 110
3.2.4.2.1 Generation of knock-out strain 110
3.2.4.2.2 Phenotype validation 113
3.3 Discussion 114
3.3.1 *S. pyogenes* CRISPR/Cas9 system in *A. woodii* 114
3.3.2 Harnessing endogenous CRISPR/Cas systems 117
3.3.2.1 Python script development and validation for PAM prediction 117
3.3.2.2 Use of endogenous CRISPR/Cas system for genome engineering in acetogens 121
3.3.2.2.1 Optimization of the workflow in *A. woodii* and *pyrE* KO in *C. autoethanogenum* 121

3.3.2.2.2 *A. woodii* Δ*hsdR1* strain phenotype 124
3.4 Key outcomes 125
CHAPTER 4: Elucidation of essential and detrimental genes during heterotrophic and autotrophic fermentations 127
4.1 Introduction 128
4.1.1 Transposon mutagenesis in *Clostridium* 128
4.1.2 Gene overexpression for improved fitness in *A. woodii* 132
4.1.3 Mutant abundance studies 134
4.1.4 Transposon mutagenesis in *A. woodii* (Dr. Jonathan Baker) 136
4.1.5 Fermentation 138
4.1.6 Aims 139
4.2 Results and discussion 140
4.2.1 Continuous fermentation on fructose or H_2/CO_2 140
4.2.2 Read mapping and essentiality studies 142
4.2.2.1 Reads mapping and coverage 142
4.2.2.2 Selective pressure and mutant abundance 145
4.2.2.3 Residual library plasmid 147
4.2.2.4 Essentiality studies 149
4.2.3 Role of *A. woodii* genes under different conditions 152
4.2.3.1 Glycolysis / gluconeogenesis, pyruvate metabolism, and tricarboxylic acid (TCA) cycle 152
4.2.3.2 Wood-Ljungdahl Pathway, energy conservation and electron bifurcation 160
4.2.3.2.1 Wood-Ljungdahl Pathway 163
4.2.3.2.2 Energy conservation 169
4.2.3.3 Glycine cleavage system, acid stress and Na^+ transport 179
4.2.3.3.1 Glycine cleavage system (GCS) 179
4.2.3.3.2 Acid stress 181
4.2.3.3.3 Na^+ transport 183
4.2.3.4 Alternative substrates 183
4.2.3.4.1 1,2-propanediol pathway 186
4.2.3.4.2 Alanine pathway: detrimental to H_2/CO_2 fermentation 188
4.2.3.4.3 Glycine betaine pathway: detrimental to fructose fermentation 191
4.2.4 Additional detrimental genes 193
4.2.4.1 Detrimental genes for heterotrophic and autotrophic fermenter growth 193
4.2.4.1.1 Awo_c35490 193
4.2.4.1.2 Awo_c26530 / Awo_c26540, bcaT1 and Awo_26510 195

4.2.4.2 Detrimental genes for heterotrophic fermenter growth only 197
4.2.4.2.1 Methyltransferase mttC7 is extremely detrimental 197
4.2.4.2.2 Miscellaneous genes 199
4.2.4.3 Detrimental genes for autotrophic fermenter growth only 201
4.2.4.3.1 Awo_c03590 is extremely detrimental 201
4.2.4.3.2 Awo_c14190 and Awo_c14200 204
4.2.4.3.3 Neu 205
4.2.4.3.4 Miscellaneous genes 207
4.3 General discussion 209
4.3.1 Residual plasmid presence and mutant emergence 209
4.3.2 Combination of essentiality and abundance study 209
4.3.3 Limitations 212
4.4 Key outcomes 214
CHAPTER 5: Towards improvements of yields in *A. woodii* 215
5.1 Introduction 216
5.1.1 Acetone production in *A. woodii* 216
5.1.1.1 Limitations of plasmid-based expression and genomic integration 218
5.1.1.2 Conversion of acetone into isopropanol 219
5.1.2 Random knock-in for identification genome loci favourable for expression of integrated genes 220
5.1.2.1 Flow cytometry in anaerobic conditions 222
5.1.2.2 Fluorescence reporter gene 222
5.1.2.2.1 LOV-domain containing genes 223
5.1.3 Aims 225
5.2 Results 227
5.2.1 Isopropanol production with *C. beijerinckii* SADH 227
5.2.2 Redirecting the carbon flux for improvement of acetone yields 231
5.2.2.1 *In silico* identification of enzymes responsible for acetone to isopropanol conversion in *A. woodii* 231
5.2.2.2 Study of function of the candidate 233
5.2.2.2.1 Gene overexpression 233
5.2.2.2.2 Gene knock-out 237
5.2.3 Transposon library for discovery of novel genome loci for gene expression 239
5.2.3.1 Fluorescent reporter genes 239
5.2.3.1.1 Reporter genes with LOV domains 239
5.2.3.1.2 FAST 240

5.2.3.2 Library construction 245
5.2.3.2.1 Terminator characterization 245
5.2.3.2.2 Plasmid design and assembly 247
5.2.3.2.3 Plasmid retention 248
5.2.3.3 Library analysis 250
5.2.3.3.1 Fluorescence 250
5.2.3.3.2 Verification of single copy of the FAST gene in the sorted cells 253
5.2.3.4 Control strains 257
5.2.3.4.1 Isopropanol production 258
5.2.3.4.2 Acetone production 258
5.3 Discussion 260
5.3.1 Use of the endogenous CRISPR/Cas system for KO and KI 260
5.3.2 Heterotrophic batch fermentation for acetone production 260
5.3.3 Fluorescence reporter genes 263
5.3.4 Library analysis 265
5.4 Key outcomes 268
CHAPTER 6: General Discussion 269
6.1 Context 270
6.2 Summary of findings 270
6.3 Further studies 274
References 279
Appendices 300

Abbreviations

5-FC	5-fluorocytosine
5-FOA	5-fluoroorotic acid
ACE	Allele-Coupled Exchange
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
ATPase	ATP synthase
BLAST	Basic local alignment search tool
BMC	Bacterial microcompartments
Bp	Base pairs
C1	One carbon compound
Cas	CRISPR-associated
CFU	Colony forming unit
Cm	Chloramphenicol
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
DEPC-H_2O	Diethyl pyrocarbonate treated water
dH_2O	Distilled water
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DR	Direct repeat
DSB	Double strand breaks
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
EMP	Embden Meyerhof Parnas
Erm	Erythromycin
FAST	Fluorescence-activating absorption-shifting tag
FbFP	FMN-binding fluorescent protein
Fd	Oxidized ferredoxin

Fd^{2-}	Reduced ferredoxin
FMN	Flavin mononucleotide
gDNA	Genomic DNA
GFP	Green fluorescent protein
GHG	Greenhouse Gas
HA	Homology arm
HMBR	4-hydroxy-3-methylbenzylidene-rhodanine
HPLC	High Performance Liquid Chromatography
IEP	Intron-encoded protein
iGEM	International Genetically Engineered Machine
IMP	Inosine monophosphate
InDel	Insertions and deletions
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
ITR	Inverted terminal repeats
kb	Kilobase pairs
KI	Knock-In
KO	Knock-Out
LB	Luria Bertani
LHA	Left homology arm
LOV	Light, oxygen, or voltage sensing
MCS	Multiple cloning site
MGE	Mobile genetic element
MTase	Methyltransferase
NAD^+	Oxidised nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide
$NADP^+$	Oxidised nicotinamide adenine dinucleotide phosphate
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NEB	New England Biolabs
NGS	Next generation sequencing
NHEJ	Non-Homologous End Joining

OD_{600nm}	Optical density at 600nm
ORF	Open reading Frame
PAM	Protospacer Adjacent Motif
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
P*i*	Inorganic phosphate
PP*i*	Inorganic pyrophosphate
RBS	Ribosome binding site
REase	Restriction endonuclease
RFP	Ribosomal protected fragment
RHA	Right homology arm
R-M	Restriction-Modification
RNA	Ribonucleic acid
rpm	Rotations per minute
RT-qPCR	Quantitative reverse transcription real-time PCR
SADH	Secondary alcohol dehydrogenase
sgRNA	single guide RNA
SMRT	Single Molecule Real-Time
SOEing PCR	Splicing by Overlap-Extension PCR
SSN	Single strand nick
Syngas	Synthesis gas
TIS	Transposon insertion sequencing
Tm	Thiamphenicol
TraDIS	Transposon-directed insertion site sequencing
UIS	Unique insertion site
UMP	Uridine monophosphate
WLP	Wood-Ljungdahl Pathway
WT	Wild type

List of Figures

Figure 1-1: Bioenergetics of acetogenesis from $H_2 + CO_2$ in *A. woodii.* 10

Figure 3-1: Schematic representation of the native Cas9 bound to target genomic DNA and the activity of its different variants; 63

Figure 3-2: Steps of native CRIPSR interference for Type I systems. 71

Figure 3-3: Steps to harness the endogenous CRISPR system. The first step is to identify potential PAM sequences..... 72

Figure 3-4: Analysis of *A. woodii* R-M systems and their impact on plasmid transformation..... 79

Figure 3-5: Transformation efficiency of WT *A. woodii* obtained with the vector containing the R-M target site (pMTL8215_targetIIP) expressed as a percentage of the efficiency obtained with the empty vector (pMTL82151) for each corresponding replicate. 81

Figure 3-6: Assessment of the impact of the Cas9 toxicity and the presence of R-M target site on plasmid transformation in *A. woodii*. 83

Figure 3-7: General overview from initial user input to final user-friendly results of the developed Python script. 85

Figure 3-8: Organisation of the Type I-B CRISPR system of *Acetobacterium woodii* DSM1030. 91

Figure 3-9: Interference assay to determine functional PAMs in *A. woodii*. (A) Table of the protospacers used in the interference assay in *A. woodii*. 93

Figure 3-10: Endogenous CRISPR array leader sequence characterization: FAST transcript level normalized with *orfB* as the reference gene measured by RT-qPCR at different stages of cell growth. 95

Figure 3-11: Endogenous CRISPR based deletion of the *A. woodii pyrE* gene. 99

Figure 3-12: Endogenous CRISPR based deletion of 354 bp in the *A. woodii pheA* gene. 103

Figure 3-13: Endogenous CRISPR based deletion of the *A. woodii hsdR1* gene 105

Figure 3-14: *A. woodii* WT and Δ*hsdR1* transformation efficiencies obtained with plasmids containing one to three R-M sites into normalized to the transformation efficiency obtained with the empty backbone pMTL82151..... 107

Figure 3-15: Endogenous CRISPR based deletion of the *C. autoethanogenum pyrE* gene..... 111

Figure 3-16: Phenotypic analysis of the *C. autoethanogenum* Δ*pyrE* strains obtained. 112

Figure 4-1: Schematic representation of the library construction in *A. woodii* with pMTL-GL15 plasmid. 136

Figure 4-2: Growth and metabolites profiles during (A) heterotrophic and (B) autotrophic (H_2/CO_2) continuous fermentations. 139

Figure 4-3: Analysis of reads mapping and discovery of unique insertion sites (UIS) for the library..... 142

Figure 4-4: Log_2 of the fold change of abundance of reads for each *A. woodii* gene between the first and second subculture in fructose and the first and last sample of both the fructose and H_2/CO_2 fermenters. 144

Figure 4-5: Percentage of plasmid-containing mutants during (A) heterotrophic batch fermentations, and (B) heterotrophic (●) or autotrophic (■) continuous fermentations. 146

Figure 4-6: Distribution of insertion index for (A) the second fructose subculture, and the final time point of both (B) heterotrophic and (C) autotrophic fermenters. 149

Figure 4-7: Venn diagram of essential genes for the fructose subcultures, fructose fermenter, and gas fermenter, (A) excluding or (B) including ambiguous genes. 150

Figure 4-8: Fructose metabolism in *A. woodii* and the genes involved..... 152

Figure 4-9: Gene-level abundance and essentiality studies of the genes involved in glycolysis/gluconeogenesis and pyruvate metabolism. 153

Figure 4-10: Insertion plots in the enolase gene for the second fructose subculture and final time point of both fermenters. 154

Figure 4-11: Classical tricarboxylic acid (TCA) cycle as identified in *A. woodii* and with the metabolic reactions linking it to glycolysis/gluconeogenesis and the WLP. 157

Figure 4-12: Bioenergetics of acetogenesis from $H_2 + CO_2$ in *A. woodii.* 160

Figure 4-13: Gene-level abundance and essentiality studies of the genes involved in the WLP, energy conservation, and electron bifurcation. 161

Figure 4-14: Insertion plots in genes of the HDCR for the second fructose subculture and final time point of both fermenters. 162

Figure 4-15: Single mutants abundance in gene *fhs1* in both fermenters. 164

Figure 4-16: Single mutants abundance in genes Awo_c10690 and Awo_c10700 in both fermenters. 166

Figure 4-17: Insertion plots in genes of the Rnf cluster for the second fructose subculture and final time point of both fermenters. 169

Figure 4-18: Insertion plots in genes of the F-type ATPase for the second fructose subculture and final time point of both fermenters. 171

Figure 4-19: *A. woodii* V-ATPase under the different conditions. 173

Figure 4-20: Insertion plots in genes of the electron-bifurcating hydrogenase for the second fructose subculture and final time point of both fermenters. 175

Figure 4-21: Gene-level abundance and essentiality studies of the genes involved in pyruvate metabolism, glycine cleavage system, acid stress and Na^+ transport. 178

Figure 4-22: Arginine deaminase pathway in *A. woodii*. 181

Figure 4-23: Gene-level abundance and essentiality studies of the genes involved in alternative substrates utilisation. 184

Figure 4-24: Analysis of the alanine pathway in *A. woodii* during H_2/CO_2 fermentation. 188

Figure 4-25: Alanine concentrations during both the (A) heterotrophic and (B) autotrophic (H_2/CO_2) continuous fermentations. 190

Figure 4-26: Gene organisation of the glycine betaine uptake and utilisation in *A. woodii*, comprised between position 860,864 and 868,984. 191

Figure 4-27: Gene Awo_c35480 and Awo_c35490 mutant abundance during fructose and H_2/CO_2 fermentations. 193

Figure 4-28: Log_2 fold change of genes detrimental to both fermenters. 195

Figure 4-29: Methyltransferase mttC7 and GntR-family regulator mutant abundance in fructose fermenter. 197

Figure 4-30: Miscellaneous genes mutant abundance in fructose fermenter. 199

Figure 4-31: Analysis of Awo_c03590 insertion mutants during fermentation. 201

Figure 4-32: Analysis of Awo_c14190 and Awo_c14200 during H_2/CO_2 fermentation. 204

Figure 4-33: Genes Awo_c24820 to Awo_c24850 mutant abundance in fructose and H_2/CO_2 fermenters. 205

Figure 4-34: Miscellaneous genes mutant abundance in H_2/CO_2 fermenter. 207

Figure 5-1: Acetone production from acetyl-CoA in the acetogen *A. woodii*. 216

Figure 5-2: Product profiles of *A. woodii* strains with the acetone operon and/or the SADH gene from *C. beijerinckii*. 227

Figure 5-3: Growth and product profiles of *A. woodii* strains with the acetone operon and/or the SADH gene from *C. beijerinckii*. 228

Figure 5-4: Protein similarity between *A. woodii* GatD and the SADH from *C. autoethanogenum* (Cau) and *C. beijerinckii* (Cbe). 231

Figure 5-5: InterPro analysis of *A. woodii* GatD protein..... 233

Figure 5-6: Growth and product profiles of *A. woodii* strains with the acetone operon and overexpressed SADH from *C. beijerinckii* or *gatD* from *A. woodii*. 234

Figure 5-7: Product profiles of *A. woodii* with the acetone operon and overexpressed SADH from *C. beijerinckii* or *gatD* from *A. woodii*. 235

Figure 5-8: Endogenous CRISPR based deletion of the *A. woodii gatD* gene..... 237

Figure 5-9: Endogenous CRISPR based insertion of P_{thl}_FAST at the *pheA* locus in the *A. woodii* genome. 241

Figure 5-10: Flow cytometry analyses of *A. woodii* using the FAST reporter gene (P_{thl}_FAST) expressed on a plasmid or in the genome. 242

Figure 5-11: Terminator characterization for use in the FAST transposon library. 245

Figure 5-12: Transposon library construction for random insertion of both an antibiotic resistance gene and the FAST fluorescent reporter using plasmid pMTL-GL15-FAST. 246

Figure 5-13: Resistance profile of colonies obtained after transformation of *A. woodii* WT and *tcdR* with plasmid pMTL-GL15-FAST for FAST transposon library creation. 247

Figure 5-14: Flow cytometry analysis of the FAST transposon library in *A. woodii*. 250

Figure 5-15: Analysis of genomic DNA samples from single colonies obtained by FACS. 253

Figure 5-16: Inverse PCR protocol performed on plasmid pMTL-GL15-FAST and sample G9 of the library of Chapter 4 (mix of mutants). 254

Figure 5-17: Knock-in of the acetone operon at the pheA locus of *A. woodii* using its endogenous CRISPR/Cas system. 257

List of Tables

Table 2-1: Number of single cells analysed for the different samples represented in Figure 5.14..... 54

Table 3-1: Comparison of the 5′ PAM consensus sequences (5′ → 3′) found in the literature with the ones from the Python script for different strains of *Clostridium spp.*. 87

Table 3-2: Details of the analyses of the arrays corresponding to CRISPR/Cas Type I-B systems of a range of *Clostridium spp*. 88

Table 3-3: Details of the analyses of the arrays corresponding to CRISPR/Cas Type I systems of a range of bacteria. 90

CHAPTER 1: General

Introduction

1.1 Climate change and the bioeconomy

Historically, chemicals have been produced through biological processes. During First World War, for instance, the Acetone-Butanol-Ethanol (ABE) fermentation was the predominant means of producing these three chemicals. However, in the 20th century, innovation centred around the petrochemical industry, as a consequence, biological processes such as the ABE fermentation soon became uneconomical resulting in their almost complete abandonment (Karimi et al., 2015; Moon et al., 2016).

The finite nature of fossil resources has long been a reason for developing sustainable routes to the production of chemicals and fuels. Additionally, the evidence that the utilisation of fossil resources and the release of CO_2 into the atmosphere is driving global warming and climate change is now irrefutable. The urgent need to reduce GHG emissions has now put sustainability and green chemicals at the forefront of scientific research and commercial innovation.

In 2015, 196 countries adopted the Paris agreement, committing to limit global warming to 1.5ºC above pre-industrial levels. To reach this objective, the United Nations estimate that emissions need to be reduced by 45% by 2030 and to as close to zero as possible by 2050 (UNFCC, 2015). Biotechnology is at the centre of the long-term innovation required to achieve those objectives. In 2017, the EU bioeconomy employed 17.5 million people and represented 4.7% of the Gross Domestic Product (GDP) (Ronzon et al., 2020).

The first generation of biofuels relies on easily degradable carbohydrates obtained from corn starch or sugar cane. While this technology reached commercial scale, it requires purpose-grown crops. As such it competes for land use with food crops, threatening food security and biodiversity. According to Oxfam international, these technologies

are not relevant for the current crisis as they are “deepening poverty and accelerating climate change” (Bailey, 2008).

Consequently, the second generation of biofuels focused on the utilisation of lignocellulosic biomass which can be sourced from agricultural and forestry waste which represents a more circular alternative. However, lignocellulosic substrates require expensive and energy intensive pre-treatments and hydrolysis steps to break down lignocellulosic biomass into fermentable substrates (Geddes et al., 2011). To address this requirement, consolidated bioprocessing (CBP) aims to convert lignocellulose into products of interest in one step. CBP relies on heterologous pathway expression to develop microbial strains capable of biomass solubilisation and production of high yields of desired products (Olson et al., 2012).

Alternatively, biomass can also be valorised by gasification into syngas, a mixture of carbon dioxide, carbon monoxide, and hydrogen. A wide range of biomass can be used, such as forestry residues, solid municipal waste, and agricultural waste, with an impact on the final gas mixture composition (Liew et al., 2016). Gasification is energy intensive as it requires high temperatures and high pressure, the high temperature produces hot steams that can be valorised to produce electricity (Handler et al., 2016). Similar gas feedstocks can also be obtained directly from waste gases of industrial processes like steel mills and refineries (Köpke & Simpson, 2020; Takors et al., 2018). Waste streams containing solely carbon dioxide such as waste from large-scale fermentations require combination with green hydrogen produced from water electrolysis. Valorisation of CO_2-only waste streams thus rely on cheap and sustainable electricity availability (Takors et al., 2018).

The Fischer-Tropsch process has historically been used to convert syngas into hydrocarbons. It requires high temperatures and pressures as well as a catalyst such as cobalt, ruthenium, and iron (Liew et al., 2016). On the other hand, gas mixtures can also be utilised by bacteria to produce chemicals of interest by gas fermentation. Unlike the Fischer-Tropsch process, biological gas fermentation has been shown to be flexible to a range of gas compositions and be tolerant to contaminants in the syngas (Köpke & Simpson, 2020; Liew et al., 2016; Takors et al., 2018). The process has already been commercialized by LanzaTech who currently operate continuous fermenters co-located with steel mills in China to produce ethanol. Life cycle analyses revealed that despite the initial energy input, the ethanol produced results in at least 60% reduction of GHG emissions compared to conventional ethanol, the process would be even more beneficial from biomass with a 90% reduction in GHG emissions (Handler et al., 2016).

However, drawbacks still hinder gas fermentation developments such as the low water solubility of gaseous substrates (Takors et al., 2018), as well as transport and storage challenges caused by their toxicity and reactivity.

To overcome these drawbacks, formate and methanol have been proposed as alternative feedstocks (Cotton et al., 2020; Yishai et al., 2016). They are both fully miscible in water and are easy to transport and store. They can both be produced by hydrogenation of CO_2 with green hydrogen (Szima & Cormos, 2018; Álvarez et al., 2017), and formate can also be produced directly from electrochemical reduction. As such, formate and methanol can act as electricity storage solutions. One of the limitations of formate and methanol as growth substrate is their toxicity at high concentrations (Cotton et al., 2020).

1.2 Acetogenic bacteria

1.2.1 Economical interest

Acetogenic bacteria represent ideal candidates for gas fermentation as they are able to fix CO_2 into acetyl-CoA via the Wood-Ljungdahl pathway (WLP), which was found to be the most efficient non-photosynthetic carbon fixation pathway (Fast & Papoutsakis, 2012). The WLP is also involved in their growth on hexoses but, more importantly, allows them to autotrophically grow on other C1 compounds like methanol and formate. As discussed above, this flexibility might be crucial for evolving feedstocks availability and prices.

While acetyl-CoA can be converted into acetate through the action of a phosphotransacetylase and an acetate kinase to generate ATP, it can also serve as a precursor for biomass formation by conversion of acetyl-CoA into pyruvate. Pyruvate or phosphoenolpyruvate can then be converted into oxaloacetate (Imkamp & Müller, 2007). However, because of the resulting ATP generation, acetate is the characteristic main product of several acetogens.

Although they all rely on the WLP, acetogens are a diverse group of bacteria with approximately 100 species in 21 genera (Imkamp & Müller, 2007). This diversity is a great opportunity for industrial processes as it results in a diversity of native products, such as, ethanol, butyrate, butanol, and 2,3-butanediol.

1.2.2 Wood-Ljungdahl Pathway

Bacterial formation of acetate from $H_2 + CO_2$ was first described in 1939 in *Clostridium aceticum* (Wieringa, 1939) but the strain was lost in the 1940s (Poehlein et al., 2015). Consequently, the Wood-Ljungdahl Pathway was elucidated through the study of *Moorella thermoacetica* (formerly *Clostridium thermoaceticum*) by H. G. Wood and

Lars G. Ljungdahl, hence the name of the pathway (Ljungdahl, 1986; Wood, 1991; Wood et al., 1986).

The Wood-Ljungdahl pathway has since been studied and reviewed extensively (Bengelsdorf et al., 2016; Bertsch & Muller, 2015a; Drake et al., 2008; Ragsdale & Pierce, 2008). It is constituted of two branches coined the 'methyl branch' and the 'carbonyl branch'.

The methyl branch reduces CO_2 into a methylated corrinoid iron-sulfur protein (CoFeSP). The first step is the reduction of CO_2 into formate by a formate dehydrogenase. Formate is then activated and condensed with tetrahydrofolate (THF) to produce formyl-THF by a formyl-THF synthase; this step is endergonic and requires one ATP. Formyl-THF is then reduced to methyl-THF through the successive action of three other THF-dependent enzymes: a formyl-THF cyclohydrase, a methylene-THF dehydrogenase, and a methylene-THF reductase. As the last step of the methyl branch, the methyl group is transferred to a corrinoid iron-sulfur containing protein (CoFeSP).

The carbonyl branch catalyses the interconversion between CO_2 and CO using a carbon monoxide dehydrogenase. During autotrophic growth on CO, it generates the required CO_2 for the methyl branch, while on $H_2 + CO_2$, it catalyses CO formation from CO_2. The bi-functional carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS) subsequently transfers the methyl group from the methylated CoFeSP to CO and CoA to form acetyl-CoA.

Acetyl-CoA is a crucial node in acetogenic metabolism. It can be used for ATP generation by acetate production through the action of a phosphotransacetylase and acetate kinase, recouping the ATP invested for formate to formyl-THF conversion. Alternatively, it can be used for biomass generation by conversion to pyruvate.

1.2.3 Energy conservation

As described above, the WLP is ATP neutral with one ATP required for formate utilisation for one ATP generated by substrate level phosphorylation during acetate production. As such, much of the initial impetus of research on acetogens focussed on elucidating energy conservation mechanisms.

Heise and co-workers first described in 1993 the fact that, in *A. woodii,* the production of acetyl-CoA through the WLP creates a sodium gradient across its membrane, this gradient then serves as a driving force for production of ATP with a membrane-bound ATP synthase (Heise et al., 1993).

Discovery of the membrane-bound Rnf (*Rhodobacter* nitrogen fixation) complex in *A. woodii* helped better understand energy conservation mechanisms (Müller et al., 2008). The Rnf complex is a ferredoxin:NAD^+-oxidoreductase, it catalyses oxidation of reduced ferredoxin (Fd^{2-}) to the reduction of NAD^+ accompanied by sodium ion translocation effectively creating a transmembrane electrochemical sodium ion gradient. This gradient is then employed by an F_1F_o-ATPase to drive ATP synthesis from sodium ion translocation (Müller et al., 2001). The sodium ion gradient is not conserved in all acetogens as others use a proton gradient generated by a ferredoxin:H+-oxidoreductase, it was discovered in *Thermoanaerobacter kivui* and coined Ech (energy converting hydrogenase) (Schoelmerich & Müller, 2019). The Rnf and Ech complexes were found to be mutually exclusive (Kremp et al., 2020).

The NADH and Fd^{2-} required for the WLP are generated by oxidation of molecular hydrogen by a cytoplasmic electron-bifurcating hydrogenases (Schuchmann & Müller, 2012). Such systems are exemplified by the [FeFe]-hydrogenases from *A. woodii* and *C. autoethanogenum* (Schuchmann & Müller, 2012; Wang et al., 2013). While the

hydrogenase from *A. woodii* was found to catalyse hydrogen-based NAD^+ and ferredoxin reduction (Schuchmann & Müller, 2012), the hydrogenase from *C. autoethanogenum* catalysed $NADP^+$ and ferredoxin reduction and is involved in CO_2 reduction to formate (Mock et al., 2015; Wang et al., 2013).

In *C. autoethanogenum*, the NAD(H) and NADP(H) pools are balanced by a second electron bifurcation system, first discovered in *Clostridium kluyveri* (Wang et al., 2010). An electron bifurcating NADH-dependent reduced ferredoxin:$NADP^+$ oxidoreductase (Nfn) catalyses electron transfer from NADPH to NAD^+ and ferredoxin (Wang et al., 2013). In *C. autoethanogenum*, the Nfn was shown to be essential during growth on syngas and H_2 + CO_2 (Marcellin et al., 2016). Similarly, a cytochrome containing hydrogenase was recently found in *Sporomusa ovata* in addition to its Rnf complex (Kremp et al., 2020). *Sporomusa* type Nfn (Stn) electron-bifurcating transhydrogenase were further identified and described in other acetogens (Kremp et al., 2020) and the authors later hypothesized that it could constitute a "third way of chemiosmotic energy conservation", "but only in addition to the well-established respiratory enzymes Rnf and Ech" (Kremp et al., 2022).

1.3 Acetobacterium woodii

A. woodii was discovered in 1977 in black sediment of an estuary (Balch et al., 1977) and its full genome was published in 2012 (Poehlein et al., 2012). *A. woodii* is strictly anaerobic, rod-shaped, non-spore-forming, and motile. Interestingly, it is a homoacetogen, it naturally produces exclusively acetate in either autotrophic or heterotrophic conditions. *A. woodii* has been studied extensively in the past decades as the model acetogen and key features of its metabolism are discussed below as well as the efforts to date to exploit it as an industrial chassis.

1.3.1 WLP and energy conservation

The carbon fixation of the Wood-Ljungdahl Pathway is conserved across acetogens. However, some enzymes of the pathway and the energy conservation mechanisms differ. The specificities of the WLP and energy conservation mechanisms of *A. woodii* are presented below and acetogenesis during $H_2 + CO_2$ growth represented in Figure 1.1.

The first step is CO_2 reduction to formate, *A. woodii* possesses a hydrogen-dependent CO_2 reductase (HDCR) with two distinct modules, a hydrogenase and a formate dehydrogenase linked by two electron-transferring subunits (Ceccaldi et al., 2017). Compared to other acetogens, the inherent specificity of this system is its ability to accept electrons directly from H_2 (Ceccaldi et al., 2017).

Both the methylene-THF dehydrogenase (Ragsdale & Ljungdahl, 1984) and the methylene-THF reductase (MTHFR) (Bertsch et al., 2015) were found to be solely dependent on NADH. The MTHFR was found to have a peculiar organisation with a MetF and a MetV but also a RnfC2 subunit. The latter was found to be the result of a fusion between the RnfC subunit from the Rnf complex and MetV (Bertsch et al., 2015). In some acetogens, MetF alone can reduce methylene-THF with NAD(P)H but the reaction is anabolic. In *A. woodii* RnfC2, and not MetF, binds and oxidizes NADH. The RnfC2 subunit was hypothesized to allow the MTHFR to interact with the Rnf complex or, more likely, to enable methyl-THF oxidation during growth on methylated substrates (Bertsch et al., 2015). This oxidation is very endergonic and prevents growth on methylated substrates for a number of acetogens, unlike *A. woodii*.

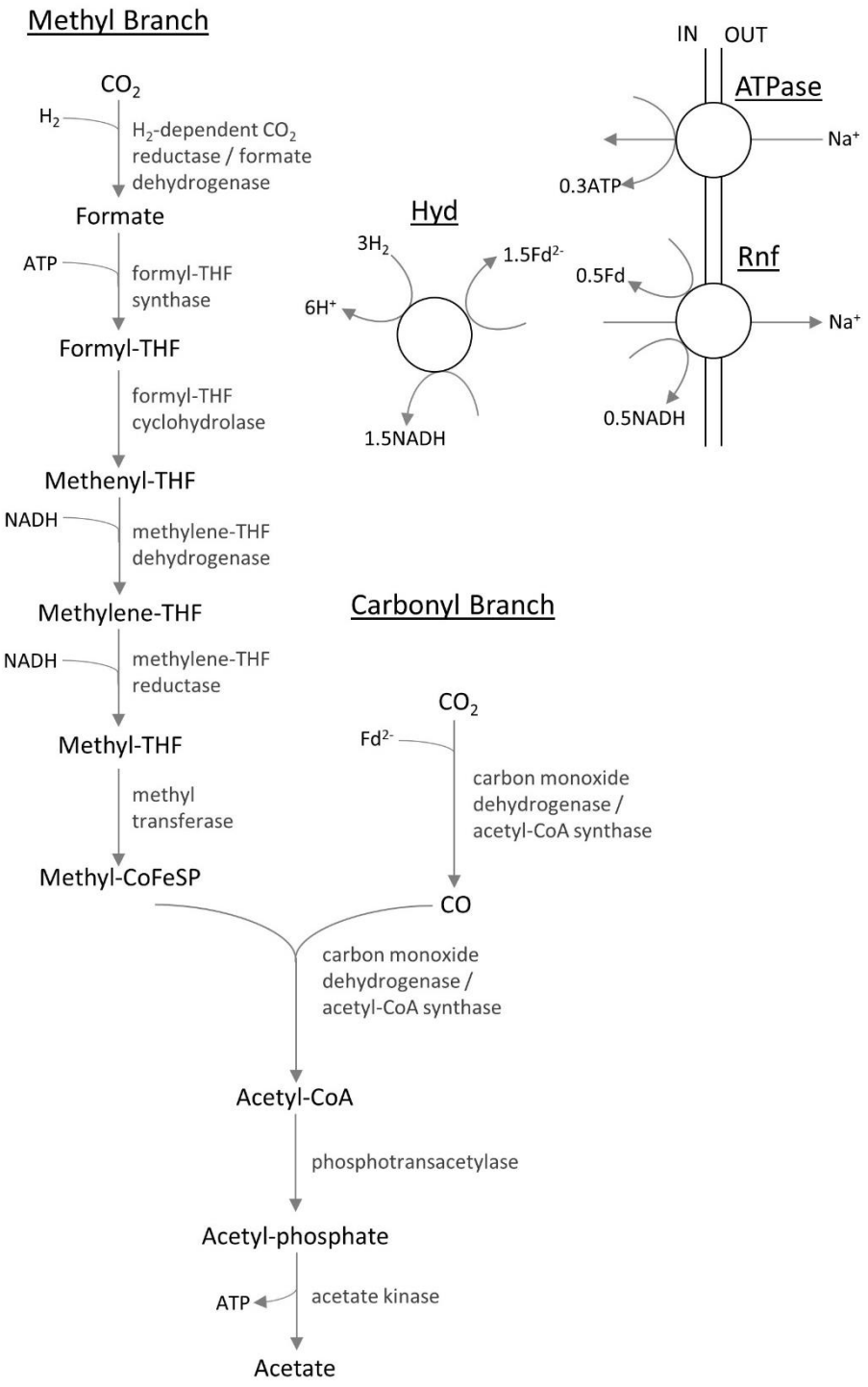


Figure 1-1: Bioenergetics of acetogenesis from H_2 + CO_2 in *A. woodii.*

Early experiments had revealed that the WLP was linked to the creation of a transmembrane sodium ion gradient in *A. woodii* and that this gradient enabled ATP generation (Heise et al., 1993). The F_1F_o ATPase responsible for this ATP generation was purified from the membrane of *A. woodii* (Aufurth et al., 2000; Reidlinger &

Müller, 1994; Spruth et al., 1995), its structure and the Na^+ cycle of *A. woodii* were elucidated in 2001 (Müller et al., 2001). Further improvements to the purification allowed for the isolation of the structure with its membrane motor (Fritz & Müller, 2007) and the expression of the ATPase gene cluster in *E. coli* allowed for elucidation of the essential role of the *atpI* gene for assembly of the ATPase (Brandt et al., 2013). Hypothesis of the existence of the Rnf complex for creation of the Na^+ gradient quickly followed (Müller et al., 2008). While evidence of its function has been reported in genetic, immunologic, and biochemical studies (Biegel & Müller, 2010; Biegel et al., 2009; Hess, Schuchmann, et al., 2013), it has not been successfully purified to date because of stability issues (Westphal et al., 2018).

Finally, during acetogenesis, the electrons are obtained from H_2 oxidation by the action of an electron-bifurcating hydrogenase, ferredoxin and NAD^+ are reduced simultaneously (Schuchmann & Müller, 2012). This hydrogenase system is described as the only source of ferredoxin reduction during acetogenesis (Bertsch et al., 2015; Schuchmann & Müller, 2014). However, *A. woodii* was recently found to contain an Stn gene cluster (Kremp et al., 2020) which was described as able to transfer the electrons from NADPH to NAD^+ and ferredoxin in the acetogen *Sporomusa ovata* (Kremp et al., 2020). In a recent study of formate metabolism in *A. woodii*, an *in silico* metabolic flux analyses predicted that the Stn complex consumes NADH and Fd^{2-} during fructose growth for NADPH production required for anabolic reactions (Neuendorf et al., 2021). The Stn complex was, however, not included in the transcriptomic and proteomic results in the same study (Neuendorf et al., 2021). In the same year, another study of the formate metabolism of *A. woodii* by transcript abundance (Moon et al., 2021) and a translational analysis of acetogenesis during heterotrophic (fructose) and autotrophic (H_2 + CO_2) growth (Shin et al., 2021) did not include the results for the Stn complex.

1.3.2 Feedstock diversity

A. woodii can utilise a variety of industrially relevant substrates as mentioned in section 1.1. Growth on $H_2 + CO_2$ has been extensively studied and *A. woodii* has been shown to be one of the most efficient autotrophs.

It was also shown to grow on 200 to 300 mM formate (Moon et al., 2021) and 60 mM methanol (Kremp et al., 2018). From four moles of formate, three moles are oxidized to H_2 and CO_2 to provide the electrons for reduction of the remaining mole of formate and one mole of CO_2. The electrons necessary are generated by the action of the hydrogenase and Rnf complex, creating a sodium ion transmembrane gradient. The gradient enables 0.3 ATP to be generated by the F_1F_o ATPase. In sum, four moles of formate result in one mole of acetate and 0.3 ATP. For growth on methanol, a methanol-specific methyltransferase system was identified, transferring methyl groups from methanol to THF to form methyl-THF. From 4 moles of methanol, one methyl group is oxidated by reversal of the methyl branch of the WLP and yields one ATP by substrate-level phosphorylation with the formyl-THF synthetase (Kremp et al., 2018). Three moles of ATP are generated by the acetate kinase during acetyl-CoA to acetate conversion (Kremp et al., 2018). The hydrogenase and Rnf complex are used for redox balancing to produce the required Fd^{2-} for CO_2 reduction. For the Rnf complex to generate Fd^{2-}, it has to import sodium ions, to maintain the sodium ion gradient, Na^+ is exported by the F_1F_o ATPase coupled to ATP hydrolysis (Kremp et al., 2018). In sum, four moles of methanol and two moles of CO_2 produce 3 moles of acetate and 2.5 ATP, this is equivalent to 0.83 ATP / mol of acetate and is this highest ATP gain reported after fructose (Kremp et al., 2018).

A. woodii cannot grow on CO alone but can co-utilise CO with $H_2 + CO_2$. However, while 5 to 15% CO led to higher growth rates, higher CO percentage were detrimental

for growth (Bertsch & Muller, 2015b). CO has been shown to reduce H_2 consumption by reversible inhibition of the HDCR responsible for CO_2 to formate conversion (Ceccaldi et al., 2017). Accordingly, when grown on formate, up to 25% CO led to higher optical densities and acetate production, higher percentages were still detrimental to growth and acetate production (Bertsch & Muller, 2015b).

Overall, *A. woodii* represents a potentially robust industrial chassis adaptable to a range of substrate combination. However, some problems persist for acetogenic gas fermentation to become prevalent on the market mainly, low product yield, low biomass density and low gas transfer into liquid (Liew et al., 2013). To try to solve these problems, fermentation optimization and genetic engineering are being investigated.

1.3.3 Culture conditions to optimise yields on $H_2 + CO_2$

To increase the rates of acetate production, fermentation conditions have been investigated. A continuous gas fermentation with a submerged membrane reactor overcame the low biomass problem with high cell retention but did not result in increased acetate concentration compared to batch process (Kantzow et al., 2015). Further work observed that immobilisation of *A. woodii* on a linen membrane successfully prevented the wash out and resulted in the highest acetate productivity during $H_2 + CO_2$ fermentation (Steger et al., 2017).

The main challenge of gas fermentation is gas-to-liquid mass transfer. The partial pressure of the headspace gases can be increased to improve their solubility. In *A. woodii*, the increase of hydrogen partial pressure, when grown on $H_2 + CO_2$, did not successfully increase acetate concentrations, rather, more formate was produced (Kantzow & Weuster-Botz, 2016). The authors concluded that the HDCR complex, responsible for formate production, is not a limiting factor while the subsequent steps

require ATP and THF, creating a bottleneck at the start of the WLP for utilisation of carbon dioxide and hydrogen (Kantzow & Weuster-Botz, 2016; Peters et al., 1999).

1.3.4 Plasmid-based strain engineering

Process optimisation through modulation of the growing conditions is limited by the intrinsic qualities of the microorganisms. Synthetic biology can provide tools for strain improvements through metabolic engineering.

The first genetic modification of *A. woodii* was reported in 1990 when a tetracycline resistance gene (*tetM*) of transposon Tn*925* was transferred from *Enterococcus faecalis* to *A. woodii* though filter mating (Strätz et al., 1990). The first successful DNA transfer in *A. woodii* was reported four years later based on the same resistance gene and the pAMβ1 Gram-positive replicon from *Streptococcus faecalis* (Stratz et al., 1994). The plasmid was transferred successfully by either electroporation or conjugation using the IncP helper protein. Plasmid presence was confirmed by gel electrophoresis and Southern hybridization, the plasmid was found to stably replicate in *A. woodii*, with 68% of plasmid maintenance after 32 generations without selection (Stratz et al., 1994).

However, 20 years passed before the first plasmid-based strain engineering of *A. woodii* was performed (Straub et al., 2014). Straub and co-workers achieved the overexpression of the four genes encoding the THF-dependent enzymes and of the combined phosphotransacetylase and acetate kinase from *C. ljungdahlii*. Both strategies resulted in increased acetate kinase activity and acetate production, demonstrating an increase in carbon flux through the WLP (Straub et al., 2014). In 2016, acetone production was achieved by expression of genes encoding thiolase A (*thlA*), acetoacetyl-CoA (*ctfAB*), and acetoacetate decarboxylase (*adc*) from *C. acetobutylicum* (Hoffmeister et al., 2016). During the development of this thesis, strategies for increased acetone production were

undertaken, which involved modification of the RBS of *adc* and *ctfA* to the native RBS and exchange of the *ctfAB* genes for the *ctfAB* genes of *C. aceticum* (Arslan et al., 2022). These studies enabled consolidation of plasmid transfer in *A. woodii* and better characterisation of the tools available.

Straub and co-workers reported transforming *A. woodii* by electroporation with a protocol adapted from *C. ljungdahlii* (Kopke et al., 2010), itself adapted from *C. tyrobutyricum* (Zhu et al., 2005). Compared to the method published in 1994 (Stratz et al., 1994), the electrical pulse (2.5 kV, 25 μF, 600 Ω compared to 10 kV, 25 μF, 400 Ω) and the recovery method differed (3 days compared to 7 h). The later engineering experiments were performed with slightly different electrical pulse again (0.625 kV, 25 μF, 600 Ω) and a 12 h recovery time. Hoffmeister and co-workers improved the competent cell preparation protocol to allow for their storage at -80ºC for up to 8 months (Hoffmeister et al., 2016), facilitating experimental procedures. The main difference between DNA transfer procedures is that while Straub and co-workers observed a requirement for methylation of the plasmid before transformation, others did not report this necessity. Straub and co-workers employed gene CLJU_c03310 from *C. ljungdahlii* (Philipps et al., 2019) for methylation and report that no plasmid transfer was obtained otherwise.

All the recently published experiments employed the chloramphenicol resistance gene (*catP*) from *C. perfringens* (Bannam & Rood, 1993). While the pIP404 Gram-positive origin of replication from *C. perfringens* led to successful gene overexpression (Bannam & Rood, 1993; Straub et al., 2014), pBP1 (from *C. botulinum*), pCB102 (from *C. butyricum*), and pCD6 (from *C. difficile*) were also found to allow plasmid replication (Hoffmeister et al., 2016). Hoffmeister and co-workers observed that replicons pCD6 and pCB102 resulted in the highest and lowest aetate concentration,

respectively. The authors noted that productivity is likely impacted by the specific copy number and segregational stability of each replicon (Hoffmeister et al., 2016). Finally, these studies demonstrated the functionality of a range of promoters in *A. woodii*. Acetate and acetone production was successful with the *pta-ack* promoter from *C. ljungdahlii*, but higher acetone yields were obtained with the thiolase promoter from *C. acetobutylicum* (Hoffmeister et al., 2016; Straub et al., 2014).

During the development of this thesis, the arginine deaminase pathway from *C. autoethanogenum* was expressed to improve ATP generation in *A. woodii* (Beck et al., 2020). Different promoters were tested for its expression. The anhydrotetracycline-inducible promoter (P_{tet}) from *Staphylococcus aureus* adapted for use in *C. difficile* (Fagan & Fairweather, 2011) and the *A. woodii* acetate kinase promoter modified with a theophylline riboswitch, both exhibited tight expression regulation. The highest expression was observed with the lactose/IPTG inducible (P_{fac}) and the constitutive promoters *pta-ack* from *C. ljungdahlii,* while the lactose-inducible promoter (P_{bgaL}) from *C. perfringens* exhibited a lower expression (Beck et al., 2020). Promoter characterisation is essential for regulation of expression of exogenous genes and later enabled lactate production by expression of the D-lactate dehydrogenase from *Leuconostoc mesenteroides* in a background strain lacking the native lactate dehydrogenase to prevent lactate degradation (see section 4.2.3.4) (Mook et al., 2022). More interestingly, butyrate production was achieved from fructose or methanol by expression of the genes coding for the thiolase (*thlA*), the 3-hydroxybutyryl-CoA dehydrogenase (*hbd*), the crotonase (*crt*), and the bifurcating butyryl-CoA dehydrogenase complex (*bcd/etfAB*) from *Eubacterium callanderi* (Chowdhury et al., 2022). While the concentrations obtained were low, this showed great promise for longer chain products in *A. woodii*.

Plasmid-based genetic engineering for acetate, acetone, lactate, or butyrate production confirmed the potential of genetic engineering for strain optimization and product diversification in *A. woodii* while consolidating the genetic tools available (Arslan et al., 2022; Chowdhury et al., 2022; Hoffmeister et al., 2016; Mook et al., 2022; Straub et al., 2014).

1.4 Genome engineering

While strain engineering by plasmid-based gene overexpression can confirm bottlenecks in the metabolic pathways (Straub et al., 2014) and confirm potential for product diversification (Arslan et al., 2022; Chowdhury et al., 2022; Hoffmeister et al., 2016; Mook et al., 2022), genetic tools are required to confirm gene function and enable further improvements of yields. Indeed, acetate was still observed in strains producing acetone, lactate, or butyrate, indicating that further yield improvements could be achieved by carbon redirection. Additionally, plasmid-based overexpression of genes requires the addition of antibiotics for plasmid maintenance and poses a risk to strain stability. Consequently, plasmid-based expression is not adequate for industrial processes and genome integration of pathways of interest is required.

Bacteria of the *Clostridium* genus either have inefficient Non-Homologous End Joining DNA repair systems (NHEJ) or do not have one (Joseph et al., 2018). Instead, they use homologous recombination for DNA repair by replacing chromosomal regions with identical or closely matching sequences. An exogenous DNA sequence can be placed between two regions homologous to the regions flanking a target locus in the genome. By recombination with the genomic locus, the exogenous DNA can be introduced in the chromosome. Each region of homology recombines separately with the genome. For successful modification of the native locus, both need to recombine. However, the

second crossover can lead to reversion of the locus to WT. For development of genome engineering tools in *Clostridium*, the bottleneck is the selection of the rare cells in which the double crossover successfully led to the desired modification of the chromosome.

1.4.1 Allelic exchange

Allelic exchange (AE), as the name suggests, is the exchange of one allele with another, often the wild-type native allele which is replaced by a mutant variant through homologous recombination. Although it is theoretically possible for two alleles to be exchanged through simultaneous recombination involving the two flanking regions of homology and a double crossover event, exchange normally takes place through two sequential steps initially involving one of the flanking regions of homology followed by a second single crossover event involving the second region of homology.

To facilitate sequential selection, the plasmids used do not contain a Gram-positive replicon and are termed suicide vectors. They integrate into the genome through homologous recombination in a single crossover event which is selected based on acquisition of a thiamphenicol resistance gene, (e.g., *catP*) by plating on media supplemented with this antibiotic. To select for the second crossover, a counter-selection, or negative selection, marker is required.

To overcome the limited availability of counter-selection markers, utilisation of the native *pyrE* (orotate phosphoribosyl transferase) gene has been demonstrated. The prior disruption of the chromosomal gene *pyrE* renders the cell an uracil auxotroph and 5-fluoroorotic acid (5-FOA) resistant. Allele-Coupled Exchange (ACE), a specialized allelic exchange method in which asymmetrical homology arms are employed, is first used to disrupt the native *pyrE* gene in the genome. The first crossover is guided by the longest homology arm and selected on thiamphenicol, the second is guided by the

second homology arm and selected on 5-FOA for cells with *pyrE* disruption. The Δ*pyrE* strain is subsequently used for genome engineering by standard Allelic Exchange, providing two homology arms for gene deletion or cargo delivery. The plasmid employed contains an exogenous *pyrE* gene for counter selection on 5-FOA of the second crossover. After generation of the desired genome modification, the *pyrE* locus is restored to WT with the ACE method used to disrupt it in the first place.

This technique has now been used in a variety of *Clostridium* species including *C. beijerinckii* (Little et al., 2018), *C. difficile (Ng et al., 2013), C. acetobutylicum* and *C. sporogenes* (Minton et al., 2016).

This method was successfully applied to *A. woodii* and published in 2022 (Baker et al., 2022). A 300 bp in-frame deletion was performed of the end of the *pyrE* gene with ACE. The resulting strain was used for different gene deletion as proof-of-concept using a suicide vector containing the *C. acetobutylicum pyrE* gene as a counter-selection marker and 700 bp homology arms (Baker et al., 2022). The method was further used to apply reverse genetics and confirm the function of the Rnf complex and of the lactate dehydrogenase by deletion of the entire Rnf gene cluster and genes *lctCDEF*, respectively (Schoelmerich et al., 2018; Westphal et al., 2018). More recently, *hydAB* from the hydrogenase gene cluster (Wiechmann et al., 2020) and the entirety of the HDCR cluster (Moon et al., 2023) were deleted. For those deletions a truncated Gram-positive replicon and the *pyrE* gene of *Eubacterium limosum* were used, but the general method is the same as described previously (Baker et al., 2022).

However, ACE requires the initial *pyrE* deletion to be successful and requires successive steps of selection for modification and for reversal of the *pyrE* locus to WT. While it

was shown to be functional in *A. woodii*, it is extremely time consuming and risks the emergence of SNPs and InDels in the genome (Baker et al., 2022).

1.4.2 CRISPR/Cas9

Overall, a more rapid and streamlined technique for genome engineering is required. In recent years, the Type II CRISPR/Cas9 from *Streptococcus pyogenes* has revolutionized the ability to manipulate the genome of many microorganisms. This adaptive prokaryotic immune-like system relies on the recognition of a target sequence matching a guide RNA, this allows for binding of the Cas9 nuclease and generation of a double-stranded break (DSB). For utilisation of CRISPR/Cas9 as a genome engineering tool in prokaryotes, homology arms are provided to allow for the repair of the DSB. As such, the CRISPR/Cas9 nuclease essentially functions as a counter-selection for the two crossovers as described in the allelic exchange method. Double crossover results in the deletion of the sequence matching the gRNA and survival of the cell.

However, Cas9 nuclease has been shown to be highly toxic to the cells and the challenge has been to modulate its expression to enable genome engineering. In *C. autoethanogenum*, no transformants could be obtained when *cas9* was expressed with a constitutive promoter but genome engineering was achieved using an inducible tetracycline promoter (Nagaraju et al., 2016). Recently, theophylline-inducible riboswitches were exploited to control the expression of Cas9, in a bio-part termed RiboCas, and led to successful engineering of the genomes of *C. pasteurianum*, *C. difficile* and *C. sporogenes* (Canadas et al., 2019).

1.5 Alternatives to CRISPR/Cas9

1.5.1 Cas9 variants

The structure of the Cas9 was elucidated (Nishimasu et al., 2014) and allowed modification of its catalytic function by mutation. Mutation of one of the nuclease domains (D10A) transforms Cas9 into a nickase (nCas9) and a mutation in both nuclease domains (D10A + H840A) results in a catalytically inactive Cas9 (dead Cas9, dCas9). These variants have been used to overcome the high toxicity of the Cas9 nuclease. Constitutive expression of the nickase enabled genome engineering of various genomic targets in *C. acetobutylicum, C. beijerinckii* (Q. Li et al., 2016), *C. cellulolyticum* (Xu et al., 2015). Interestingly, the Cas9 employed for successful modification of *C. ljungdahlii* (Huang et al., 2016) and *C. difficile* (Ingle et al., 2019) were found to be nickases, highlighting the toxicity of the unaltered Cas9 in *Clostridium*.

In contrast, dCas9 has no nucleolytic activity, while it cannot be used for genome engineering, it can still participate in reverse genetics thanks to its ability to bind DNA. The bound Cas9 blocks RNA polymerase, halting the transcription elongation, resulting in efficient silencing of transcription of the target gene (Larson et al., 2013), this technique is called CRISPR interference (CRISPRi). It was successfully applied to *C. acetobutylicum* and *C. beijerinckii* for study of gene function (Q. Li et al., 2016). In *C. ljungdahlii*, it allowed for redirection of the carbon flux (Woolston et al., 2018).

1.5.2 CRISPR/Cas12a

Type V CRISPR/Cas12a systems, also called CRISPR/Cpf1, have recently been explored. The Cas12a nuclease recognizes the PAM sequence TTN, an interesting alternative for AT-rich organisms compared to the NGG PAM sequence of Cas9 (Zetsche et al., 2015). Consequently, the CRISPR/Cas12a system from *Acidaminococcus* sp.

BV3L6 (AsCpf1) has recently been utilised for genome engineering of *C. difficile* (Hong et al., 2018) and *C. beijerinckii* (Zhang, Hong, et al., 2018). In those instances, the Cas12a nuclease was put under the control of a lactose inducible promoter to reduce its toxicity. When attempting to use this system in *C. ljungdahlii*, AsCpf1 was found highly toxic and Cas12a genes from other organisms were screened to enable plasmid transformation and genome engineering (Zhao et al., 2019).

1.5.3 Endogenous CRISPR/Cas systems

The main barrier to the widespread utilisation of CRISPR/Cas systems is the required expression of a toxic nuclease. In 2016, 74% of sequenced *Clostridium* genomes were found to contain an endogenous CRISPR system (Pyne et al., 2016). Consequently, harnessing of endogenous CRISPR systems could circumvent expression of the Cas9 or Cas12, as only the editing template and spacer for sequence targeting would be required for genome engineering. The necessity for prior characterisation of the endogenous system of interest can be time-consuming and prevents transferability between different bacteria. However, characterisation and leveraging of the system for genome engineering was achieved in *Clostridium pasteurianum* (Pyne et al., 2016), *Clostridium tyrobutyricum* (Zhang, Zong, et al., 2018), *Clostridium saccharoperbutylacetonicum* (Atmadjaja et al., 2019), *Clostridium butyricum* (Zhou et al., 2021), *C. difficile* (Maikova et al., 2019), and *Clostridium thermocellum* (Walker et al., 2020). In an interesting example, utilization of Cas9, nCas, or AsCpf1 was unsuccessful for genome engineering of *C. tyrobutyricum*, but the harnessing of its endogenous CIRPSR system was successful with editing efficiencies up to 100% (Zhang, Zong, et al., 2018).

However, these studies demonstrated that elucidation of the functional Protospacer Adjacent Motif (PAM) allowing for targeting of the DNA is a bottleneck for the method, requiring extensive *in silico* data mining or *in vivo* library construction and sequencing.

Strategies for spacer expression varied and the resulting editing efficiencies as well, likely because of variations in homologous recombination efficiencies and spacer copy number in the cells.

Despite these challenges, utilisation of endogenous CRISPR/Cas system represents an exciting alternative for bacteria displaying low transformation efficiencies of vector containing the Cas9 nuclease.

1.6 Forward genetics

As described above, Allelic exchange allowed for confirmation of the phenotype of Rnf, Hyd, and HDCR *A. woodii* deletion strains. This approach is an example of reverse genetics, for which the phenotype of a deliberately created genotype is determined. However, reverse genetics relies on preliminary characterisation of strains to provide targets for rational genetic engineering. In practice this limits the scope for its application to known features of the genome.

The knowledge of the WLP and energy conservation in *A. woodii* led to acetate yield improvements by gene overexpression (Straub et al., 2014). Other studies have managed to fine tune exogenous gene overexpression for acetone production (Hoffmeister et al., 2016). In both cases, the yields appeared to be limited by the intrinsic qualities of the bacteria studied. In 2021, transcriptomic and proteomic studies have allowed identification of genes upregulated for growth on different substrates (Moon et al., 2021; Neuendorf et al., 2021; Shin et al., 2021), but a better understanding of gene sets impacting strain fitness is required.

As opposed to reverse genetics, forward genetics aims at elucidating the genomic basis of a given phenotype. This approach can provide a more in-depth understanding of the dynamics active during growth and result in identification of bottlenecks for growth

kinetics and metabolite production. Forward genetics relies on the creation of a random mutant library from which individual cells with a demonstrable phenotype are selected and the identity of the gene affected is determined. In other words, the genotype of a particular phenotype is identified.

This approach requires that each mutant contains only one mutation, the mutation needs to be unbiased, and taken together the pooled mutants need to represent multiple mutations for each gene of the genome. An adequate screening step subsequently selects for the mutants displaying the phenotype of interest, sequencing of those mutants enable the identification of the genomic basis for the phenotype studied.

1.6.1 Transposon mutagenesis

The most commonly used method for mutation is transposon mutagenesis. Transposons are mobile genetic elements able to integrate into the genome through the action of a transposase. The transposase recognizes specific inverted terminal repeats (ITR) flanking a transposon sequence and relocates this transposon sequence into a different genomic locus. Transposition events thus disrupt the genomic features the transposon inserts into, allowing for characterisation of that gene or genomic region.

Next-generation sequencing (NGS) methods, such as Illumina, generate millions of sequencing reads. In the context of transposon libraries, this enables the simultaneous sequencing of all the mutants present in a pooled library by transposon insertion sequencing (TIS). In 2009, four separate TIS methods were published describing the high-throughput sequencing of transposon library samples (Gawronski et al., 2009; Goodman et al., 2009; Langridge et al., 2009; van Opijnen et al., 2009). Despite some differences, they rely on the same principles (van Opijnen & Camilli, 2013). Genomic DNA is extracted from the samples and fragmented, with some of those fragments

containing the transposon sequence. Transposon-genome junctions are subsequently enriched by PCR to enable sequencing of the loci of insertion.

As each mutant contains one insertion, the proportion of reads at each insertion site is proportional to the frequency of the mutant in the pooled library. Consequently, TIS methods are quantitative, and the composition of the population can be determined and genes can be assessed for their impact on strain fitness. Genes with low and high reads abundance can be hypothesized to impact strain fitness positively and negatively, respectively. With this method, genes important for specific conditions can be identified by sequencing of the pooled library before and after growth in selective conditions.

1.6.2 Transposon libraries and Fluorescent-Activated Cell Sorting (FACS)

In recent years, transposon libraries have been combined to different technologies in order to respond to intricate biological questions, notably with microfluidics systems such as flow cytometry and Fluorescence-Activated Cell Sorting (FACS). Multidrug efflux systems in *Acinetobacter baumannii* were identified by application of ethidium bromide to the cells before FACS (Hassan et al., 2016) and genes involved in toxin production were identified in *S. enterica* Typhi by building a library in a strain with a reporter gene in place of the toxin in the genome (Fowler & Galán, 2018).

Genome integration necessary for the development of commercial strains has been shown to lead to lowered yields. Additionally, in *E. coli*, transcription propensity was shown to depend on genomic position and random insertion of a fluorescent reporter gene resulted in 300-fold variations in fluorescence depending on insertion locus (Bryant et al., 2014; Scholz et al., 2019).

Transposon libraries combined with FACS can provide a new avenue for determination of genomic loci resulting in high gene expression. In *Bacillus subtilis*, this technique was applied by randomly inserting the Green Fluorescent Protein (GFP). A genome locus with 2.9-fold higher fluorescence compared to the commonly used *amyE* insertion site was identified (Jeong et al., 2018). Existence of a sensor strain with isobutanol-dependent fluorescence later enabled a more complex study for improved isobutanol yields in *E. coli*. A part or the complete isobutanol pathway was randomly inserted into the *E. coli* genome (Saleski et al., 2021). Individual cells from the resulting libraries were encapsulated with individual cells of the sensor strain and sorted with FACS. This method allowed for identification of insertion loci with isobutanol yields comparable to the yields obtained by plasmid-based expression of the pathway (Saleski et al., 2021).

1.6.3 *mariner*-based transposon system in *Clostridium*

In *Clostridium*, transposon libraries have been created using the *Himar1* transposable element. It was first successfully used for mutagenesis in *C. difficile* to elucidate spore formation, crucial for pathogenicity (Cartman & Minton, 2010). This system relied on expression of the transposase with the P_{tcdB} toxin promoter from *C. difficile*, active only in the presence of its cognate sigma factor (TcdR), avoiding transposition events in the *E. coli* cloning cells. It was then adapted to non-pathogenic *Clostridium* by inserting the *tcdR* element in the *pyrE* locus of the strains prior to the addition of the transposon delivery plasmid. It has been used successfully with this design in *C. beijerinckii* (Little et al., 2018), *C. acetobutylicum*, *C. sporogenes* (Zhang et al., 2015), and *C. autoethanogenum* (Woods et al., 2022). The presence of genome-integrated TcdR did not appear to modify the phenotype of these strains, but targeted mutagenesis in a clean wild-type background is required to confirm the gene functions (Minton et al., 2016).

1.6.4 Fluorescent reporter genes for anaerobic conditions

Enzymatic reporters such as the chloramphenicol acetyltransferase (*catP*) or the glucuronidase (*gusA*) genes have been predominantly used for tools development in acetogens (Mordaka & Heap, 2018). The β-galactosidase from *Thermoanaerobacterium thermosulfurigenes* was also used in *C. acetobutylicum* after the *E. coli* gene had been used widely in other bacteria (Tummala et al., 1999). For the combined application of transposon libraries with single cell measurements, a fluorescent reporter gene is required. Fluorescent reporters have been instrumental for the development and characterisation of synthetic biology techniques. They have allowed for characterisation of promoter expression with promoter assays or protein tagging. In acetogens, the required anaerobic conditions prevent the use of the most commonly used fluorescent proteins, such as the green fluorescent protein (*gfp*).

For this reason, oxygen-independent blue-light photoreceptors have been developed based on Light Oxygen Voltage (LOV) domains. Upon irradiation they change conformation to form a chromophore. The genes from *Bacillus subtilis*, *Pseudomonas putida*, and *Arabidopsis thaliana* have been shown to be active in a range of Gram-positive bacteria (Landete et al., 2015; Landete et al., 2017) and *Clostridium* (Bruder et al., 2016; Cho & Lee, 2017; Molitor et al., 2016). Notably, FACS was performed using the *Bacillus subtilis* gene in *C. acetobutylicum* (Cho & Lee, 2017).

More recently, the Fluorescence-Activating and Absorption-Shifting (FAST) tag protein was developed. Contrary to the LOV-domain proteins, it requires the reversible binding of a ligand for fluorescence measurements (Tebo et al., 2018). It was employed in *C. acetobutylicum* for microscopy, promoter assay and flow cytometry applications (Streett et al., 2019).

1.7 Project aims

This project revolves around exploiting acetogens as industrial chassis through development of synthetic biology tools and better strain characterisation, with an emphasis on *Acetobacterium woodii*.

Plasmid-based engineering strategies have shown that metabolic bottlenecks inherent to *A. woodii* hinder further developments. One of the aims of the project is to better characterise the required metabolic pathways and gene networks during heterotrophic and autotrophic fermentations to identify potential targets for genome engineering.

Additionally, the requirement for genome integration of exogenous pathways is set to become a challenge for production yields. As such, a second aim is to develop a tool to further the knowledge on the link between genomic loci, transcription levels, and resulting yields in acetogens.

Finally, to enable validation and leveraging of the results obtained with the aims discussed above, a reliable and streamlined genomic engineering tool needs to be developed for *A. woodii*.

CHAPTER 2: Materials and methods

2.1 Python script development

The Python script described in Supplementary Figure 1 was created in the PyCharm IDE (Integrated Development Environment) with Biopython modules. An object called Spacer was created to easily compile all the information collected on each hit for each spacer (Supplementary Table 1 for a full list of attributes collected) and to facilitate submission, retrieval and filtering of information from databases. INPUT - The user needs to enter an identification email and the GenBank accession number of the studied organism as well as a set of information about the CRISPR system which can be retrieved from the CRISPRFinder database (Couvin et al., 2018; Grissa et al., 2007a, 2007b). ARRAY – The array direction is set either with the CRISPRFinder predictions or an internal function submitting the flanks to BPROM (Solovyev & Salamov, 2010). If no orientation can be found the user is prompted to enter an orientation. BLAST – The protospacer list is submitted to BLASTn with a filtering function: global mismatch lower than 20% and no more than one mismatch in the first 7 nucleotides. PHASTER – Information on the hits is retrieved to update the phage attribute, whether directly from the position of the hits or after submission to PHASTER (Arndt et al., 2019; Zhou et al., 2011). PAM - If at least two of the hits are in phage regions, a consensus sequence and WebLogo graphical representation are then created to represent the potential PAMs (Crooks et al., 2004). OUTPUT – Confirmation of intermediate steps are displayed in the console, the final output of the script is an Excel summary document.

2.2 Chemicals

Unless otherwise stated all the chemicals were purchased from Sigma Aldrich.

2.3 Bacterial strains and plasmids

The following strains were used, *Escherichia coli* TOP10 (Invitrogen), WT *Acetobacterium woodii* strain DSM1030 (DSMZ) and WT *Clostridium autoethanogenum* strain DSM10061 (DSMZ). Strains are listed in Supplementary Table 2.

2.4 Growth media and conditions

E. coli was grown aerobically at 37°C, in a shaking incubator (225 rpm) when grown in broth, in LB medium supplemented with the appropriate antibiotic: chloramphenicol 25 µg/mL in plates and 12.5 µg/mL in liquid cultures, spectinomycin 250 µg/ml, erythromycin 500 µg/mL. *A. woodii* was grown anaerobically and statically at 30°C in either liquid *A. woodii* medium (AWM) or modified ATCC medium 1019 plates supplemented when needed with thiamphenicol 15 µg/mL or clarithromycin 6 µg/mL. For phenotypic work, ATCC media lacking yeast extract was used supplemented, when needed, with uracil or phenylalanine at 20 µg/mL. The trace element solution SL9, the selenite-tungstate solution and the vitamin solution were made separately and stored at 4°C. *C. autoethanogenum* was grown anaerobically and statically at 37°C in YTF media supplemented with the appropriate antibiotic: 15 µg/mL thiamphenicol in plates and 7.5 µg/mL in liquid cultures, 250 µg/ml D-cycloserine in plates. For phenotypic work, PETC MES agar plates were used supplemented when needed with uracil at 20 µg/mL. The list of ingredients for the media and the different solutions can be found in Supplementary Table 3, Supplementary Table 4, and Supplementary Table 5. Liquid and solid media to be used anaerobically were placed into anaerobic cabinets at least 24 h before use.

E. coli strains were stored using Microbank™ vials (Pro-Lab Diagnostics), by adding 700 µL of exponentially growing cultures, incubating at room temperature for 10 min, removing all the liquid, and finally storing at -80ºC. *A. woodii* and *C. autoethanogenum* strains were stored using 900 µL of exponential phase culture and 100 µL of DMSO (10% (v/v) DMSO), stocks were stored at -80ºC.

2.5 DNA manipulations

All the kits and enzymes described below were used according to the manufacturer's instructions unless otherwise stated.

Plasmid DNA was extracted from *E. coli* cultures using the Monarch® Plasmid Miniprep Kit (NEB). Genomic DNA extraction of *A. woodii* was performed using the DNeasy Blood and Tissue kit (Qiagen).

High fidelity DNA amplification was performed using Q5® High-Fidelity Master Mix (NEB) and colony PCRs with Taq polymerase with standard buffer (NEB). The melting temperatures of oligonucleotides were calculated using the NEB Tm Calculator online tool.

Agarose gels were prepared with 1-2% agarose (Sigma-Aldrich) in 1xTAE buffer (Thermo Fisher Scientific), and 0.001% v/v SYBR Safe DNA gel stain (Thermo Fisher Scientific). SeaPlaque™ low melt agarose (Lonza Bioscience, USA) was used for visualisation and size selection of the TraDIS PCR samples. DNA samples were mixed with 6x Gel Loading Dye (NEB) and loaded in the gels along with a 2-Log DNA ladder (NEB). Gels were run in a Gel Electrophoresis Tank (Axygen) at 90-120V for 30-45 min. Gels were imaged on ChemiDoc XRS+ (Bio-Rad).

Restriction enzymes and buffers were obtained from NEB. The digestions were set up in 20 µL with 1 µg of DNA and 1 µL of restriction enzyme(s), they were incubated at 37ºC for 1 to 2 h unless stated otherwise. DNA was visualized on agarose gels and purified using Monarch® DNA Gel Extraction Kit (NEB). When purification was not performed, the enzymes were inactivated according to the manufacturer's instructions.

When needed, dephosphorylation was performed using Antarctic phosphatase (NEB).

Ligations were performed using T4 DNA Ligase and buffer (NEB). The reactions were set up in in 20 µL and incubated for 16 h at 16ºC.

When restriction ligation techniques were not possible, DNA fragments were assembled by HiFi assembly using NEBBuilder® HiFi DNA Assembly Master Mix (NEB).

DNA sequencing was performed by Eurofins or Genewiz interchangeably.

2.6 Plasmid construction

All the oligonucleotides used for molecular biology are listed in Supplementary Table 6, they were ordered from Sigma-Aldrich and Integrated DNA Technologies interchangeably. Some of the plasmid backbones were taken from the culture collection of the laboratory. All the plasmids were built according to the pMTL80000 shuttle plasmid standards (Heap et al., 2009). All plasmids used are listed in Supplementary Table 7.

Plasmids needed for this study and the colony PCR reactions to verify their assembly or their presence are described hereafter, when relevant. Colony PCR reactions to verify the corresponding genome locus in *A. woodii* or *C. autoethanogenum* are also described.

2.6.1 Chapter 3

2.6.1.1 Validation of R-M recognition site impact on *A. woodii* transformation

The target site together with its surrounding sequence in *cas9* was generated by annealing primers FW_targetIIP and RV_targetIIP and inserting the 60 bp fragment formed between the NotI and XbaI sites of plasmid pMTL82151 by restriction digestion followed by ligation with T4 DNA ligase. *A. woodii* was transformed in triplicate with the resultant plasmid, pMTL8215_targetIIP. pMTL82151 was used as a control. Transformation efficiencies of pMTL8215_targetIIP were expressed as percentage of the transformation efficiency of pMTL82151.

Colony PCRs were performed with primers ColE1+tra-F2 / pBP1-R1 and plasmids confirmed by Sanger sequencing.

2.6.1.2 Validation of *cas9* nuclease toxicity

To remove the 5′-TAAGN_5TCC-3′ R-M recognition site from the *cas9*, a silent mutation was introduced at position Y1001, creating the modified motif 5′-TAAGN_5CCC-3′. Primer pairs RV_Pfdx_XbaI/FW_IIPmut and RV_IIPmut/FW_NotI_cas9 were used to amplify two halves of the *cas9* gene from the pMTL8215_pRECas1_MCS vector with the mutation introduced in the middle. The halves were then assembled by SOEing PCR (Splicing by Overlap-Extension PCR) with primer pair RV_Pfdx_XbaI/FW_NotI_cas9. The mutated *cas9* gene was inserted back in the pMTL8215_pRECas1_MCS vector by XbaI / NotI digestion and ligation with T4 DNA ligase to generate vector pMTL8215_pRECas1_IIPmut.

To introduce the two mutations, D10A and H840A, abolishing the nuclease activity of the Cas9, the gene was amplified from the pMTL8215_pRECas1_MCS vector as three

separate DNA fragments with primer pairs RV_Pfdx_XbaI/FW_D10A, RV_D10A/FW_H840A and RV_H840A/FW_NotI_cas9. The three fragments were assembled by SOEing PCR with primer pair RV_Pfdx_XbaI/FW_NotI_cas9. The mutated *cas9* gene was inserted back in the pMTL8215_pRECas1_MCS vector by XbaI / NotI digestion and ligation with T4 DNA ligase to generate vector pMTL8215_pREdCas1_MCS.

To create the dead *cas9* vector with the R-M site mutated, the gene was amplified from the pMTL8215_pRECas1_MCS vector as four DNA fragments with primer pairs RV_Pfdx_XbaI/FW_D10A, RV_D10A/FW_H840A, RV_H840A/FW_IIPmut and RV_IIPmut/FW_NotI_cas9. The four fragments were then assembled by SOEing PCR with primer pair RV_Pfdx_XbaI/FW_NotI_cas9. The mutated *cas9* gene was inserted back in the pMTL8215_pRECas1_MCS vector by XbaI / NotI digestion and ligation with T4 DNA ligase to generate vector pMTL8215_pREdCas1_IIPmut.

A vector in which *cas9* was absent was created as a control, *cas9* and its P_{fdx-E} promoter were removed from the pMTL8215_pRECas1_MCS vector. XbaI / NotI digestions followed by blunt-ending with T4 DNA polymerase and re-ligation with T4 DNA ligase yielded vector pMTL8215_P1339_MCS.

A. woodii WT was transformed with all the vectors described in triplicate.

2.6.1.3 Interference assays plasmids

The following primer pairs were annealed with Q5: FW_proto8 / RV_proto8, FW_proto8-5′ / RV_proto8, FW_proto20 / RV_proto20, FW_proto20-5′ / RV_proto20, FW_proto22 / RV_proto22 and FW_proto22-5′ / RV_proto22 to yield the three protospacers with and without their PAM candidate. They were then inserted by HiFi in

a SacII / SalI digested pMTL82151 backbone to create pMTL-MPD1 to 6. Colony PCRs and Sanger sequencing were performed with primers ColE1+tra-F2 and pBP1-R1.

2.6.1.4 Endogenous leader characterization

An MCS from pMTL82151 was inserted after the leader sequence in pMTL-MPD21 by digestion with SacI / NheI and ligation with T4 DNA ligase creating pMTL-MPD15. The FAST gene and the *catP* gene were inserted in the MCS after the leader sequence for characterization of the leader sequence by RT-qPCR and to serve as a negative control, respectively. The *catP* and FAST genes were amplified with primer pairs FW_catP_SacI / RV_catP_NheI and FW_FAST_SacI / RV_FAST_NheI, respectively. The pMTL-MPD15 backbone and both genes were digested by SacI and NheI. The *catP* and FAST genes were ligated in the backbone with T4 DNA Ligase (NEB), yielding pMTL-MPD30 and pMTL-MPD29, respectively. All plasmids were confirmed by Sanger sequencing with primers ColE1+tra-F2 and pCD6-R1.

2.6.1.5 *pyrE* KO in *A. woodii*

The editing template consisted of two homology arms of 471 bp (LHA) and 491 bp (RHA). The spacer was designed to be 36 bp and at a distance of 29 bp from the RHA.

The editing cassette was assembled in plasmid pMTL8215_P1339_MCS. HAs were amplified with primer pairs FW_LHA_pyrE / RV_LHA_pyrE+BM4 and FW_BM4+1+RHA_pyrE / RV_RHA_pyrE from *A. woodii* genomic DNA, assembled by SOEing PCR ligated to the vector after AscI / AatII digestion, creating pMTL-MPD7.

A. woodii leader sequence was amplified with primers FW_leader / RV_leader, the synthetic array was created by annealing primers FW_pyrE_endo2 and RV_pyrE_endo2. Both leader sequence and synthetic array were assembled by HiFi into pMTL-MPD7 digested by SacII / AatII, yielding pMTL-MPD8.

The editing cassette from pMTL-MPD8, amplified with primers FW_leader and RV_RHA_pyrE+MCS, pMTL82151, and pMTL84151 were digested by AatII / NheI, SacII / AatII or SacII / NheI and ligated together to obtain constructs with HAs alone, expressed spacer alone or the whole editing cassette, respectively, with either Gram-positive replicon, yielding plasmids pMTL-MPD9 to 14.

Colony PCRs were performed with primers ColE1+tra-F2 / pBP1-R1 or ColE1+tra-F2 / pCD6-R1 and plasmids confirmed by Sanger sequencing. The same primers were used to check for plasmid presence in *A. woodii* and the *pyrE* locus was analysed with primers AW_CRISPR_pyrE_armF and AW_pyrEcomp_RHAR.

2.6.1.6 *pheA* KO in *A. woodii* with 500 bp, 1.0 kb and 1.5 kb HA

The spacer was designed to be 35 bp and at a distance of 26 bp from the RHA. Homology arms of 0.5, 1.0 and 1.5 kb were designed. The resulting editing cassettes were inserted into a backbone with the pCD6 Gram-positive replicon.

The spacer sequence was created by annealing primers FW_pheA_spacer and RV_pheA_spacer by PCR and digestion by SacI and AatII; the leader sequence was created by amplification with FW_leader / RV_leader and SacII and SacI digestion; both fragments were assembled in pMTL84151 digested by SacII / AatII. Homology arms of 0.5 kb, 1.0 kb and 1.5 kb were designed. Left homology arms were amplified with primers FW_L0.5 / L1.0 / L1.5 and RV_LHA_pheA; right homology arms were amplified with primers FW_RHA_pheA and RV_R0.5 / R1.0 / R1.5. LHA and RHA were assembled by SOEing PCR with primer pairs FW_L0.5 / RV_R0.5, FW_L1.0 / RV_R1.0 and FW_L1.5 / RV_R1.5, respectively. Assembled HAs were ligated in both pMTL84151 and pMTL84151 with assembled leader and spacer after AatII / NheI digestion, yielding three plasmids containing the HAs alone (pMTL-MPD16 to 18) and

three editing plasmids containing the expressed spacer and the HAs (pMTL-MPD19 to 21).

Colony PCRs were performed with primers ColE1+tra-F2 / pCD6-R1 and plasmids confirmed by Sanger sequencing. *A. woodii* colonies were analysed by colony PCR both for plasmid presence with primers ColE1+tra-F2 and pCD6-R1 and for the *pheA* locus with primers FW_pheAHA1.5_out and RV_pheAHA1.5_out.

2.6.1.7 *hsdR1* KO in *A. woodii*

Homology arms of 1,455 bp (LHA) and 1,489 bp (RHA) and a 36 bp spacer starting 40 bp from the LHA were designed.

Primers FW_hsdR1_spacer and RV_hsdR1_spacer were annealed by PCR and inserted in the pMTL84151 vector with assembled leader and spacer used for *pheA* KO by SacI and AatII digestion and ligation with T4 DNA ligase, resulting in vector pMTL8415_leader_hsdR1_spacer. The left and right homology arms were amplified with the primers FW_LHA_hsdR1 / RV_LHA_hsdR1and FW_RHA_hsdR1 / RV_RHA_hsdR1 respectively and assembled by SOEing PCR with the FW_LHA_hsdR1 / RV_RHA_hsdR1 primer pair. The vector containing the spacer and the assembled HAs were digested by AatII and NheI and ligated together with T4 DNA ligase, resulting in the final editing vector for *hsdR1* KO, pMTL-MPD22.

Colony PCRs were performed with primers ColE1+tra-F2 / pCD6-R1 and plasmids confirmed by Sanger sequencing. Cell densities and amount of DNA were investigated as potential factors to increase editing efficiency as follows: 100 µL of cells and 1 µg of DNA (T1); 100 µL of high density cells and 1 µg of DNA (T2); 200 µL of high density cells and 4 µg of DNA (T3). *A. woodii* colonies were analysed by PCR for plasmid

presence with primers ColE1+tra-F2 / pCD6-R1 and the *hsdR1* locus with primers FW_hsdR1_screen_1.5 and RV_hsdR1_screen_1.5.

2.6.1.8 *pyrE* KO in *C. autoethanogenum*

Homology arms of 868 bp and 797 bp were used and the synthetic array was designed to mimic *C. autoethanogenum* native array 4, with the native leader sequence controlling the expression of a 35 bp spacer flanked by direct repeats. The editing cassette was inserted in a pMTL83151 backbone.

pMTL83151 and CRISPR array leader amplified with FW_leader4_CLAU / RV_leader4_CLAU were digested by SacII / SacI and ligated together to yield pMTL-MPD25. Primers FW_CLAUpyrE_spacer1 and RV_CLAUpyrE_spacer1 were annealed by PCR and inserted in pMTL-MPD25 by SacI and BamHI digestion and ligation with T4 DNA ligase yielding pMTL-MPD27. This plasmid and homology arms amplified with FW_CLAUpyrE_HA / RV_CLAUpyrE_HA from plasmid vFS67 (provided by Dr. Francois Seys) were digested by BamHI / NheI and ligated together to yield the final editing vector, pMTL-MPD28. An intermediate vector (pMTL-MPD26) containing the HAs alone was obtained by digesting pMTL83151 and the HAs by BamHI / NheI and ligation.

Colony PCRs were performed with primers ColE1+tra-F2 and pCB102-R1 and plasmids confirmed by Sanger sequencing. *C. autoethanogenum* colonies were analysed by PCR for plasmid presence with ColE1+tra-F2 and pCB102-R1 and the *pyrE* locus analysed with FW_pyrECLAU_screening and RV_pyrECLAU_screening.

2.6.2 Chapter 5

2.6.2.1 Isopropanol production with *C. beijerinckii* SADH

Isopropanol production was established in *A. woodii* using the SADH from *C. beijerinckii*. The gene was inserted either at the end of the acetone operon in plasmid pMTL24-1 (pMTL24-1+Adh) or alone, under the control of the *C. difficile* P_{tcdB} promoter (pMTL84151_PtcdB_Cbe_Adh).

Primers FW_hifi_adc_Cbe_adh and RV_hifi_Cbe_adh were used to amplify the *adh* gene along with its RBS from *C. beijerinckii* NRRL B592 genomic DNA. The amplicon and pMTL24-1 digested by Nhe1 were assembled by HiFi, resulting in pMTL24-1+Adh. Colony PCRs were performed with primers FW_seq_thlA and FW_seq_Cbe_adh. Plasmids were confirmed by Sanger sequencing with primers FW_seq_thlA, FW_seq_Cbe_adh, and pCD6-R1.

Primers FW_hifi_Pfdx_Cbe_adh and RV_hifi_Cbe_adh were used to amplify the *adh* gene from *C. beijerinckii* NRRL B592 genomic DNA. The amplicon and plasmid pMTL24-1 were digested by NdeI and NheI and assembled by ligation with T4 DNA ligase resulting in pMTL84151_PtcdB_CbeSADH. Colony PCRs were performed with primers ColE1+tra-F2 and pCD6-R1. Plasmids were confirmed by Sanger sequencing with the same primers.

2.6.2.2 Overexpression of *A. woodii* gene *gatD*

Primers FW_AWO_gatD and RV_AWO_gatD were used to amplify the *gatD* gene from *A. woodii* genomic DNA. The amplicon and plasmid pMTL24-1 were digested by NdeI and NheI and assembled by ligation with T4 DNA ligase resulting in pMTL84151_PtcdB_AwoGatD. Colony PCRs were performed with primers ColE1+tra-F2 and pCD6-R1. Plasmids were sent for Sanger sequencing with the same primers.

2.6.2.3 *gatD* KO in *A. woodii*

Homology arms of 630 bp and 1,000 bp and a 36 bp spacer were designed. Primers FW_gatD_spacer and RV_gatD_spacer were annealed by PCR and inserted in the pMTL84151 vector with assembled leader and spacer used for *pheA* KO by SacI and AatII digestion and ligation with T4 DNA ligase. The left and right homology arms were amplified with primer pairs FW_gatD_LHA / RV_gatD_LHA and FW_gatD_RHA / RV_gatD_RHA respectively and assembled by SOEing-PCR with the FW_gatD_LHA / RV_gatD_RHA primer pair. The assembled HAs were inserted in the vector containing the spacer by AatII and NheI digestion and ligation by T4 DNA ligase, resulting in the KO vector pMTL8415_GatD_KO.

Plasmids were confirmed by Sanger sequencing with ColE1+tra-F2 / pCD6-R1

A. woodii with *tcdR* at the *pyrE* locus and P_{tcdB}_acetone was transformed with the editing vector following the protocol named T3 in the previous section, with 200 µL of high density cells and 4 µg of DNA. *A. woodii* colonies were analysed by colony PCR for plasmid presence with primers ColE1+tra-F2 / pCD6-R1 and the *gatD* locus with primers FW_gatD_seq and RV_gatD_seq.

2.6.2.4 Study of creiLOV, Bs2 and Pp1 fluorescent proteins

Genes creiLOV, Pp1, and Bs2 were ordered on IDT as GeneBlocks (Supplementary Table 8), the first attempts at expressing them was in the backbone pTML83251. This strategy resulted in numerous mutations and the cloning strategy was adapted to assemble them in a vector with the low copy number *E. coli* replicon p15a.

The P_{thl} promoter was amplified from pMTL84422 with primers FW_Pthl_backbone / RV_Pthl and the creiLOV, Pp1, and Bs2 genes with FW_creiLOV_Pthl / RV_creiLOV_backbone, FW_evoglow-Pp1_Pthl / RV_evoglow-Pp1_backbone, and

FW_evoglow-Bs2_Pthl / RV_evoglow-Bs2_backbone, respectively. Promoter and genes were assembled by SOEing PCR and assembled in the pMTL82221 by NdeI and XhoI digestion and ligation with T4 DNA ligase, yielding plasmids pMTL82222_creiLOV, pMTL82222_Pp1, and pMTL82222_Bs2. Sequencing with p15a-F1 and pBP1-R1 revealed numerous mutations once more and fluorescence measurement on a TECAN M1000 Pro (Tecan) plate reader confirmed abolishment of fluorescence when the mutations are present (Supplementary Figure 2).

After characterisation of the R-M system in section 3.2.1.1 and confirmation of the riboflavin role in mutation generation during the cloning process, The genes were cloned again in the pMTL84153 backbone, the GeneBlocks for creiLOV, Pp1, and Bs2 were amplified with FW_NdeI_creiLOV_corrected / RV_AatII_creiLOV, FW_NdeI_Pp1_corrected / RV_AatII_Pp1, and FW_NdeI_Bs2_corrected / RV_AatII_Bs2, respectively. They were assembled in backbone pMTL84153 by NdeI / AatII digestion and ligation with T4 DNA ligase. Pp1 was also assembled in pMTL84152 by NdeI / AatII digestion and ligation with T4 DNA ligase. Colony PCRs and Sanger sequencing was performed with primer pair ColE1+tra+F2 and pCD6-R1.

2.6.2.5 FAST gene KI at the *pheA* locus in *A. woodii*

The spacer is 36 bp, homology arms are 1,496 bp and 1,465 bp and were designed for insertion of the cargo in place of the PAM. The cargo consists of the FAST reporter gene flanked by the thiolase (*thl*) promoter (P_{thl}) and *pyrE-hydA* (orotate phosphoribosyltransferase and hydrogenase I) terminator from *C. acetobutylicum* and was inserted between the homology arms.

Primers FW_AWO_pheA_CCG_KI_spacer and RV_AWO pheA_CCG_KI_spacer were annealed by PCR and inserted in the pMTL84151 vector with assembled leader

and spacer used for *pheA* KO by SacI and AatII digestion and ligation with T4 DNA ligase. The left and right homology arms were amplified with primers FW_AWO_LHA_pheA_cargo / RV_AWO_LHA_pheA_CCG_cargo_MCS and FW_AWO_RHA_pheA_CCG_cargo_MCS / RV_AWO_RHA_pheA_cargo respectively and assembled by SOEing PCR with the FW_AWO_LHA_pheA_cargo / RV_AWO_RHA_pheA_cargo primer pair. The assembled HAs were inserted in the vector containing the spacer by AatII and NheI digestion and ligation by T4 DNA ligase, resulting in a KI vector (pMTL-MPD23) containing an MCS with BamHI, NotI, NdeI and XbaI restrictions sites between the two homology arms.

A P_{thl}_FAST_term cassette was amplified with primer pair FW_BamHI_FAST_cassette / RV_FAST_NotI_AscI from plasmid pMTL84152_P_{thl}_FAST with the FAST gene flanked by the *thl* (thiolase) promoter and *pyrE-hydA* (orotate phosphoribosyltransferase and hydrogenase I). The cassette was then inserted into the MCS of pMTL-MPD23 vector by XbaI and NotI digestion and ligation with T4 DNA ligase, yielding pMTL-MPD24.

Plasmids were confirmed by Sanger sequencing with ColE1+tra-F2 / pCD6-R1 / FW_pheA_cargo_seq / RV_pheA_cargo_seq.

WT *A. woodii* was transformed with the final editing vector, pMTL-MPD24 following the protocol named T3 in the previous section, with 200 µL of high density cells and 4 µg of DNA. *A. woodii* colonies were analysed by colony PCR for plasmid presence with primers ColE1+tra-F2 / pCD6-R1 and the *pheA* locus with primers FW_pheA KI_seq and RV_pheA KI_seq.

2.6.2.6 Terminator characterisation

To evaluate the *pyrE-hydA* (orotate phosphoribosyltransferase and hydrogenase I) terminator from *C. acetobutylicum* and a terminator from *C. tetani* (the terminators are referred as "T" and "CT" in the following paragraphs, respectively), the following plasmids were assembled. Testing of transcriptional termination was assessed by the ratio of transcript level on either side of each terminator. *catP* was placed between the thiolase promoter and the terminator, while FAST was placed after the terminators. The cassettes containing these elements were assembled in a backbone containing the pCD6 Gram-positive replicon and the erythromycin resistance gene (*ermB*).

The terminator from *C. tetani* was inserted in pMTL84151_Pthl_FAST. The terminator was retrieved from CRG4680 by NotI digestion and vector pMTL84151_Pthl_FAST was digested by NotI and dephosphorylated. The terminator was inserted in the vector by ligation with T4 DNA ligase, yielding pMTL84151_Pthl_FAST_CT, it was confirmed by sequencing with primers ColE1+tra-F2 and pCD6-R1.

The *catP* gene was amplified from a pMTL84151_Pthl_FAST with primers FW_catP_NdeI / RV_catP_NdeI and digested by NdeI. Terminator T was amplified from pMTL84151_Pthl_FAST with primers FW_T and pCD6-R1 and terminator CT from pMTL84151_Pthl_FAST_CT with primers FW_CT and pCD6-R1, both were digested by NdeI and AscI. The catP gene and either terminator were assembled in vector pMTL84252 digested by NdeI and AscI, yielding pMTL84252_catP_T and pMTL84252_catP_CT. They were confirmed by Sanger sequencing with primers ColE1+tra-F2 and pCD6-R1.

FAST amplified with primers FW_RBS_FAST_AscI and RV_FAST_AscI. The amplicon and the two vectors pMTL84252_catP_T and pMTL84252_catP_CT were

then digested by AscI, the vectors were dephosphorylated, the FAST gene was subsequently ligated in both vectors with T4 DNA ligase, yielding pMTL84252_catP_T_FAST and pMTL84252_catP_CT_FAST. They were confirmed by Sanger sequencing with primers catPrep-F1 and pCD6-R1.

2.6.2.7 Transposon library plasmid with FAST reporter gene

Primers FW_Pthl_ITR_GL15 and RV_pyrEhydA_GL15 were used to amplify the FAST cassette with the thiolase promoter P_{thl} and the *pyrE-hydA* terminator from the pTML84151_Pthl_FAST plasmid. Primers FW_catp_cassette_corrected and RV_XbaI_CatP. catP and FAST were assembled by SEOing PCR with primers FW_cassette_GL15 and RV_cassette_GL15. The assembled catP-FAST cassette and plasmid pMTL-GL15 were digested by EcoRI and XbaI and ligated together by ligation with T4 DNA ligase resulting in pMTL-GL15-FAST. Plasmids were confirmed by Sanger sequencing with the primers ColE1+tra-F2, catP-R1, and Eurofins M13 rev -29.

2.6.2.8 *C. beijerinckii* SADH gene KI at the *purD* locus in *A. woodii*

First, a vector for *purD* KI was created with an MCS between the HAs. The *purD* spacer was generated by annealing primers FW_purD_spacer and RV_purD_spacer. The LHA and RHA were amplified from *A. woodii* genomic DNA from primer pairs FW_purD_LHA / FW_purD_LHA and FW_purD_RHA / RV_purD_RHA, respectively. The HAs were then assembled by SOEing PCR with primers FW_purD_LHA and RV_purD_RHA. The spacer was digested by SacI / AatII and the HAs by AatII / NheI, both were assembled in pMTL-MPD1 digested by SacI / NheI and dephosphorylated by ligation with T4 DNA ligase. Colony PCRs and sequencing of the resulting plasmid were performed with primers ColE1+tra-F2 and pCD6-R1.

The PtcdB_Cbe_adh fragment was amplified with primers FW_NotI_PtcdB / RV_XbaI_adh from the pMTL84151_PtcdB_CbeSADH vector. The fragement and the plasmid described above were digested by NotI and XbaI and ligated together by T4 DNA ligase, yielding pMTL84151_purD_KI_PtcdB_CbeSADH. The plasmid was confirmed by sequencing with primers ColE1+tra-F2, pCD6-R1, FW_purD_cargo_seq, and RV_purD_cargo_seq.

A. woodii *tcdR*/P_{tcdB}_acetone was transformed with plasmid pMTL84151_purD_KI_PtcdB_CbeSADH following the protocol named T3 in the previous section, with 200 µL of high density cells and 4 µg of DNA. *A. woodii* colonies were analysed by colony PCR for plasmid presence with primers ColE1+tra-F2 and pCD6-R1 and the purD locus with primers FW_purD_KI_seq and RV_purD_KI_seq.

2.6.2.9 Acetone operon KI at the *pheA* locus in *A.woodii*

pMTL84151_pheA_KI_PtcdB_Cbe_adh was digested by NotI and AscI, resulting in three fragments, the fragment containing PtcdB_Cbe_adh was discarded. The fragment containing the backbone was dephosphorylated and re-assembled with the fragment containing the RHA and a PtcdB_acetone fragment obtained by digestion of pMTL24-1 with NotI and AscI by ligation with T4 DNA ligase, yielding pMTL84151_pheA_KI_PtcdB_acetone. Plasmid was confirmed by Sanger sequencing with ColE1+tra-F2 / pCD6-R1 / FW_seq_ctfA / RV_seq_ctfA / FW_seq_thlA / FW_pheA_cargo_seq / RV_pheA_cargo_seq / JPB_475 / JPB_476 / JPB_657 / JPB_693 / JPB_694

A. woodii tcdR was transformed following the protocol named T3 in the previous section, with 200 µL of high density cells and 4 µg of DNA. *A. woodii* colonies were analysed by colony PCR for plasmid presence with primers ColE1+tra-F2 / pCD6-R1

and the *pheA* locus with primers FW_pheA_KI_seq / RV_pheA_KI_seq. The colony PCR amplicons were confirmed by Sanger sequencing with primers FW_pheA_KI_seq / JPB_476 / RV_pheA_cargo_seq / JPB_694 / JPB_657 / FW_pheA_cargo_seq / JPB_475 / RV_pheA_KI_seq.

2.7 Strain engineering

2.7.1 *Escherichia coli*

2.7.1.1 Chemically competent cells

An overnight culture of *E. coli* TOP10 was used to freshly inoculate 100 mL of LB Medium. Once the OD_{600nm} reached 0.4 - 0.5, the culture was incubated at 4°C for 30 min. The cells were then collected by centrifugation at 5,000 rpm for 5 min at 4°C. Cells were washed and incubated 15 min with ice cold 0.1M $MgCl_2$. A second wash and incubation were performed with ice cold 0.1M $CaCl_2$. Finally, the cells were resuspended in 1 mL of ice cold 0.1M $CaCl_2$ with 20% Glycerol. Cells were then aliquoted for storage and kept at -80°C until needed.

2.7.1.2 Transformation and colony PCR

Briefly, 50 µL of chemically competent *E. coli* TOP10 cells were incubated with DNA for 30 min on ice, then 30 sec at 42°C, 5 min on ice, and a final 1 h recovery in SOC medium before plating on selective LB agar plates. Colonies were checked by colony PCRs to verify the presence of the plasmids using Taq with standard buffer (New England Biolabs). The plasmids were then extracted using Monarch® Plasmid Miniprep Kit (New England Biolabs) and confirmed by Sanger sequencing.

2.7.2 *Acetobacterium woodii*

2.7.2.1 Competent cells: standard and high density

A 48 h liquid culture of the desired strain of *A. woodii* was used to inoculate a 300 mL culture to reach an OD_{600nm} of around 0.3 and 0.4 overnight. The cells were collected and then washed twice with filter-sterilized SMP5.8 buffer (1 mM sodium phosphate pH5.8, 1 mM MgCl2, and 270 mM sucrose). In a final step they were resuspended in 2 mL of SMP5.8 buffer for standard competent cells or in 666 µL for high density competent cells and 100% DMSO was added to reach a 10% concentration. Cells were aliquoted for storage and kept at -80°C until needed.

2.7.2.2 Transformation and colony PCR

A. woodii was transformed by electroporation using either 100 µL of competent cells and 1 µg of plasmid in 20 µL of dH_2O or 200 µL of high density competent cells and 4 µg of plasmid in 20 µL of dH_2O. Electroporation was done on a Gene Pulser Xcell™ electroporator (Bio-Rad). After 5 h (unless stated otherwise) of recovery in AWM, transformations were plated on Modified AWM medium containing the appropriate selection or supplementation. After 10 days, the transformation efficiencies were recorded. Colonies were then cultured in liquid AWM, after 48 h colony PCRs were performed for plasmid presence and genome locus, when appropriate, using Taq DNA Polymerase (NEB). The PCR reactions were subsequently confirmed by Sanger sequencing before any further work on the strains.

2.7.2.3 Plasmid loss

After mutant strain confirmation, a succession of inoculations into liquid broth and agar plates were performed for plasmid loss. Colonies were inoculated onto non-selective plates either directly from the liquid cultures with selection used for colony PCRs or

from a subculture of it without selection. From those non-selective plates, single colonies were grown in liquid cultures without selection and then streaked on non-selective plates. Single colonies were streaked on plates with and without selection to check for plasmid loss.

2.7.3 *Clostridium autoethanogenum*

2.7.3.1 Conjugation and colony PCR

YTF was inoculated to an OD_{600nm} of 0.02 from a late exponential (48-72 h) preculture of the *C. autoethanogenum* recipient strain and incubated overnight. *E.coli* NEB sExpress conjugation donor (Woods et al., 2019) harbouring the shuttle vector was inoculated and incubated overnight in LB supplemented with Spectinomycin and the additional antibiotic for the shuttle vector.

The next day, LB broth containing appropriate antibiotic selection was inoculated with the *E. coli* donor strain and incubated until OD_{600nm} 0.2 - 0.4 was reached. A 1 mL aliquot of the *E. coli* donor strain culture was collected at 3,000 x g for 3 min, washed in 500 µL of PBS through gentle flicking of the tube and centrifuged again at 3,000 x g for 3 min. The pellet was then transferred into the anaerobic cabinet and resuspended in 200 µL of the *C. autoethanogenum* overnight culture. The mixture was spotted onto non-selective YTF agar plates and incubated overnight (16-24 h on mating plate).

The next day, the growth on the non-selective mating plate was collected using 500 µL of anaerobic PBS and a wedge-shaped spreader. The slurry was then spread onto agar plates with appropriate selection for the shuttle vector and D-cycloserine as *E. coli* counter-selection. Plates were incubated under anaerobic conditions for 3-4 days. Once colonies were visible, they were inoculated onto fresh agar plates with selection, single colonies were inoculated in liquid cultures with appropriate antibiotic. Glycerol stocks

and colony PCRs were realized after 48 h using Taq DNA Polymerase (NEB). The PCR reactions were subsequently confirmed by Sanger sequencing.

2.7.3.2 Plasmid loss

The protocol described for *A. woodii* plasmid loss (section 2.7.2.3) was performed for *C. autoethanogenum* plasmid loss.

2.8 Transcript levels

2.8.1 RNA extraction

Acetobacterium woodii cells were collected by centrifugation at 7,000 g for 5 min at 4°C and the pellet was resuspended in 2 mL of RNAprotect (Qiagen). After a 5 min incubation at room temperature, the cells were collected again and the pellet stored at -80°C.

RNA extraction was carried out using a FastRNA Pro Blue kit (MP Bio). Cells were homogenized at 6,400 rpm for 45 sec, all the centrifugation steps were carried out for 15 min at 4°C, the RNA was precipitated for 2 h at -80°C, and the RNA pellet was dried for 30 min at room temperature and resuspended in 50 µL of DEPC-H_2O.

To reduce genomic DNA contamination, samples were digested by TURBO DNase (Ambion). The resulting RNA was purified using the RNeasy Mini kit (Qiagen). 350 µl of RLT buffer containing β-mercaptoethanol and 200 µL of absolute ethanol were added to the digested samples before being purified according to the manufacturer's handbook.

To check for genomic DNA contamination, a PCR (35 cycles, 50°C, 2:00 min extension) with primers gDNA_FW and gDNA_RV (See Supplementary Table 9), amplifying a 1,997 bp fragment in the *A. woodii pyrE* region, was carried out with Taq DNA Polymerase (NEB). When genomic DNA was present, samples were treated with

TURBO DNase and purified with the RNeasy Mini kit a second time. RNA samples were stored at - 80°C.

The RNA concentration was assessed with a Nanodrop Lite spectrophotometer (Thermo Fisher). The quality was evaluated using an Agilent RNA 6000 Nano kit on an Agilent 2100 Bioanalyzer, the RIN (RNA Integrity Number) was recorded for all the samples.

2.8.2 Quantitative reverse transcription real-time PCR (RT-qPCR)

Complementary DNA was synthesized from 1 µg of RNA using Omniscript RT kit (Qiagen), Random hexamer (Thermo Fisher), RNase inhibitor (NEB) following the manufacturer's instructions.

Power SYBR Green (Thermo Fisher) was used for the qPCR and primer pairs targeting the *gyrA* gyrase gene in the *A. woodii* genome, *orfB* of the plasmid backbones, the *catP* gene and the FAST gene were used, they are listed in Supplementary Table 9. Standard curves allowed for primer pair efficiencies to be calculated. Gene expression was calculate for each sample using the Pfaffl method without a calibrator gene:

$$Expression = \frac{E_{target}{}^{-Cp_{target}}}{E_{reference}{}^{-Cp_{reference}}}$$

2.9 Fluorescence measurements

2.9.1 Plate reader for creiLOV, Pp1, and Bs2 in *Escherichia coli*

Cells were grown in the appropriate medium then washed twice in PBS and resuspended in PBS. 200 µL of each sample was then loaded into a Corning 96-well black plate with clear bottom. Measurements were taken using a TECAN M1000 Pro (Tecan) plate reader. Measurements were taken with a 450 nm excitation and 495 nm emission in the bottom-reading mode. The optical density at 600 nm was also measured with the same

plates to normalize the measured fluorescence. Triplicates of PBS-only measurements were done to delete the background for each value.

2.9.2 Plate reader for Pp1 protein emission and excitation scans in *A. woodii*

Cells were grown in the appropriate medium, each sample was then loaded into a Corning 96-well black plate with clear bottom. Measurements were taken using a CLARIOstar (BMG Labtech) plate reader. Measurements were performed in duplicate for each sample and a blank with medium alone. The excitation scan was performed a 495 nm emission and excitation ranging from 400 nm to 464 nm with 1 nm increments. The emission scan was performed with a 450 nm excitation and emission measurement from 480 nm to 600 nm by increments of 1 nm.

2.9.3 Flow cytometry for FAST protein

A. woodii cells were grown until early- to mid-exponential phase and then pelleted at 7,000 g for 5 min. Cells were washed twice in PBS before being resuspended in PBS. 100 µL of 25 µM ^{TF}Amber (The Twinkle Factory) or PBS were added to 400 µL of resuspended cell samples immediately before fluorescence analysis. Cell fluorescence was analysed with an Astrios EQ cell sorter (Beckman Coulter) with a 488 nm laser and a 576/21 nm filter. Data was analysed using the Kaluza 2.1 analysis software (Beckmann Coulter). The population was gated to eliminate debris and agglomerates of cells, a first gate to select cells and a second gate to select single cells when represented on a Side Scatter / Forward Scatter plot.

For the preliminary experiments represented in Figure 5.10, the percentage of cells gated was represented as a function of the FAST Amber fluorescence intensity. The number of events for each graph is as follow; WT without / with ligand: 549,749 / 441,371;

Plasmid without / with ligand: 2,568,561 / 6,791,597; KI without / with ligand: 1,285,872 / 5,630,864.

For the library sorting, the number of single cells analysed and represented in the graphs of Figure 5.14 are indicated in Table 2.1 below.

Strain	**Ligand present**	**Number of single cells**
TcdR	No	1,010,937
pMTL8415_P*thl*_FAST	No	1,176,632
FAST KI	No	1,231,132
Library	No	1,003,791
TcdR	Yes	1,045,551
pMTL8415_P*thl*_FAST	Yes	1,124,879
FAST KI	Yes	1,178,844
Library sample 1	Yes	1,019,635
Library sample 2	Yes	1,011,950
Library sample 3	Yes	999,863
Library sample 4	Yes	1,015,593

Table 2-1: Number of single cells analysed for the different samples represented in Figure 5.14.

2.10 Inverse PCR

Genomic DNA is extracted using the DNeasy Blood and Tissue kit (Qiagen). 300 – 500 ng of the resulting gDNA is digested with restriction enzyme HindIII (NEB) for 2 h at 37ºC and the digestion inactivated for 20 min at 80ºC. A volume of 10 µL of the inactivated digestion is ligated by T4 DNA ligase in a total volume of 20 µL for 1 h at room temperature, the ligation is subsequently inactivated for 10 min at 65ºC. Finally, Q5 is used to amplify the DNA with primers catP_INV_F1 bis / catP_INV_R1 bis with an annealing temperature of 65ºC and 6 min extension.

2.11 Heterotrophic batch fermentation in serum bottles

The strains to be studied were grown until late stationary phase in 10 mL of AWM with 20 mM of fructose. The cultures were subsequently used to inoculate serum bottles containing 20 mL AWM with 60 mM fructose to OD_{600nm} between 0.02 and 0.03.

Samples were taken at regular intervals using a syringe. 1 mL of culture was collected, cells were pelleted at 9,000 g for 3 min and the supernatants and cell pellets were stored at -20ºC. Measurements of acetate, acetone, ethanol, and isopropanol concentrations were performed by HPLC (High-Performance Liquid Chromatography). For this purpose, the supernatants were thawed at room temperature, 300 µL of sample was mixed with 300 µL of diluent and filtered before analysis.

2.12 TraDIS library

2.12.1 Continuous fermentation in stirred tank bioreactor

As a control, the pooled library stocks were restored in fructose and sub-cultured twice in fructose in triplicates (experiment performed by Dr. Jonathan Baker).

Fermentations were performed in a BioFlo®/CelliGen® 115 (New Brunswick) bioreactor with a working volume of 1.5 L, pH was set at 7.0 and controlled by addition of 2 M sodium hydroxide. The vessel was initially autoclaved with 1.5 L medium and flushed with nitrogen for at least 24 h before inoculation. For continuous medium addition, 5 L of 5X buffered fermentation medium (Supplementary Table 3) and 15 L of dH_2O were pumped and filtered in a 20 L Duran bottle, the bottle was stirred and flushed with nitrogen for the whole duration of the fermentation.

All the cultures, including the fermentations, were performed in presence of thiamphenicol selection to ensure maintenance of *A. woodii* containing transposon

insertions. The pooled freezer stocks were used to inoculate a 100 mL fructose liquid cultures. After 72 h of growth, the cultures were used to inoculate the bioreactors. An initial batch growth was performed for 20-23 h, in 20 mM fructose for the fructose fermenter and 2 mM fructose for the H_2/CO_2 fermenter, before starting continuous mode. The fructose fermenter was grown on 20 mM fructose with 60 mL/min nitrogen to maintain the anaerobic conditions. For the gas fermenter, a N_2:H_2:CO_2 (100:70:30 v/v/v) gas mix was used with 70 mL/min of H_2, 30 mL/min of CO_2, and 100 mL/min of N_2.

Samples for genomic DNA extraction of the precultures were collected to establish the initial library mutant pool. Subsequently, samples for OD_{600nm}, potential genomic DNA extraction, and HPLC analysis were collected regularly during the fermentations. Cell pellets were collected by centrifugation at 7,000 g for 10 min, both the cell pellets and the supernatants were stored at -20ºC.

2.12.2 Library preparation for sequencing

Genomic DNA was extracted using a DNeasy Blood & Tissue kit (Qiagen) as described in section 2.5. 1 µg of each genomic DNA samples was fragmented via sonication by Deep Seq (University of Nottingham) using a S-series SonoLab (Covaris) with the following settings: 6W, 200 cycles/burst, intensity 3, 5% duty cycle. A 1X bead purification was performed using NEBNext® sample purification beads (NEB) before fragments were end repaired and dA-tailed using End Prep Enzyme Mix and Reaction Buffer (NEBNext® Ultra™ II DNA Library Prep kit, NEB). A custom splinkerette adapter (Barquist et al., 2016) was created by annealing primers SplA5_top and SplA5_bottom (Supplementary Table 10) at 95ºC and cooling at room temperature. The splinkerette at was ligated to the ends of fragments using the Ligation Master Mix and Ligation Enhancer reagents Buffer (NEBNext® Ultra™ II DNA Library Prep kit, NEB).

Adapter-ligated samples were purified again with a 1X bead purification step before a PCR was performed with Kapa HiFi HotStart ReadyMix (Roche). The PCR amplified the transposon-genomic DNA junctions with the library primer "5′ Tn1" and the adapter primer SplAP5.x (Supplementary Table 10) where x represents the index number, different for each sample prepared. SplAP5 primers bind to the adapter and contain a unique region allowing indexing of samples. The library primer contains the P5 flow-cell binding region, while the adapter primer contains the P7 flow-cell binding region. PCRs fragments were size-selected to be 350 bp to 650 bp on a 2% low-melt agarose gel and gel extracted using a Monarch® DNA gel extraction kit (NEB), the gel pieces were melted at room temperature. Gel extracted samples were analysed on a 2100 Bioanalyzer (Agilent) with the High Sensitivity DNA kit (Agilent) to assess the size distribution of the library. The concentration of DNA going onto the sequencing cartridge was assessed using a Qubit 4 fluorometer (Thermo Fisher Scientific).

Samples from heterotrophic growth in static bottles were prepared by Dr. Craig Woods.

2.12.3 Sequencing

Quality control by qPCR and subsequent Illumina sequencing was performed by Deep Seq (University of Nottingham).

Samples were run on a MiSeq using a custom sequencing primer "5′ SeqLong" (Supplementary Table 10) which was designed to read out of the transposon giving initially 12 bp of the transposon tag and thereafter genomic DNA. A custom MiSeq recipe was also used which employed dark cycles over the first 12 bp of reads hence preventing run failure from an inability to accurately detect clusters. The dark cycles results in accurate calling of cluster location since after 12 bp the bases are not homogenous throughout the whole sequencing lane.

2.12.4 Bioinformatics

Reads were filtered for quality and trimmed using TrimGalore with the default parameters (https://github.com/FelixKrueger/TrimGalore). The MiSeq custom recipe already filters for reads that contain the specified transposon tag. Bioinformatic analyses were then performed using methods based on those outlined in Barquist et al., 2016. Quality control was performed using the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reads were processed and mapped against *A. woodii* DMS1030 genome (CP002987.1) using the bacteria_tradis script with the smalt mapper, kmers of 10 bp, and a step size of 1 bp. Reads mapping to multiple locations in the genome were mapped randomly to one of those locations.

Extraction of reads abundance at the gene level and at the single mutant level, and gene essentiality, were performed by Dr. Craig Woods. Subsequently, the $\log_2$ fold change of abundance was calculated between the two subcultures in fructose bottles and between the last and first time point of each fermenter as $\log_2(Y/X)$ with Y the end abundance and X the start abundance.

CHAPTER 3: CRISPR/Cas systems for genome engineering of *A. woodii*

3.1 Introduction

3.1.1 Genetic engineering of *Acetobacterium woodii*

3.1.1.1 Plasmid-based engineering

Plasmid-based expression of endogenous or exogenous genes often constitutes the first experiment performed in organisms where DNA transfer is possible. It does not require genome engineering capabilities and is thus the first approach to study metabolism, improve yield, or express an exogenous pathway in a new bacterium of interest.

From the earliest *A. woodii* transformation reported, Gram-positive replicons for the efficient maintenance of plasmids are described. In 1994 a plasmid containing pAMβ1 was successfully electroporated (Stratz et al., 1994). More recent articles detail the electroporation protocol and describe a range of replicons functional in *A. woodii* including pIP404 (Straub et al., 2014), pBP1, pCB102, and pCD6 (Baker et al., 2022; Hoffmeister et al., 2016).

In *A. woodii*, the WLP was studied by plasmid-based overexpression of the genes encoding its enzymes from *C. ljungdahlii* to identify potential bottlenecks for acetate production (Straub et al., 2014). Two plasmids were constructed with either the phosphotranacetylase and acetate kinase genes or formyl-THF-synthetase, methenyl-THF-cyclohydrolase, methylene-THF-dehydrogenase, and methylene-THF-reductase. The recombinant strains showed that gene expression is a bottleneck for high carbon flux during autotrophic growth and both strategies resulted in increased carbon flux through the WLP.

In 2016, production of acetone in *A. woodii* was achieved with the promising yield of up to 26.4 mg/L/h in continuous gas fermentation (Hoffmeister et al., 2016). This

production was achieved using a synthetic Acetone Production Operon (APO) assembled with genes from *Clostridium acetobutylicum*.

While those two studies demonstrate how plasmid-based metabolic changes can generate important insights on the strains studied, they also highlight key limitations. Plasmids are present in multiple copies in each cell, resulting in a high copy number of the genes they contain. This can lead to a beneficial overexpression of exogenous genes, but it can also result in unnecessary burden to the cells. Overexpression of the WLP genes led to a lower biomass and acetate titer when grown heterotrophically, potentially explained by a higher burden to the cells (Straub et al., 2014). Hoffmeister and co-workers also showed that replicons had an impact on acetone production, replicons pBP1 or pCD6 resulted in the same concentration of acetone as pIP404 whereas pCB102 resulted in a lower acetone concentration (Hoffmeister et al., 2016). Replicons have an impact on plasmid copy number and stability, the results presented by Hoffmeister and coworkers for acetone production confirm that burden to the cells is a determining factor of plasmid-based expression of genes.

Additionally, in industrial processes, antibiotic supplementation comes with added cost and associated risks, as maintenance of the plasmid requires the use of antibiotic selection.

3.1.1.2 Allelic Exchange

To overcome the shortcomings of plasmid-based expression and for the development of commercial strains, genomic integration is crucial. A commonly used technique for genome engineering is allelic exchange (AE). Homology regions are designed to enable deletion of a genomic region or integration of a cargo sequence and placed on a plasmid lacking Gram-positive replicon. Upon transformation, homologous recombination

results in a first crossover and integration of the plasmid in the genome. This event is selected for based the acquisition of a resistance gene. To select for the second crossover and excision of the plasmid, a counter-selection marker is required.

To overcome the limited availability of counter-selection markers in *Clostridium*, a specific AE method, the Allele Coupled Exchange (ACE) method, is utilised to delete the *pyrE* gene. This method employs asymmetrical homology regions to control the order in which homologous recombination occurs. The first crossover is selected with thiamphenicol resistance gene (*catP*), and the second crossover (having a 4-fold shorter homology region than the first one) is selected because of its disruption of the *pyrE* and resulting 5-FOA resistance (Heap et al. 2012). With this technique, *pyrE* can subsequently be utilised as a counter-selection method for genome engineering of other loci. This technique has now been used in *C. beijerinckii* (Little et al. 2018), *C. difficile* (Ng et al. 2013)*, C. acetobutylicum* and *C. sporogenes* (Minton et al. 2016). The genome of *A. woodii* has been successfully engineered with the ACE method, yielding a diversity of gene knock-outs (Baker et al., 2022).

This method requires high efficiency transformations required with the use of replication-deficient vectors and adequate selection markers for selection of the successive homologous recombination events. While those requirements were successfully met in *A. woodii*, the process of restreaking and extensive strain screening to obtain the desired strain is lengthy and may generate unwanted mutations in the genome (Baker et al., 2022).

3.1.2 Barriers to the utilization of *Streptococcus pyogenes* Type II CRISPR/Cas9

The Type II CRISPR/Cas9 system from *Streptococcus pyogenes* (Nishimasu et al., 2014) has played a leading role in recent advances in genetic engineering. The expression of the Cas9 endonuclease from *S. pyogenes* along with a single guide RNA (sgRNA), a trans-activating RNA fused with a guide RNA, has been shown to effectively cleave both strands of DNA in the sequence complementary to the guide RNA. Similar to AE, it relies on homologous recombination directed by homology arms and cleavage of the genome by the *cas9* endonuclease to counter select against the WT genomic locus. It has simplified the rapid isolation of markerless knockout (KO), knock-in (KI) and point mutations, primarily in eukaryotic cells but also in multiple bacterial species.

There are reports of successful application of CRISPR/Cas9 in several *Clostridium* spp. Successful knock-outs were achieved in *C. beijerinckii* (Wang et al., 2015; Wang et al., 2016) and *C. ljungdahlii* (Huang et al., 2016), with *cas9* under the control of constitutive promoters, either native or the thiolase promoter from *C. acetobutylicum*, respectively. However, with a widening scope for its application in prokaryotes, the system suffered from a number of drawbacks mainly linked to the toxicity of the Cas9 nuclease expression.

3.1.2.1 Toxicity of the Cas9 nuclease

The *S. pyogenes* Cas9 has been shown to be highly toxic to cells due to its nucleolytic activity causing unspecific double-strand breaks in genomes (Jinek et al., 2012; Q. Li et al., 2016). In 2014, the crystal structure of Cas9 in the presence of a guide RNA and target sequence was elucidated (Nishimasu et al., 2014). That study allowed for a better

understanding of the Cas9 interactions with the different CRISPR/Cas9 elements. Two main domains were identified; the REC lobe and the NUC lobe for recognition and cleavage of the target sequence, respectively (Nishimasu et al., 2014). The nuclease lobe contains the subunits of a RuvC nuclease and a HNH nuclease, they have been shown to cleave the non-complementary and complementary strand of target DNA, respectively (Figure 3.1). One critical catalytic residue has been identified in each nuclease which, when mutated, completely abolished nucleolytic activity. Mutation of one of the nucleases (D10A) transforms Cas9 into a nickase (nCas9) and mutation in both nucleases inactivate it completely. The latter is called 'dead Cas9' (dCas9; D10A + H840A) as it has no nucleolytic activity *in vitro* (Jinek et al., 2012) but is still able to bind DNA.

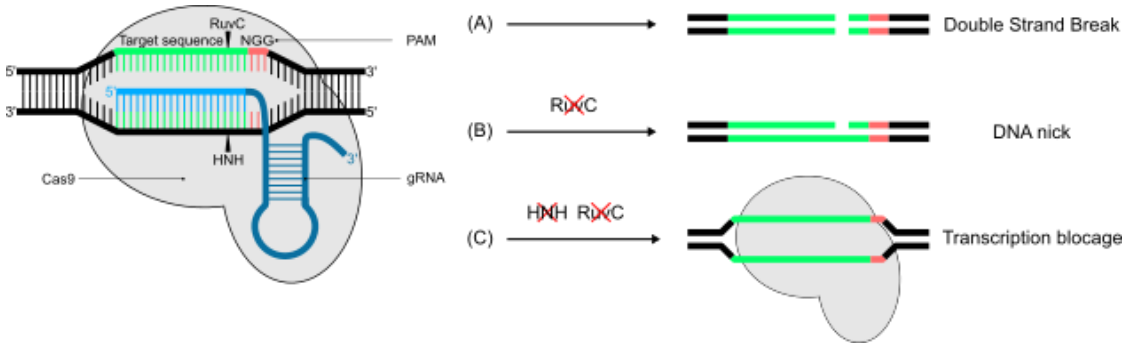


Figure 3-1: Schematic representation of the native Cas9 bound to target genomic DNA and the activity of its different variants; (A) action of native Cas9; **(B)** action of the nickase, Cas9 variant D10A with RuvC domain inactive, nCas9; **(C)** action of the dead Cas9, Cas9 variant D10A H840A with RuvC and HNH domains inactive, dCas9.

To overcome Cas9 toxicity for genome engineering applications in *Clostridium* spp., recent developments include the modulation of the Cas9 nuclease activity by regulation of its expression (Canadas et al., 2019) or mutation of its sequence, yielding nickases (Ingle et al., 2019; Q. Li et al., 2016; Xu et al., 2015) or dead Cas9 (Q. Li et al., 2016; Woolston et al., 2018).

3.1.2.1.1 DNA Nickase

In *C. acetobutylicum*, the presence of Cas9 reduced the transformation efficiency almost 5-fold compared to the empty vector backbone (Q. Li et al., 2016). The replacement of Cas9 by a nickase (D10A) restored the transformation efficiency to levels comparable to the transformation efficiency of the empty vector backbone, confirming that single strand nicks (SSN) are significantly less detrimental to cells than double strand breaks (DSB). The expression of that nickase gene with strong constitutive promoters allowed the successful editing of various genomic targets in *C. acetobutylicum* and *C. beijerinckii* (Q. Li et al., 2016). Cas9 toxicity was relieved in the same manner in *C. cellulolyticum*, allowing for generation of knock-out and knock-in mutants in the strain (Xu et al., 2015).

In a more recent study, application of the CRISPR/Cas9 system in *C. difficile* allowed for successful generation of knock-out strains (Ingle et al., 2019). The necessary editing vectors were constructed from the plasmids that allowed for genome editing in *C. ljungdahlii* (Huang et al., 2016), with *cas9* under the control of the constitutive thiolase promoter. However, sequencing of those editing vectors revealed a frameshift in *cas9* present in the *C. difficile* vectors as well as their parental plasmids (Ingle et al., 2019). The observed frameshift resulted in the truncation of the RuvC domain and conversion of the Cas9 into a nickase. As a result, genome engineering experiments successfully carried out in *C. ljungdahlii* (Huang et al., 2016) were actually obtained with a nickase rather than the fully functional Cas9. This highlights the toxicity of the unaltered Cas9 and the need for the modulation of its activity for applications in *Clostridium* spp.

3.1.2.1.2 Inactive Cas9

Contrary to nCas9, dCas9 has no nucleolytic activity, while it cannot be used for direct genome engineering, it can still provide important insight in cell metabolism. The dCas9 (D10A + H840A) ability to bind DNA without introducing cuts was employed to establish an efficient method to knock-down gene expression. The bound Cas9 blocks RNA polymerase, halting the transcription elongation, resulting in efficient silencing of transcription of the target gene (Larson et al., 2013), this technique is called CRISPR interference (CRISPRi). It was successfully applied to *C. acetobutylicum* and *C. beijerinckii* to repress the expression of the sporulation gene *spo0A* with a significant reduction of relative gene expression (Q. Li et al., 2016). In *C. ljungdahlii,* CRISPRi successfully enabled transcriptional regulation of genes (Woolston et al., 2018). To optimize the system further, Cas9 was put under the control of an inducible promoter to have better control of gene regulation for improved phenotypic characterisation. Multiplexing was also achieved to regulate expression levels of multiple genes in one step. Overall, CRISPRi proved useful for elucidation of metabolic pathways and redirection of the carbon flux towards 3-hydroxybutyrate in a recombinant strain (Woolston et al., 2018).

3.1.2.1.3 Regulation of *cas9* expression

Aside from mutation of *cas9*, the tight control of its expression was explored to reduce toxicity. Riboswitches are mRNA regions that are able to bind different small molecules, they can play an important role in regulating naturally occurring events in cells. Binding triggers a conformational change in the mRNA that affects downstream gene expression, most commonly by making the RBS available to the ribosome (Serganov & Nudler, 2013). Under those circumstances, riboswitches can be selected or engineered to act as inducible expression systems for the expression of the protein of interest. In 2010 a

riboswitch was identified that binds to theophylline, an inexpensive small molecule. A range of five different riboswitches was engineered to have different activation strength and basal expression by modifying their RBS or the stability of their secondary structure when no ligand is present (Topp et al., 2010). These riboswitches were found to be functional in the Gram-positive bacteria *S. pyogenes*, *M. smegmatis,* and *B. subtilis* (Topp et al., 2010). Recently one of these riboswitches was further engineered for exploitation in *Clostridium*, it resulted in a range of riboswitches and promoter combinations exhibiting a range of expression levels (Canadas et al., 2019). The RiboCas system combined those riboswitches and the CRISPR/Cas9 technology to overcome the toxicity problem in multiple clostridia. The system consists of a plasmid containing a *cas9* under the control of a theophylline-induced riboswitch linked to *C. sporogenes* ferredoxin promoter (P_{fdx}) and a sgRNA expressed either under the control of the constitutive arabinose promoter from *Clostridium acetobutylicum* (P_{araE}) or a synthetic promoter J23119, sourced from the iGEM registry. The cells are allowed to recover before inducing the expression of Cas9 with theophylline (Canadas et al., 2019), allowing more time for homologous recombination and subsequently increasing efficiency of genome editing. Efficiency of the RiboCas tool was demonstrated in *C. pasteurianum*, *C. difficile*, *C sporogenes*, and *C. botulinum* (Canadas et al., 2019).

3.1.2.2 Restriction-Modification systems

The early discovery that phage transfer success was dependent on the donor strain in the 50s (Murray, 2000) led to further studies to decipher the mechanisms protecting bacterial cells from foreign DNA. Restriction-Modification (R-M) systems have since been pinned as the failsafe recognizing and degrading foreign DNA, as such they are largely recognized as the “primitive immune system of bacteria” (Vasu & Nagaraja, 2013). R-M systems are based on methylation patterns recognized as endogenous DNA.

Those systems are based on methyltransferases (MTases) and restriction endonucleases (REases), performing methylation and degradation of unmethylated DNA, respectively. A classification of these systems was started in 1973 and ratified in 2003 (Roberts et al., 2003).

There are three main types of R-M systems (Roberts et al., 2003). Type I systems are organised as subunits working as a protein complex; subunit S (Specificity) recognizes the DNA motif, subunit M (Methylation) methylates DNA and subunit R (Restriction) cleaves the DNA. M_2S_2 complexes have a methylation function and $R_2M_2S_2$ have both methylation, and restriction functions, the balance between the different subunits is thus crucial (Mruk & Kobayashi, 2014; Roer et al., 2015). With Type I, restriction usually occurs around 1,000 bp away from the motif. Type II systems are comprised of an endonuclease and a methylase that act as separate proteins, the DNA cut is initiated at a specific nucleotide in the recognition motif. Type III have the methylase and endonuclease in a single protein and cut close to the motif. Not all the R-M systems fit in those three main types, for example Type IV contains enzymes that cleave only methylated DNA (Roberts et al., 2003).

The above classification led to the creation of a database, the REBASE database lists all the potential R-M systems, with their enzymes and recognition sites for a specific organism. That database predicts R-M systems in new genomes added to GenBank based on sequence similarities with known R-M systems (Roberts et al., 2010). MTases are more prevalent in that database because REases have less sequence similarity and are thus harder to predict with bioinformatics tools (Murphy et al., 2013). However, the database openly has a “liberal approach” by listing all potential candidates until proven otherwise, i.e., the enzymes are proven to be RNA or protein methyltransferases (Roberts et al., 2010).

Multiple studies have focused on improving the motif and methyltransferase prediction. That was facilitated by Single Molecule Real-Time (SMRT) DNA sequencing (Flusberg et al., 2010) to detect methylation motifs in genomes. Consequently, in 2016, a study significantly improved the number of motifs identified in prokaryotes by systematically studying the SMRT sequencing data for 230 bacterial and archeal species (Blow et al., 2016).

For the use of *S. pyogenes* CRISPR/Cas9, the relatively large size (4.1 kb) of the nuclease gene, *cas9*, makes it likely to contain restriction endonuclease (REases) recognition sites that will be cleaved by the target organism. To overcome native R-M systems, plasmids must be first appropriately methylated before being transferred into the target microorganisms. *E. coli* strains have been created for *in vivo* methylation (Jensen et al., 2019; Straub et al., 2014) and some enzymes are commercially available for *in vitro* methylation.

The former have been used to overcome R-M systems and led to successful genome engineering with CRISPR/Cas9 in *C. pasteurianum* (Bruder et al., 2016; Canadas et al., 2019) and .*C. acetobutylicum* (Bruder et al., 2016; Q. Li et al., 2016). Consequently, successful transfer of *cas9*-containing plasmids in bacteria of interest relies on adequate characterisation of the R-M systems present in the strain and generation of an appropriate methylation strain.

3.1.3 Endogenous CRISPR/Cas systems

3.1.3.1 Distribution and description of endogenous CRISPR/Cas systems

The increasing knowledge on CRISPR systems, present in approximately 45% of sequenced bacterial genomes (Grissa et al., 2007b; Rousseau et al., 2009), led to the

creation and improvement of databases of predicted CRISPR loci and tools to allow users to predict CRISPR loci in newly sequenced genomes (Couvin et al., 2018; Grissa et al., 2007a, 2007b; Rousseau et al., 2009). The use of endogenous CRISPR/Cas systems for genome editing could be a promising new avenue for the easy engineering of numerous organisms.

In Type I systems, the CRISPR locus is composed of a Cas cluster containing all the effector genes and one or more CRISPR arrays. The arrays contain multiple spacers derived from previous invading elements and separated by identical direct repeats (DR)(Nishimasu & Nureki, 2017). In both Type I and II systems, adaptation refers to the integration of new spacers in the array; a Cas1-Cas2 complex allows for sequence-specific cleavage of invading elements and integration of the resulting spacer in the array. It was shown that new spacers are inserted at the beginning of the CRISPR array (Alkhnbashi et al., 2016), making the new spacers likely to have a higher expression. The leader sequence controlling the transcription of the CRISPR array is usually located between the cluster and the array (Moon et al., 2019; Zhang, Zong, et al., 2018). Unlike in Type II systems, after transcription of the CRISPR array, pre-crRNA is cleaved by Cas6 after each 3′ stem loop (Figure 3.2) and not by tracrRNA-mediated RNase III. Cas6 remains tightly bound to the mature crRNA and is the base of the Cascade complex assembly, formed by Cas6-Cas8-Cas7-Cas5 (Figure 3.2).

Similarly to Type II systems, foreign DNA recognition is based on the presence of a protospacer adjacent motif (PAM) in the invading element to distinguish it from endogenous DNA (Nishimasu & Nureki, 2017). In the presence of an invading element containing both a protospacer matching one of the spacers and an appropriate PAM, the Cascade-mature crRNA complex binds the target DNA. This process is called interference and was found to be directed by a seed sequence adjacent to the PAM rather

than the full spacer (Semenova et al., 2011; Wiedenheft et al., 2011), possibly enabling a faster screening of DNA entering the cells. During interference, Cas5 binds the 5′ handle, Cas8 recognizes the PAM sequence and Cas7 binds to the spacer and the corresponding protospacer. Upon successful binding to the target DNA, the Cascade complex recruits the Cas3 nickase, it unidirectionally reels the DNA and nicks both strands of DNA upstream of the protospacer (Mulepati & Bailey, 2013).

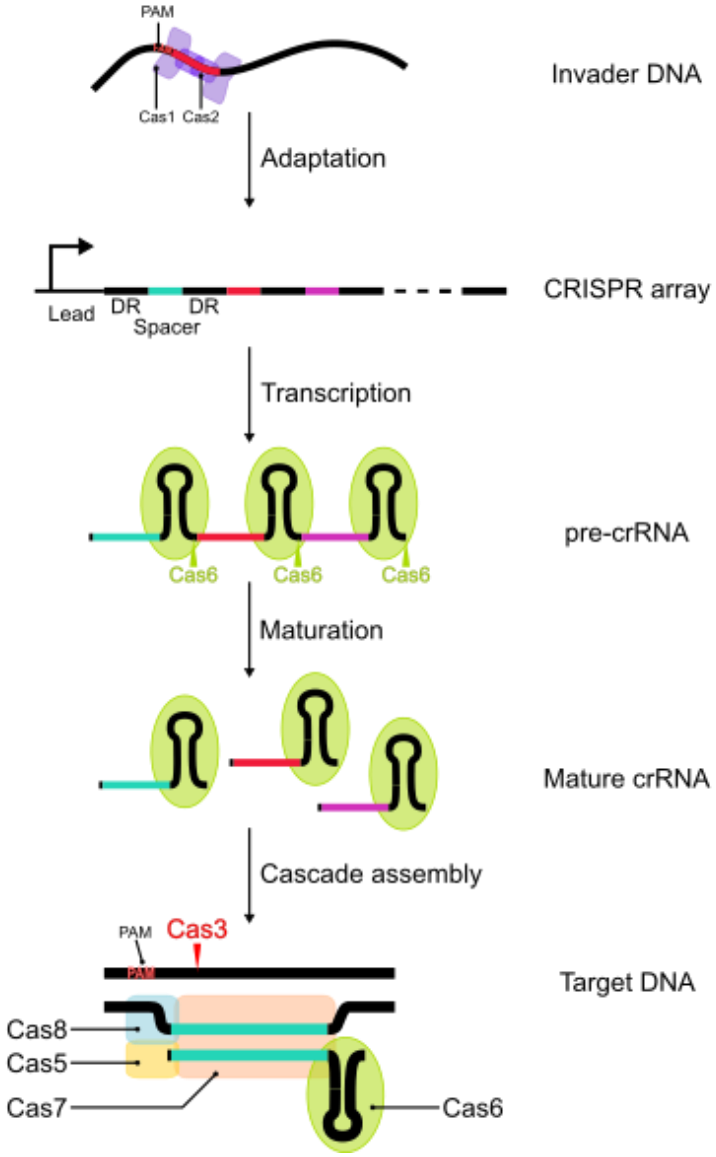


Figure 3-2: Steps of native CRIPSR interference for Type I systems. The Cas1-Cas2 complex detects new invader DNA, cuts and integrates the new spacer (red) in the CRISPR array. The CRISPR is built of direct repeats (DR) flanking the spacers, the lead sequence allows the transcription of the array. After transcription the pre-crRNA is processed by Cas6 that cuts it after each 3′ stem loop. The Cas6 remains tightly bound after cutting and is the base of the assembly of the cascade complex. Cas5 binds the 5′ handle, Cas8 recognizes the PAM sequence and Cas7 binds to the spacer and the corresponding protospacer. Upon binding the complex recruits the Cas3 for DNA nicking. Type I CRISPR interference results in a DNA nick. (Makarova & Koonin, 2015)

3.1.3.2 Harnessing of endogenous CRISPR/Cas systems for genome engineering

In *Clostridium* 74% of strains have an endogenous CRISPR system (Pyne et al., 2016), identified Type I-B systems have successfully been used to genetically modify a wide range of strains. Among them, *Clostridium pasteurianum* (Pyne et al., 2016), *Clostridium tyrobutyricum* (Zhang et al., 2018b), *Clostridium saccharoperbutylacetonicum* (Atmadjaja et al., 2019), *Clostridium butyricum* (Zhou et al., 2021), *C. difficile* (Maikova et al., 2019) and *Clostridium thermocellum* (Walker et al., 2020).

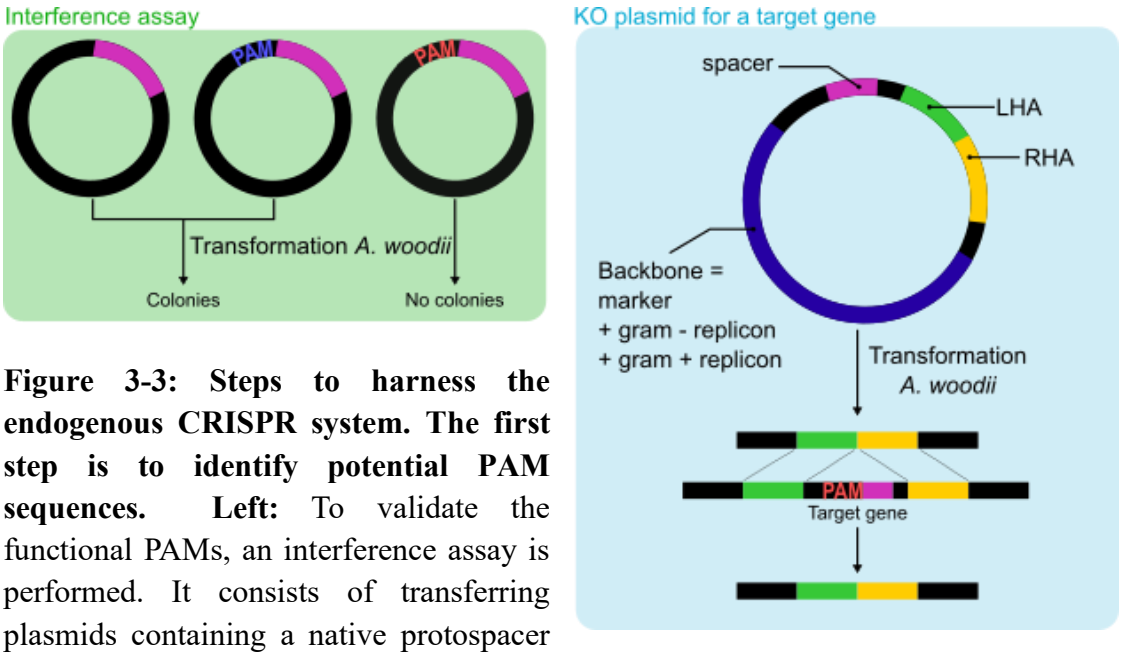


Figure 3-3: Steps to harness the endogenous CRISPR system. The first step is to identify potential PAM sequences. Left: To validate the functional PAMs, an interference assay is performed. It consists of transferring plasmids containing a native protospacer either without a PAM or with the different PAM candidates (blue and red). Functional PAMs should trigger interference and result in reduced transformation efficiencies (red one) compared to the protospacer alone or non-functional PAMs (no PAM and blue). **Right:** Genetic engineering experimental design. An appropriate spacer for the target gene (purple) is selected and expressed on a plasmid containing template DNA for homologous recombination (green and yellow). Only the cells with the deleted gene, and no protospacer-PAM sequence present survive.

3.1.3.2.1 Functional PAM identification

The CRISPR system is either unveiled by sequencing and analysing a newly sequenced genome or by querying the CRISPR-Cas++ database for a specified strain (Couvin et al., 2018; Grissa et al., 2007a, 2007b). The first step to harness a CRISPR system is to identify one or multiple PAM sequences able to trigger interference in the organism studied.

3.1.3.2.2 *In silico* strategies

The most utilised strategy to identify functional PAMs is *in silico*. Spacers matching phages, prophages and plasmids are identified using BLAST and PHASTER (Arndt et al., 2019; Zhou et al., 2011). To ensure that the spacer allows for interference, BLAST hits are filtered with mismatches accounting for less than 15% of the spacer (Zhang, Zong, et al., 2018) or a well-conserved seed sequence (Pyne et al., 2016; Zhang, Zong, et al., 2018), with no more than one mismatch in the first 7 nucleotides (Pyne et al., 2016). For Type I, it was shown that the PAM is upstream of the protospacer (Makarova & Koonin, 2015), so the 5′ sequence is retrieved to find the consensus sequence. A variant of that strategy is the use of CRISPRTarget, a tool created to predict the targets of CRISPR RNAs (Biswas et al., 2015; Biswas et al., 2013). That tool was created to better characterise the features of CRISPR arrays and the exposure history of the strains (Biswas et al., 2013). It has been used to better understand evolutionary relationships between related strains and the role of CRISPR systems (Negahdaripour et al., 2017), however, it has also been used recently to elucidate functional PAM sequences (Xu et al., 2020; Zheng et al., 2019).

In silico methods require subsequent *in vivo* validation of the PAM candidates. This is achieved with an interference assay (Figure 3.3). This consists of transforming into the microorganism of interest plasmids containing a spacer found in its genome flanked

either by the sequence native to the CRISPR array or one of the PAM candidate at its 5′ end (Boudry et al., 2015). Upon transformation, plasmids containing an active PAM should trigger interference by the endogenous CRISPR system and result in lower transformation efficiencies than plasmids containing part of the native DR, enabling confirmation of recognized PAM.

3.1.3.2.3 *In vivo* strategies

In vivo techniques are also employed to identify functional PAMs. Plasmid depletion assays rely on the construction of a plasmid library each containing a possible combination of five nucleotides and sequencing of the resulting transformation to find which combination triggered interference (Walker et al., 2020). Alternatively, PAM-SCANR relies on the cloning of the endogenous Cas genes in an *E. coli* strain. A NOT gate with the *lacI* gene and repressor is triggered (with subsequent expression of GFP) in the presence of a functional PAM upstream of the promoter and binding of the Cas proteins. The resulting cells are then sorted by FACS (Fluorescence-Activated Cell Sorting) to identify the functional PAM by sequencing (Leenay et al., 2016).

3.1.3.2.4 Need for improved PAM prediction

The CRISPRTarget tool was not optimized for the prediction of PAM. The manual *in silico* analysis to identify PAM candidates requires analysing a vast volume of BLAST hits and PHASTER submissions to obtain a relevant idea of the potential PAM candidates. It is thus labour intensive and does not yield a global landscape of possible PAMs. Both *in vivo* techniques require extensive library construction or cloning and rely on a large scale sequencing. For those reasons, the PAM identification step is recognized as one of the challenges for the use of endogenous CRISPR systems for genetic engineering (Zhou et al., 2021). A dedicated bioinformatics tool is needed to be able to

systematically predict potential PAMs for different bacteria (Y. Li et al., 2016), it would allow to analyse all the BLAST hits and have a global result for PAM candidates.

Despite the challenges of PAM prediction and verification, the consensus sequence was confirmed in different *Clostridium*, CCW and CCN for *C. difficile* (Boudry et al., 2015) and *C. saccharoperbutylacetonicum* (Atmadjaja et al., 2019) respectively, TCA for *C. pasteurianum*, *C. tetani*, *C. thermocellum* (Pyne et al., 2016) and *C. tyrobutyricum* (Zhang, Zong, et al., 2018), and TAA for *C. autoethanogenum* and *C. tetani* (Pyne et al., 2016).

3.1.3.3 Editing vector construction

On the editing vectors, only two elements are essential: a synthetic CRISPR array containing an adequate spacer flanked by DR, and an editing template to allow for homologous recombination (Figure 3.3). The spacer is designed to target the knock-out region and adjacent to a functional PAM.

Spacer expression, editing template length, Gram-positive replicon, and transformation parameters were shown to have an impact on editing efficiencies.

Optimal spacer size is easily determined by the average length of protospacers in the native array. In *C. butyricum*, protospacer are between 34 bp and 37 bp, it was shown that spacers of 34 bp and 37 bp allow genome engineering whereas spacers of 28 bp and 31 bp did not (Zhou et al., 2021).

Different strategies exist for expression of the synthetic array; however, the most straightforward solution is to use the native leader sequence (Hidalgo-Cantabrana et al., 2019; Pyne et al., 2016; Zheng et al., 2019). Other studies found that it led to unsuccessful genome editing and it was hypothesized that expression of the spacer was inadequate, either too high or too low. The former led to cell death, likely because spacer

expression happened before successful double crossover had occurred (Maikova et al., 2019; Zhou et al., 2021) and the latter led to insufficient DNA degradation (Walker et al., 2020; Xu et al., 2020). In those cases, successful genome editing was achieved by expression of the array with an inducible promoter or a strong constitutive promoter, respectively. In Pyne et al. (2016), the downstream sequence of the native array was added to the other end of the synthetic array, which might have had a role for adequate CRISPR array expression. Inducible promoters, constitutive promoters, and addition of the end sequence of the native arrays all constitute alternative strategies that can be tested depending on the results obtained in the strain of interest.

The editing template length is crucial for successful homologous recombination; longer HAs facilitate homologous recombination but increase plasmid size. HAs of around 500 bp (Xu et al., 2020; Zhou et al., 2021) and 1 kb (Hidalgo-Cantabrana et al., 2019; Pyne et al., 2016) each were used successfully in multiple organisms and for different applications including KO, KI and SNP introduction. In 2019, Zheng et al. (2019) successfully engineered *Zymomonas mobilis* with 700 bp HAs and then tested the impact of HA length by testing HA of 100 bp, 200 bp, 300 bp and 400 bp. Decreasing HA length significantly reduced the number of transformants obtained and the editing efficiency of the system. However, HAs of 300 bp constituted an effective compromise, with good editing efficiency and reduction of plasmid size.

Together, Gram-positive replicons and recovery times are also parameters that can be optimised to allow the strain to retain the plasmid during recovery, allowing for the homologous recombination to happen. Plasmid loss is required after genome editing, the Gram-positive replicon should therefore also be chosen to be easily lost for efficient downstream characterization.

3.1.4 Aims

The aim of this study was to establish CRISPR-based tools in the acetogen *A. woodii* to facilitate genome engineering.

Both the Cas9 toxicity and the presence of R-M recognition sites were investigated as barriers to the use of the exogenous Type II CRISPR/Cas9 system from *S. pyogenes*.

To facilitate exploitation of endogenous CRISPR systems, an *in silico* tool was developed for PAM prediction. Known CRISPR systems were used to validate the tool. As a proof-of-concept, the *A. woodii* CRISPR system was analysed to produce potential PAMs for genetic engineering. Other candidates relevant for industrial purposes were also analysed, to have wider application, as well as confirm and refine the script developed.

In *A. woodii*, an interference assay and RT-qPCR were performed to confirm the candidate PAMs and the functionality of the leader sequence, respectively. For genome engineering, the requisite knock-out vectors were built with two components, a synthetic array and an editing template. The synthetic array was built to mimic the native arrays, editing template length, Gram-positive replicon, and transformation parameters were investigated to increase editing efficiency.

Finally, to show the transferability of the techniques discussed, the CRISPR arrays of *C. autoethanogenum* was analysed with the Python script and genetic engineering using the endogenous CRISPR systems was undertaken.

3.2 Results

3.2.1 *S. pyogenes* CRISPR/Cas9 system in *A. woodii*

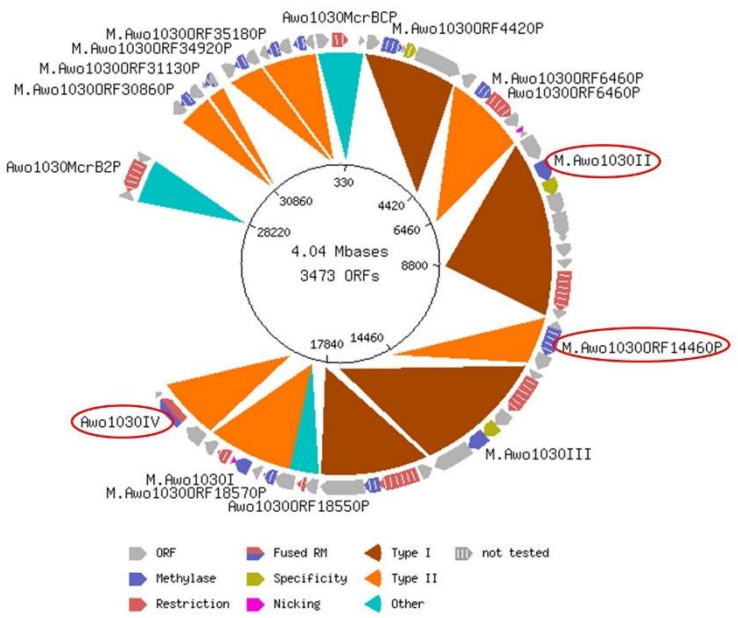
3.2.1.1 Restriction-Modification (R-M) system in *A. woodii*

3.2.1.1.1 *In silico* identification of R-M system specific to *cas9*-containing plasmid transformation

In 2014, Straub and co-workers first reported the necessity of the prior methylation of DNA before transformation into *A. woodii* (Straub et al., 2014). The plasmid used for methylation contained the gene for a Type I methyltransferase from *Clostridium ljungdahlii* (CLJU_c03310) identified previously (Leang et al., 2013). However, that gene has been further characterised (Philipps et al., 2019) and its recognition motif was found in most of the plasmids successfully transformed in *A. woodii* in our laboratory without prior methylation. Other *A. woodii* transformation protocols do not mention methylation requirements (Hoffmeister et al., 2016). This suggests that the methylation reported by Straub et al. (2014) was likely dependent on *E. coli* cloning strain choice and their different methylation phenotypes.

Data on the methylation motifs in *A. woodii* obtained in 2016 (Blow et al., 2016) were improved by linking those motifs to detected active methyltransferases (Jensen et al., 2019). Some motifs were slightly different from the predicted ones and they had diverse modification Quality Values (modQV, representing the probability of a modified base). Analysis of the REBASE data on the *A.woodii* genome (Figure 3.4A), can enable identification of R-M barriers to its transformation with exogenous DNA. In our laboratory, previous experiments have shown that *A. woodii* cannot be transformed with some plasmids, these plasmids were analysed for the presence of putative restriction motifs (Figure 3.4B).

(A)



(B)

Gene(s)	c08800-08810^	c17130-17140^	c18590^^	c19740^^	c14460^^
REBASE ID	II	III	I	IV	ORF14460
Motif	TAAGN5TCC	GATN6TGC	CCWGG	GCCRAG	CAAAAAR
modQV	1914	1567	48	1986	2185
Plasmid name	**Number of motifs on plasmid**				
pMTL8225_Csp_Pfdx_catP	0	0	11	3	6
pMTL8215_P1339_MCS	0	0	10	2	7
*pMTL82211_PthI	2	0	7	1	4
*pMTL8215_pRGCas1_MCS	2	0	10	6	21

Figure 3-4: Analysis of *A. woodii* R-M systems and their impact on plasmid transformation. (A) REBASE data available for *A. woodii* with the different R-M systems confirmed or hypothesized in its genome. The systems discussed below are circled in red. (Data available on the 17/04/2020 on the website: http://rebase.neb.com/cgi-bin/onumget?8566). **(B)** Analysis of *A. woodii* R-M systems. **Top:** List of active methyltransferases in *A. woodii* according to Jensen et al. (2019); genes ID designated in that study, ^ marks Type I R-M, ^^ marks Type II R-M; The REBASE ID corresponds to the ones used in Figure 1; the methylated nucleotides or their complements are underlined in the recognized motif. **Bottom:** Number of the previously listed motifs in different plasmids constructed and transformed in *A. woodii*; red if site present; green if site absent from plasmid; * identifies the plasmids that were not successfully transformed in *A. woodii*.

From this analysis, it is noteworthy that the site methylated by Straub et al. (2014) does not appear in REBASE. The impact observed on transformation efficiency in that study was likely due to an unexpected variation of cloning strains or other transformation methods.

All the plasmids were built in K-12 *E. coli* sub-strains (Dam^+ Dcm^+), the Dcm enzyme methylates the second cytosine of the motif 5′-CCWGG-3′ not the first, but that site has a low modQV in *A. woodii* (Figure 3.4B). A high number of this recognition site in plasmids does not seem to affect transformation; it indicates a low probability of this recognition site being a barrier for transformation in *A. woodii*.

The putative Type II methyltransferases recognizing motifs 5′-GCCRAG-3′, and 5′-CAAAAAR-3′ (Jensen et al., 2019) do not seem to have a detrimental effect on the transformation success of the studied plasmids. The latter does not appear to have a linked REase (M.Awo1030ORF14460P circled in Figure 3.4A), 48% of MTases are shown to be "orphan" MTases with no cognate REase (Blow et al., 2016). In most organisms, these enzymes play a role in DNA mismatch repair, replication initiation or gene regulation (Adhikari & Curtis, 2016; Blow et al., 2016; Murphy et al., 2013; Vasu & Nagaraja, 2013). In contrast, the former is a Type IIG system with fused MTase and REase domains (see Awo1030IV circled in Figure 3.4A). The methylation of the predicted motif for this system was experimentally verified but not its restriction activity (Jensen et al., 2019). This recognition site should, in theory, impede the transformation of plasmids. In practice, however, such plasmids have been successfully transformed in *A. woodii*.

Only a small sample size was analysed in Figure 3.4B, however, the results indicate that the methyltransferase recognition site linked to the genes Awo_c08800 and Awo_c08810 as a likely barrier to DNA transformation in *A. woodii* (R-M system M. AWO1030II circled in Figure 3.4). These two genes encode the methylase (M) subunit and the specificity (S) subunit of a Type I N^6-adenine DNA methyltransferase, recognizing 5′-TAAGN_5TCC-3′. The restriction subunit of this enzyme is found in gene Awo_c08870, also called *hsdR1*.

That site is present in specific modules of the studied plasmid: two sites are located in the low copy number Gram-negative replicon p15a, one in *cas9* and one in the riboswitch G.

The other Type I methyltransferase system, recognizing 5′-GATN$_6$TGC-3′ and linked to genes Awo_c17130 and Awo_c17140, might also hinder *A. woodii* transformation, but the recognition site was not found in any plasmid of interest in our laboratory.

3.2.1.1.2 *In vivo* validation of the R-M system identification

To confirm the impact of the presence of the 5′-TAAGN$_5$TCC-3′ recognition site on *A. woodii* transformation efficiency, *A. woodii* was transformed with plasmids with and without the target site in their MCS and their comparative transformation efficiencies were recorded (Figure 3.5).

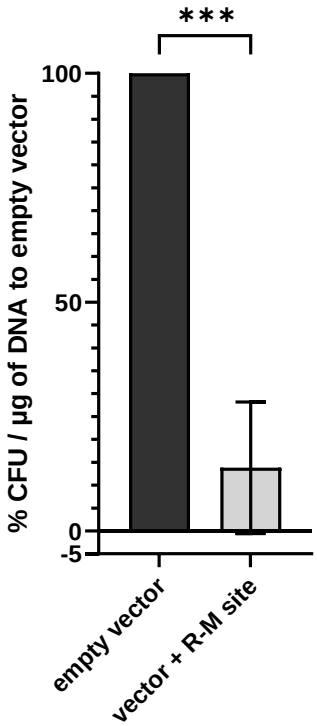


Figure 3-5: Transformation efficiency of WT *A. woodii* obtained with the vector containing the R-M target site (pMTL8215_targetIIP) expressed as a percentage of the efficiency obtained with the empty vector (pMTL82151) for each corresponding replicate. Two of the replicates resulted in 18.42 ± 0.12% transformation efficiency for the vector containing the R-M target site; the third replicate resulted in 0% transformation efficiency for this vector. n = 3, unpaired t-test, ***: -value < 0.001.

For two of the replicates, the transformation efficiency of the plasmid containing the R-M target site were 18.51% and 18.34% of the transformation efficiency of the control

plasmid lacking the target site. The third replicate did not produce any colonies after transformation of the plasmid containing the R-M site. Colony PCR amplicons for four colonies per condition for one replicate were confirmed by Sanger sequencing.

These results confirm that the R-M system encoded by genes Awo_c08800 and Awo_c08810 is a barrier to *A. woodii* transformation with DNA containing its recognition site.

3.2.1.2 Cas9 nuclease toxicity in *A. woodii*

Plasmids were built to assess the *cas9* toxicity isolated and in combination with the impact of the R-M site (5′-TAAGN$_5$TCC-3′) presence. The toxicity of the Cas9 protein is usually linked to its nuclease activity. Mutations D10A and H840A were previously shown to convert Cas9 into a dead Cas9 (dCas9) with no nucleolytic activity (Jinek et al., 2012). In this study, variants of the *cas9* gene were generated with those mutations and/or a silent SNP to remove the R-M recognition site. To remove the 5′-TAAGN$_5$TCC-3′ R-M recognition site, a silent mutation was introduced at position Y1001, creating the modified motif 5′-TAAGN$_5$CCC-3′. All the variants were placed under the control of the P$_{fdx}$ promoter associated with riboswitch E (Canadas et al., 2019). As a control, a plasmid lacking the *cas9* gene was also generated (Figure 3.6A). WT *A. woodii* was transformed with all the vectors constructed and the transformation efficiencies recorded (Figure 3.6B).

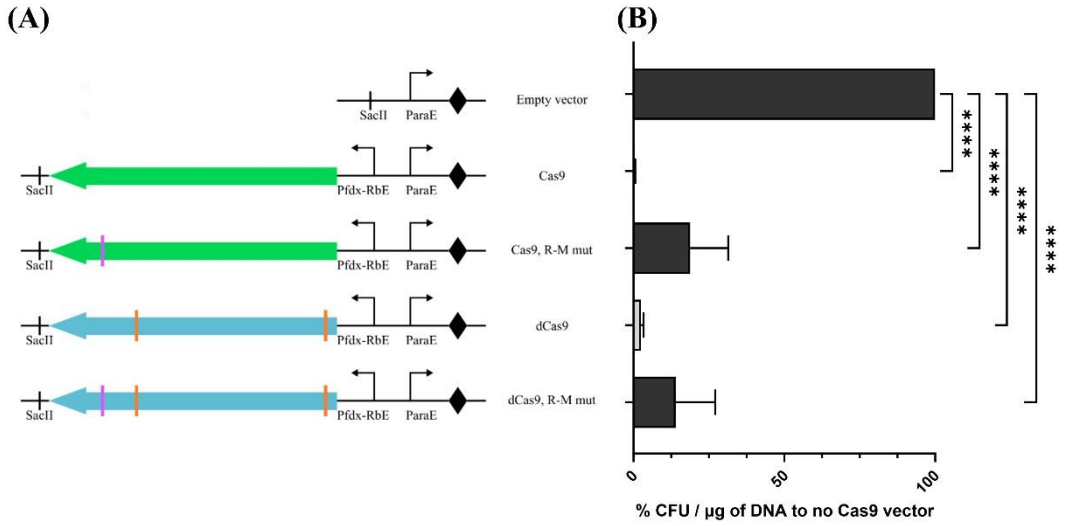


Figure 3-6: Assessment of the impact of the Cas9 toxicity and the presence of R-M target site on plasmid transformation in *A. woodii*. **(A)** Representation of the application-specific modules of the vectors containing the Cas9/dCas9 genes with or without mutation of the HsdR1 target site TAAGN$_5$TCC expressed. The empty vector contains the arabinose promoter (P_{araE}) controlling a MCS (diamond) from the pMTL8215_pRECas1_MCS vector. In subsequent vectors, Cas9 (green) and dCas9 (blue) are under the control of the $P_{fdx\text{-}RbE}$. Silent mutation Y1001 to remove the R-M site is represented in purple and mutations D10A and H840A to remove Cas9 nucleolytic activity are represented in orange. Empty vector: pMTL8215_P1339_MCS, Cas9: pMTL8215_pRECas1_MCS, Cas9, R-M mut: pMTL8215_pRECas1_IIPmut, dCas9: pMTL8215_pREdCas1_MCS, dCas9, R-M mut: pMTL8215_pREdCas1_IIPmut. **(B)** Transformation efficiency obtained with the described vectors as a percentage of the transformation efficiency obtained with the corresponding empty vector into wild type *A. woodii*. n = 3, error bars show standard deviations, ANOVA test, ∗∗∗∗: p <0.0001.

The presence of the unadulterated *cas9* gene under the control of the non-induced $P_{fdx\text{-}RbE}$, and in the absence of any gRNA, yielded the lowest average transformation efficiency; it represented a 99.58% decrease in transformation efficiency compared to the empty vector. The removal of the nucleolytic activity still exhibited a 97.46% decrease in transformation efficiency compared to the one obtained with the empty vector control.

The presence of the R-M recognition site was shown to hinder *A. woodii* transformation in the previous section but mutation of the R-M site in *cas9*-containing plasmids (including a dead cas9) was not sufficient to significantly increase the transformation efficiencies obtained. Two hypotheses could explain this, mutation of the R-M site in the *cas9* gene might not be sufficient to completely abolish recognition by the R-M system or Cas9 toxicity might be the prevalent factor preventing *A. woodii* transformation. Since both the Cas9 and dCas9 are inhibitory to transformation, the Cas9 toxicity could be a result of a combination of the nucleolytic or DNA-binding activities. Taken together, these results showed the inadequacy of CRISPR/Cas9 for *A. woodii* genome engineering.

3.2.2 Optimization of *in silico* steps for harnessing of endogenous CRISPR systems

3.2.2.1 Python script development

A Python script was developed to automate identification of candidate PAMs, facilitating the first step in harnessing endogenous CRISPR/Cas system. The script was designed to take as input the host organism genome identifier along with the information found on the CRISPRFinder database (Couvin et al., 2018; Grissa et al., 2007a) and compile all the information collected on each match for each spacer (See Supplementary Table 1 for a full list of attributes collected). A schematic overview is presented in Figure 3.7 and a complete diagram is available in Supplementary Figure 1.

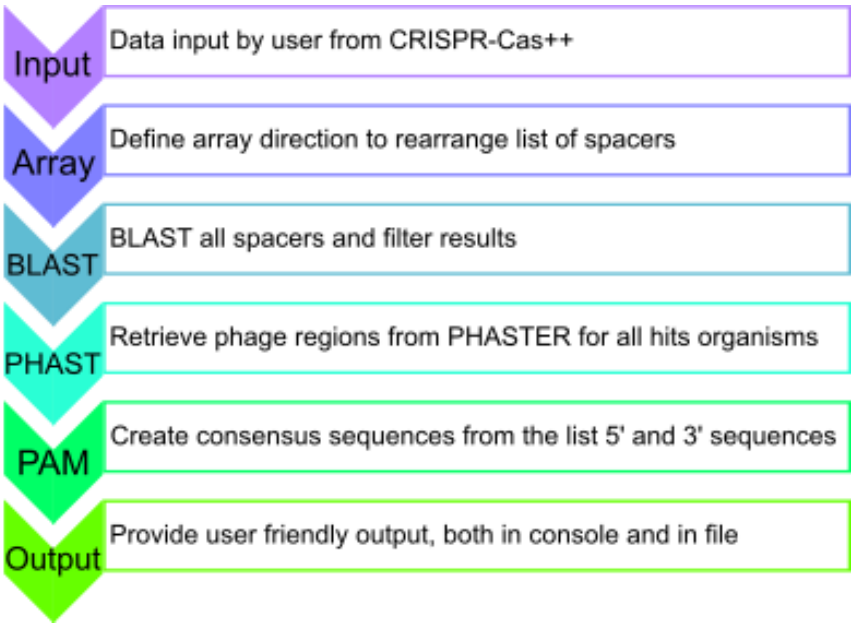


Figure 3-7: General overview from initial user input to final user-friendly results of the developed Python script.

After determining the array direction by promoter prediction, the first step is to run the spacers through BLAST to obtain all their matches. Filtering of those matches is then essential to remove false-positives as well as retain only the most biologically relevant results. Elimination of false-positives is based on identity: only hits with mismatches accounting for less than 20% of the spacer and a well-conserved seed sequence (Pyne et al., 2016; Zhang, Zong, et al., 2018), with no more than one mismatch in the 7 nucleotides adjacent to the PAM (Pyne et al., 2016) are kept. As CRISPR/Cas systems protect bacteria from invading elements, biologically relevant hits are retained if they belong to Mobile Genetic Elements (MGEs) sequences. For this purpose, three elements can be used to identify them, the title of the hit containing the keywords “phage” or “plasmid”, their presence in a phage sequence predicted by PHASTER, and the annotation of the hit containing the keywords “transposase”, “phage”, “integrase” and “terminase”. Adjacent sequences to the filtered hits are retrieved and output in an Excel sheet along with all the information obtained on that hit. A WebLogo is also compiled with all the candidate PAMs obtained.

3.2.2.2 Python script performance evaluation against published literature

The performance of the script was assessed using data available in the literature on Type I-B systems of *Clostridium* spp. that have been successfully used for genome editing (Table 3.1).

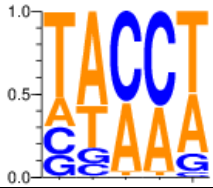
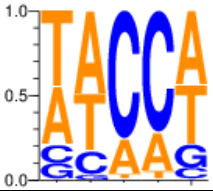
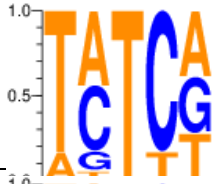
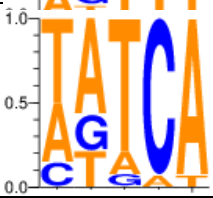
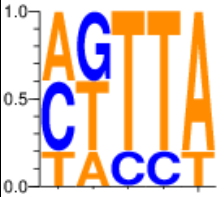
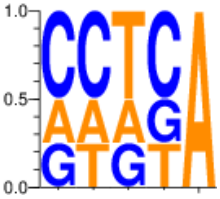
Strain	#	# in MGEs	Script WebLogo	Script consensus 5′ → 3′	Consensus (source) 5′ → 3′
C. difficile 630	105	234		CCW / AAK*	**CCW** (Boudry et al., 2015; Maikova et al., 2019)
C. difficile R20291	96	147		CCW	**CCW** (Boudry et al., 2015; Maikova et al., 2019)
C. pasteurianum ATCC 6013	45	18		TCW / TTG / TCG* / TTA*	**TCW / TTG** (Pyne et al., 2016)
C. tetani 12124569	31	14		TCA	TNA (Pyne et al., 2016)
C. thermocellum ATCC 27405	168	5		TTA / TTG / TCA	NCA (Pyne et al., 2016) **TTN / TNA** (Walker et al., 2020)
C. tyrobutyricum	25	4		TCA / GGA* / ATA*	**TCR** (Zhang, Zong, et al., 2018)

Table 3-1: Comparison of the 5′ PAM consensus sequences (5′ → 3′) found in the literature with the ones from the Python script for different strains of *Clostridium spp.*. Bold: PAMs verified experimentally; *: sequence that have not been experimentally tested. The total indicates the number of spacers analysed; Hits in phage, the number of spacers that were found in invading elements, all the consensus were done by WebLogo 3.7.4 (Crooks et al., 2004). (Ambiguous nucleotides codes: W for A or T; R for A or G; K for T or G; D for A, G or T)

The script successfully identified the majority of the functional PAMs. For *C. tyrobutyricum* it failed to identify the TCG alternative (Zhang, Zong, et al., 2018) and for *C. thermocellum* it identified only TTA/TTG/TCA instead of TTN and TNA which have been experimentally validated as functional (Walker et al., 2020). For a majority of these organisms the script predicted more PAMs not revealed by manual analysis in the literature, marked with asterisk in Table 3.1. *In vivo* tests are necessary to determine if these PAMs allow for interference and to conclude on the performance of the script.

3.2.2.3 Analyses of other *Clostridium* spp.

Clostridium spp. of interest containing a Type I-B system, *C. butyricum*, *Clostridium limosum* and *Clostridium novyi*, are of particular interest in our laboratory and were analysed using the script (Table 3.2).

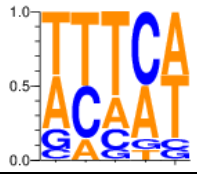
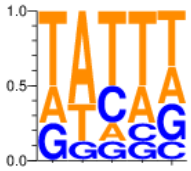
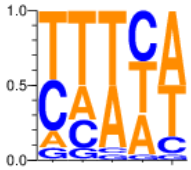
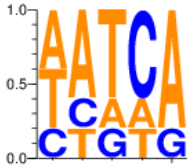
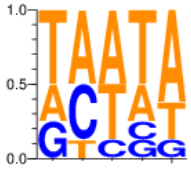
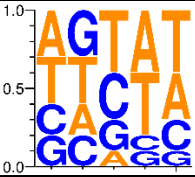
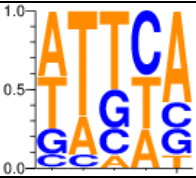
Strain	Array	#	# in MGEs	5′ WebLogo	5′ PAM 5′→ 3′
C. butyricum KNU-L09	3	57	14		TCM / TAW / GTT / CCG / TGT / CAT / ACT / AAA
C. limosum 14S0207	1	29	8		unclear
	2	35	26		TCT / ATA / TAA / ATT
	3	40	3		GTG / TCA / AAA
C. novyi NT	2	20	8		TTD / GYA / TCA / SAC / AAG / CCA
	3	26	10		HAT / TTM / GTW / CGC / TCA
	4	50	12		TTT / AAA / TGG / CCA / ATA

Table 3-2: Details of the analyses of the arrays corresponding to CRISPR/Cas Type I-B systems of a range of *Clostridium spp*. The array numbers correspond to the ones that can be found in the CRISPRFinder database. The number of spacers found in invading elements corresponds to the number found by the script to be in plasmid / phages / prophages / mobility-related genes. The 5′ PAMs indicated were compiled manually to account for separate consensus sequences possible for each organism. . (Ambiguous nucleotides codes: Y: C or T; S: G or C; M: A or C; W: A or T; D: not C; H: not G)

The results obtained indicate that PAMs containing the nucleotide pairs TC or TT are common in Type I-B CRISPR/Cas systems of the *Clostridium* genus and could potentially be tested in new organisms before any in-depth *in silico* analyses as previously suggested in the literature (Pyne et al., 2016). The number of candidate PAMs for each organism is quite high, confirming that those CRISPR/Cas systems are likely to recognise multiple PAMs (Fischer et al., 2012).

3.2.2.4 Analyses of bacteria harbouring CRISPR Type I systems

Finally, a wider range of bacteria harbouring different CRISPR Type I systems was analysed. Thermophiles and methanotrophs are studied in our laboratory, they are industrially relevant organisms but have limited genetic toolsets available, restricting their potential as biotechnological chassis. They are, therefore, interesting candidates for the use of their endogenous system as a gene-editing tool. *Moorella thermoacetica* has a Type I-B and a Type I-C system with 23 and 40 spacers respectively. *Methylocystis parvus* has a Type I-C system with two arrays of 53 and 87 spacers. *Methylococcus capsulatus* has a Type I-C system and a Type I-E system with one array each of 62 and 7 spacers respectively. Finally, *Methylomicrobium buryate* has a Type I-C system with one array of 19 spacers. The arrays of those four microorganisms were analysed with the script (Table 3.3).

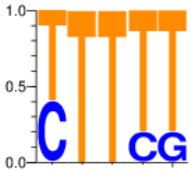
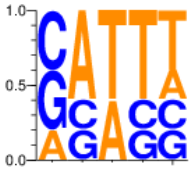
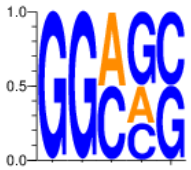
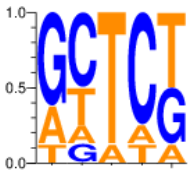
Strain	Type	Array	# of spacers	# in invading elements	5′ WebLogo	5′ PAM 5′ → 3′
Moorella thermoaceticа DSM 521	I-C	1	23	5		TTK / TCT
	I-B	3	40	0	N/D	N/D
Methylocystis parvus BRCS2	I-C	3	53	5		TTC / TGA / ATT / TCG
		4	87	4		CMG / AGC
Methylococcus capsulatus str. Bath	I-C	1	62	N/D	N/D	N/D
	I-E	2	7	0	N/D	N/D
Methylomicro-bium buryate 5GB1C	I-C	2	19	8		TCK / ATT / TAA

Table 3-3: Details of the analyses of the arrays corresponding to CRISPR/Cas Type I systems of a range of bacteria. The array numbers correspond to the ones that can be found in the CRISPRFinder database. The number of in spacer found in invading elements corresponds to the number found by the script to be in plasmid / phages / prophages / mobility-related genes. The 5′ PAMs indicated were compiled by hand to account for separate consensus sequences possible for each organism. (Ambiguous nucleotides codes: M: A or C; K: T or G)

The script failed to analyse the array of the Type I-C system of *Methylococcus capsulatus*, this error is likely due to the volume of queries submitted with the efetch function. It will be investigated in the future. None of the spacers in its Type I-E system

array were found to be in invading elements, this was also the case for the spacers of *Moorella thermoacetica* Type I-B system.

The script successfully identified candidate PAMs for the Type I-C systems of *M. thermoacetica*, *M. parvus*, and *M. buryate*. Nucleotide pairs TT and TC appeared to be essential for the three Type I-C analysed.

3.2.3 Harnessing *A. woodii* Type I-B CRISPR/Cas system

3.2.3.1 *In silico* analysis of *A. woodii* Type I-B CRISPR system

The CRISPRFinder database (Couvin et al., 2018; Grissa et al., 2007a) identified a Type I-B system in *A. woodii* (Figure 3.8A). The Cas locus is located between the positions 3,381,013 and 3,389,311 in reverse direction (Figure 3.8B). This system contains a Csx8, a Cas8a homolog, confirmed by protein BLAST, with homology to multiple Cas8a proteins from *Clostridium* spp. Cas8a is specific to Type I-A systems but the gene organisation of *A. woodii* Cas locus is characteristic to Type I-B systems (Makarova & Koonin, 2015).

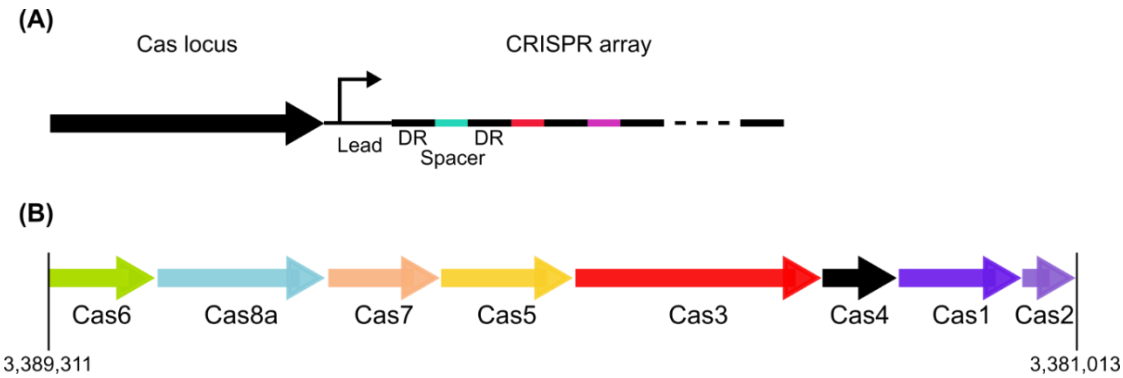


Figure 3-8: Organisation of the Type I-B CRISPR system of *Acetobacterium woodii* DSM1030. (A) Cas locus and CRISPR array relative positions. Both the Cas locus and CRISPR array are in reverse direction in the genome and the CRISPR array is directly downstream of the Cas locus. **(B)** Gene organisation of the Cas locus.

Only one CRISPR array, named array 2, with an evidence level of 4 is identified in the database. It is located directly downstream of the Cas locus; it contains 47 spacers and has a direct repeat "ATTTACATTCCAATATGGATCTACTCAAAT". The CRISPR array was analysed using the Python script, it identified multiple matches with invading elements but none with less than 17% mismatch (Supplementary Table 11). It computed no clear consensus sequence for the 5′ PAM, however, 3 out of the 8 matches contain two consecutive cytosines. Interestingly, a match with a 5′ CCA PAM in a phage protein was found and this PAM was shown to be functional for the Type I-B system of *C. difficile* (Boudry et al., 2015; Maikova et al., 2019). It is thus the main candidate for harnessing the endogenous CRISPR system in *A. woodii*.

3.2.3.2 PAM validation by interference assay

Interference assays allow to quantify the functionality of PAMs based on their ability to prevent strain transformation with plasmids that contain them. Transformation of a WT strain with a plasmid containing a protospacer and a functional PAM triggers interference and results in lower transformation efficiencies compared to non-functional PAMs. In *A. woodii*, three PAM candidates were experimentally tested including the CCA PAM candidate identified in the previous section. Mismatches adjacent to the PAM reduce the likelihood of effective interference, for this reason, the two other PAMs to test were chosen to have no mismatch in the first position and only one mismatch in the first ten nucleotides. Each spacer was assembled into a plasmid either with the native 5′ sequence "CAAAT" of the direct repeat or with its corresponding PAM candidate. All constructs contained the native 3′ sequence from the direct repeat "ATTTA", the protospacer sequences used are presented in Figure 3.9A.

(A)

5′ PAM	Protospacer
ATGCC	GAATCAAAATCATTAATAGCTTTAATTAAATCGTC
GGGTG	TTATTCTATGAGGTGATGAGATGAATGTTATTAAGA
AACCA	AATAACTATCAGGCTATCAGCTTCGGGGAGACTTT

(B)

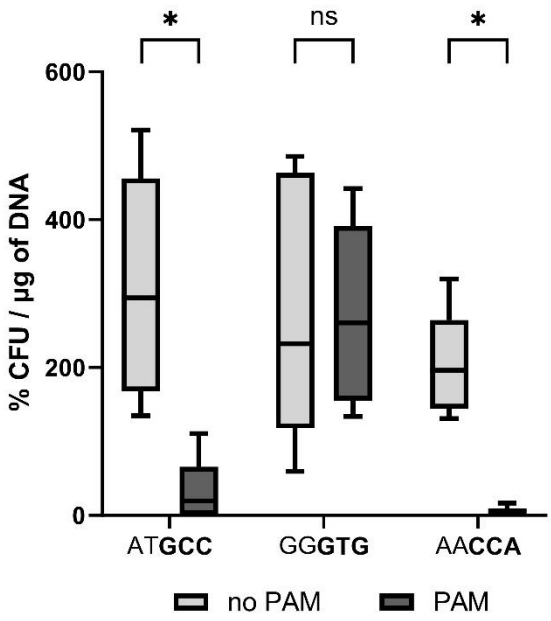


Figure 3-9: Interference assay to determine functional PAMs in *A. woodii*. (A) Table of the protospacers used in the interference assay in *A. woodii*. Three of the protospacers identified in Supplementary Table 11 were inserted in a backbone with, in 5′ position, either the native 5 nucleotides from the direct repeats "CAAAT" or candidate PAMs. The native 3′ sequence from the direct repeat "ATTTA" was used. Potential functional trinucleotides PAMs are underlined. **(B)** Transformation efficiency of the plasmids containing the different combinations of protospacer with (dark grey) or without (light grey) their corresponding candidate PAMs (bolded) into wild type *A. woodii*. Data normalized and expressed in percentages with the following parameters: transformation efficiency of the empty backbone control pMTL82151 set to 100% and smallest value of the replicate as 0%, the 5th and 95th percentile are represented. n = 5; two-way ANOVA test, ns: non-significant; ∗: $p<0.05$.

WT *A. woodii* was transformed independently with the six plasmids as well as the empty backbone pMTL82151 as a positive control for DNA transfer. The transformation efficiencies were recorded and expressed as a percentage of the transformation efficiency obtained with the control (Figure 3.9B).

Although high variations were observed between replicates, the reduction in transformation efficiency in the presence of the GCC and CCA PAMs was shown to be statistically significant (p-value = 0.038 and p-value = 0.01, respectively) compared to the presence of the protospacers alone. Colonies PCRs were performed on the colonies obtained and Sanger sequencing the amplicons showed prevalent mutations in the CCA PAM with one of the cytosines missing. It can be hypothesized that this mutation results from the selective pressure of Cas3 and that the cytosine deletion abolishes targeting of the plasmid by the endogenous CRISPR/Cas system. These colonies are thus escape mutants of the system and the cytosine deletion validates the hypothesis that the double cytosine is essential for efficient interference by *A. woodii* Type I-B CRISPR/Cas system. The CCA 5′ PAM was identified as the most likely candidate by *in silico* analysis (previous section) and it led to the highest level of interference, it was, therefore, selected for use in the following genetic engineering experiments in *A. woodii*.

3.2.3.3 Characterisation of the leader sequence

In endogenous CRISPR systems, the CRISPR array is transcribed into pre-crRNA before maturation and Cascade assembly. The assembled complex is then able to bind to target DNA and recruit the Cas3. In order to harness *A. woodii* CRISPR system, characterisation of the leader sequence responsible for the transcription of the array is necessary. For that purpose, the plasmid pMTL8415_leader_FAST was built with the FAST reporter gene placed downstream of the leader of array 2 described above. Control

plasmids were built with either the *catP* reporter gene replacing the FAST reporter gene downstream of the leader sequence (pMTL8415_leader_catP) or with the thiolase promoter replacing the leader sequence for expression of the FAST-encoding reporter gene (pMTL8415_Pthl_FAST). Characterisation was performed by RT-qPCR and these controls were used to confirm primer specificity and efficiency, respectively. FAST gene expression was normalized to plasmid copy numbers by using *orfB* as the reference gene (Figure 3.10).

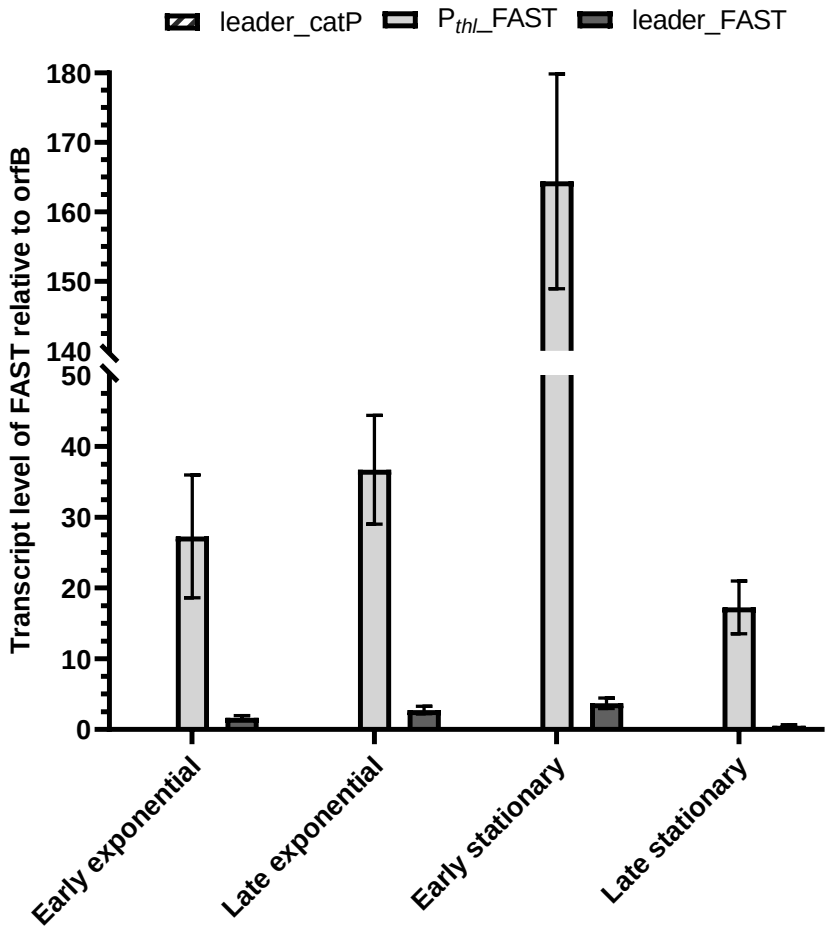


Figure 3-10: Endogenous CRISPR array leader sequence characterization: FAST transcript level normalized with *orfB* as the reference gene measured by RT-qPCR at different stages of cell growth. Samples were taken in early and late exponential stage and early and late stationary for the pMTL-MPD30 plasmid (**dark grey**), the negative control pMTL-MPD29 (**dashed**), and positive control pMTL8415_P*thl*_FAST (**light grey**). The transcript levels were calculated with the Pfaffl method without a calibrator sample. Error bars represent technical triplicates.

No transcript level was detected for the *catP*-containing construct, confirming specificity of the primers for the FAST gene. Transcript levels of the FAST gene expressed with the thiolase promoter confirmed adequate annealing of the primers. Interestingly, the leader sequence allowed for a low level of transcription in all cell growth stages with a clear reduction of RNA level in late stationary. This experiment confirmed the functionality of the leader sequence of array 2 for expression of CRISPR arrays in *A. woodii*.

3.2.3.4 Genome engineering capacity validation by *pyrE* knock-out generation

To establish the genetic engineering capabilities of the endogenous Type I-B CRISPR/Cas system of *A. woodii*, a deletion of 300 bp in *pyrE* that led to uracil auxotrophy through gene inactivation (Baker et al., 2022) was targeted. The editing cassette consisted of a synthetic array and two homology arms of 471 bp and 491 bp, respectively. The synthetic array contained a 36 bp spacer targeting the KO region 29 bp from the RHA and flanked by DR. It was expressed under the native leader of array 2 analysed in the previous section. Six plasmids were constructed (Figure 3.11A) with HAs alone, spacer alone or the whole editing cassette and either the pBP1 or pCD6 Gram-positive replicon. The vectors with HAs and spacer alone act as controls to assess the impact of homologous recombination and self-targeting by Cas3, respectively. Only the whole editing cassette with HAs and expressed spacer allows for homologous recombination and selection against WT by self-targeting. WT *A. woodii* was separately transformed with those plasmids in biological duplicate and plated after 5 h and 7.5 h of recovery. Colony numbers and transformation efficiencies were recorded (Figure 3.11B) and when possible, up to six colonies were analysed by colony PCRs to assess plasmid presence and functionality and the status of the genomic *pyrE* locus. The data obtained

by colony PCRs was analysed to determine the impact of the following parameters: presence of HAs, recovery times, and Gram-positive replicon. Colonies were clustered, in turn, according to each parameter regardless of the other parameters to assess their impact on the quality of the colonies (Figure 3.11C). Details of each colonies analysed is also provided as reference (Figure 3.11D).

The presence of the expressed spacer targeting the *pyrE* gene, with either replicon pCD6 or pBP1, with or without HAs, and with either 5 h or 7.5 h of recovery, significantly reduced transformation efficiencies to between 0.06% and 2.17% of the transformation efficiency of the corresponding plasmid with HAs alone (Figure 3.11B). This showed the efficient self-targeting by the endogenous CRISPR/Cas system of the *pyrE* gene leading to cell death. Analysis of the escape mutants revealed single crossover events and frequent deletion of the spacer; this can be explained by the presence of the identical direct repeats flanking the spacer facilitating recombination under Cas3 selective pressure.

As expected, the HAs are a requisite for successful KO generation (Figure 3.11C). Surprisingly, their presence did not significantly increase transformation efficiencies (Figure 3.11B), showing the low frequency of homologous recombination events. Colony numbers obtained varied significantly between the different replicates but successful gene knockout was obtained with the full editing cassette, with either Gram-positive replicon pBP1 or pCD6, and only with 7.5 h of recovery time.

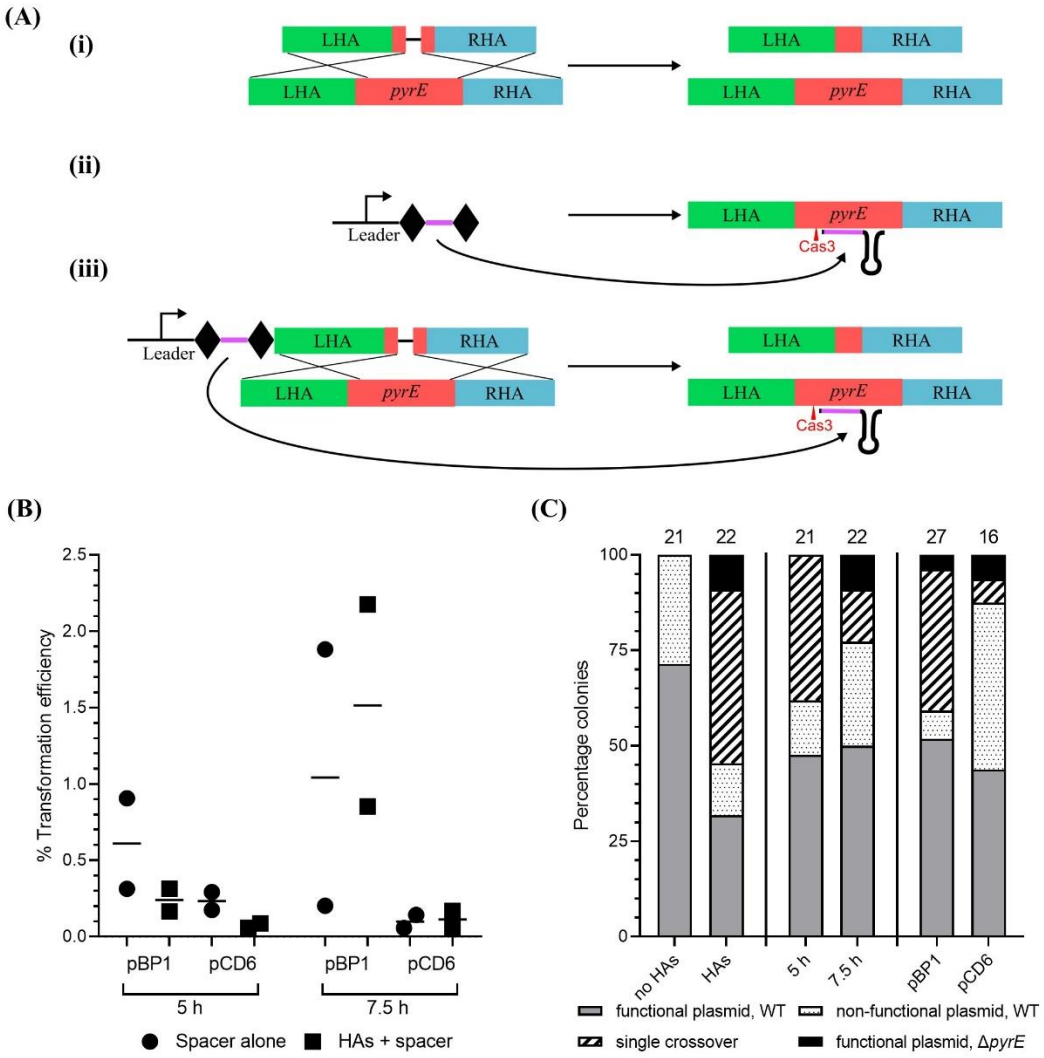

(A)
(i)
LHA
RHA
pyrE
(ii)
Leader
Cas3
(iii)
(B)
% Transformation efficiency
pBP1
pCD6
5 h
7.5 h
Spacer alone
HAs + spacer
(C)
Percentage colonies
no HAs
HAs
functional plasmid, WT
non-functional plasmid, WT
single crossover
functional plasmid, ΔpyrE
(D)
Colonies analysed
HA
spacer
KO
5h recovery
7.5h recovery
functional plasmid, WT
non-functional plasmid, WT
single crossover
functional plasmid, ΔpyrE

Figure 3-11: Endogenous CRISPR based deletion of the *A. woodii pyrE* gene. (A) *pyrE* (red) deletion strategy in *A. woodii* using the native Type I-B system. The constructions represented correspond to the application-specific module of the following plasmids, with "X" the Gram-positive replicon number: **(i)** pMTL8X15_HA pyrE, HAs alone (471 bp LHA in green, 491 bp RHA in blue) allow for homologous recombination in some of the cells of the population; **(ii)** pMTL8X15_leader_pyrE_spacer, the expression of the 36 bp spacer (purple) flanked by DRs (diamonds) alone leads to self-cleavage by Cas3 (red) at the *pyrE* gene and cell death; **(iii)** pMTL8X15_pyrE KO, the full editing cassette allows for homologous recombination, deletion of the 300 bp region and selection of the successful KO. **(B)** Transformation efficiencies obtained when transforming WT *A. woodii* with plasmids containing the spacer alone (**round**) or the spacer and HAs (**square**), expressed as a percentage of the transformation efficiency obtained with the plasmid containing the HAs alone. Transformation efficiencies with 5 h and 7.5 h recovery times and with both pBP1 and pCD6 Gram-positive replicons. The experiment was performed in duplicate and the line represents the mean. **(C)** Analysis of the data obtained by colony PCRs with primers AW_CRISPR_pyrE_armF and AW_pyrEcomp_RHAR after transformation of *A. woodii* with the plasmids containing the spacer targeting *pyrE*. When possible up to 6 colonies were analysed per condition. Colonies were clustered, in turn, according to the following parameters: presence of HAs, recovery times, and Gram-positive replicon, regardless of the other parameters. Total number of colonies is indicated above each column. **Grey**: wild type *pyrE* locus with functional corresponding plasmid; **Dots**: wild type *pyrE* locus with non-functional corresponding plasmid; **Dashed**: single crossover of the plasmid in the *pyrE* locus; **Black**: Δ*pyrE* with functional corresponding plasmid. **(D)** Details of colonies analysed for the 300 bp deletion in the *pyrE* gene.

For the plasmids containing the spacer, more colonies were obtained with the pBP1 replicon, the percentages of transformation efficiency obtained varied between 0.16% and 2.17% with pBP1, while they varied between 0.06% and 0.29% with the pCD6 replicon (Figure 3.11B). Gram-positive replicon also appeared to have an impact on the status of colonies obtained when transforming *A. woodii* with spacer-containing plasmids (Figure 3.11C), to decipher this impact, the following analysis was performed based on the details of all the colonies analysed (Figure 3.11D) regardless of recovery time. Although pCD6 resulted in more non-functional plasmids than pBP1 when only the spacer is present (55.55% against 0%), it led to more knock-out colonies (20%

against 5.88%) and less single crossover events (20% against 58.82%) when HAs are present as well.

Taken together, the relative colony numbers and their quality suggests that the pCD6 replicon leads to a plasmid copy number favourable for homologous recombination and genome targeting with the spacer it contains. The pCD6 replicon was also shown to be more unstable than pBP1, it is lost in 100% of the population after one subculture without selection (compared to 70% for pBP1) (Baker et al., 2022), which is a desirable parameter when designing editing vectors.

These results are the first reported successful gene editing using *A. woodii* endogenous Type I-B CRISPR/Cas system, they demonstrated that replicon pCD6 is a good option for future editing vectors and they constitute a very promising basis for further improvements of the technique.

3.2.3.5 Impact of homology arms lengths on editing efficiency for *pheA* knockout

A 354 bp in-frame deletion in the *pheA* gene that was shown to lead to phenylalanine auxotrophy through gene inactivation (Baker et al., 2022) was targeted. To assess the impact of homology arms lengths on editing efficiency, HAs of 0.5 kb, 1.0 kb, and 1.5 kb were tested (Figure 3.12A). The editing cassette was comprised of homology arms and a 35 bp spacer starting 26 bp from the RHA. The editing cassettes were inserted in a backbone with pCD6 replicon and native *A. woodii* leader sequence.

WT *A. woodii* was transformed with six plasmids, three containing the HAs alone and three editing plasmids containing the expressed spacer as well. The transformation efficiency obtained with the editing vectors were normalized to the transformation efficiency obtained with the vector containing only their respective HAs.

Low transformation efficiencies and high variations in colony numbers between each replicate were observed. The presence of the spacer reduced the transformation efficiency to averages (for triplicates) of between 0.053% and 0.32% for each condition (HAs length and recovery time) of the transformation efficiencies obtained with the corresponding vectors with the HAs alone. All but one individual value were comprised between 0% and 0.17%, with 0.57% obtained for one replicate with 1.5 kb HAs. This reduction of transformation efficiencies when the spacer is present demonstrated efficient targeting of the *A. woodii* genome by the designed spacer.

Characteristics of the colonies analysed for the three replicates of this experiment are summarized in Figure 3.12B, and Figure 3.12C shows the results of the colony PCRs for the *pheA* locus for the first replicate of this experiment.

Both at 5 h and 7.5 h recovery, the 1.5 kb HA resulted in more consistent knockout results with between 33.3% and 100% of gene editing efficiency, showing the importance of the editing template length on editing efficiency. Contrary to the previous section, this experiment showed that a 7.5 h recovery time is not essential for successful genome engineering.

After implementation of the plasmid loss protocol, 91.7% of colonies tested (72 in total) had successfully lost the plasmid. The phenotype of the strains was confirmed by plating on minimal media with and without phenylalanine supplementation (Figure 3.12D).

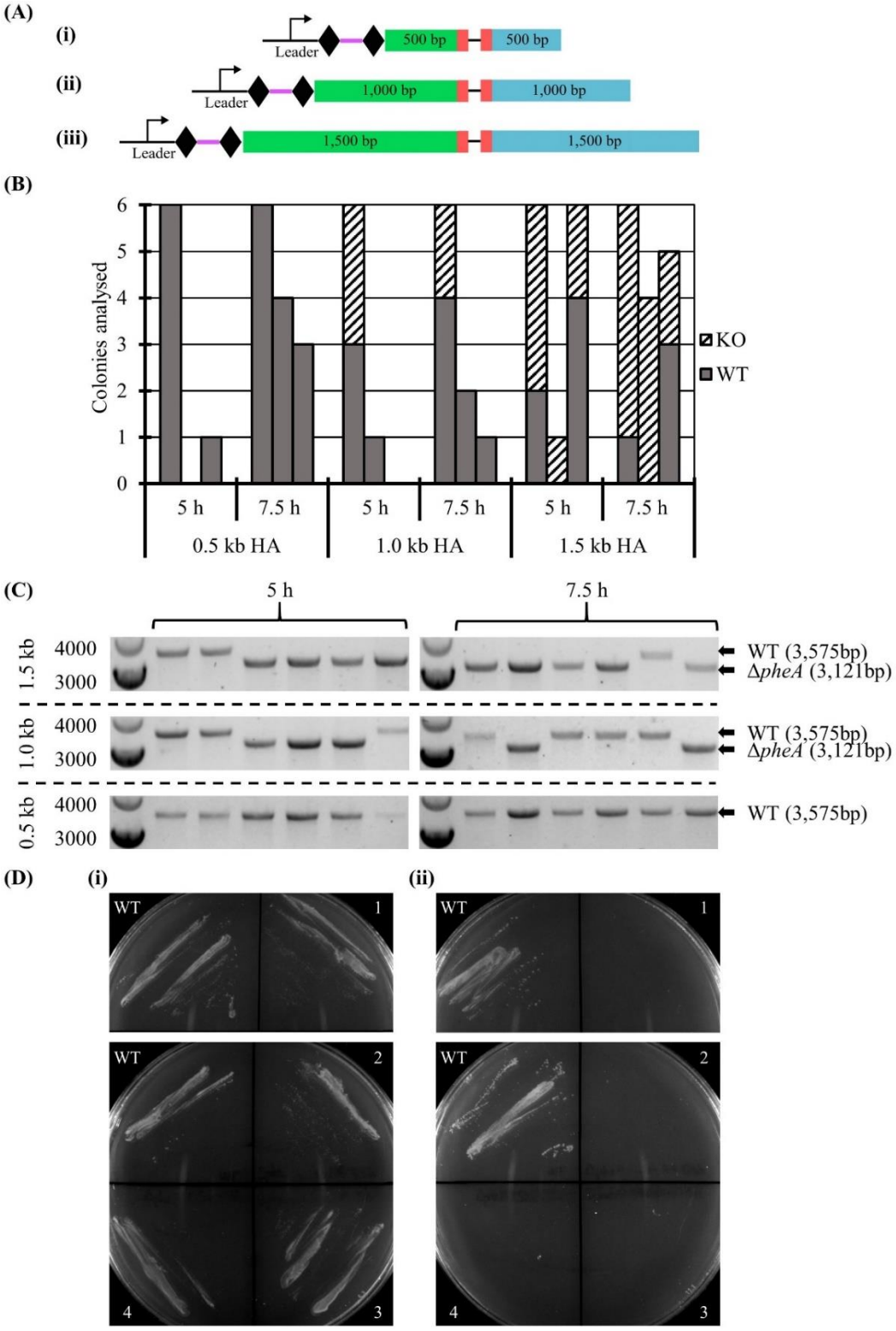

(A)
(i)
Leader
500 bp
500 bp
(ii)
Leader
1,000 bp
1,000 bp
(iii)
Leader
1,500 bp
1,500 bp
(B)
Colonies analysed
KO
WT
5 h
7.5 h
0.5 kb HA
1.0 kb HA
1.5 kb HA
(C)
5 h
7.5 h
1.5 kb
1.0 kb
0.5 kb
4000
3000
WT (3,575bp)
ΔpheA (3,121bp)
(D)
(i)
(ii)
WT
1
2
3
4

Figure 3-12: Endogenous CRISPR based deletion of 354 bp in the *A. woodii pheA* gene. (A) Representation of the application-specific modules of the three different editing plasmids containing different sizes HAs for *pheA* (red) KO in *A. woodii*. The 35 bp spacer (purple) is flanked by the native DR (diamond) and expressed under the control of the native leader sequence. HAs (LHA in green, RHA in blue) of different lengths allow for homologous recombination and deletion of the 354 bp region: **(i)** pMTL8415_pheA KO 0.5, **(ii)** pMTL8415_pheA KO 1.0, **(iii)** pMTL8415_pheA KO 1.5. **(B)** Summary of colonies analysed for the 354 bp deletion in the *pheA* gene by colony PCRs with primers FW_pheAHA1.5_out and RV_pheAHA1.5_out. Data show 3 replicates of transformation assays with homology arms of 0.5 kb, 1.0 kb and 1.5 kb for both 5 h and 7.5 h recovery times. When possible up to 6 colonies were analysed, missing bars show conditions where no colonies were obtained. **Grey**: wild type pheA locus; **Dashed**: Δ*pheA*. **(C)** Colony PCR results for the first replicate of the *pheA* KO experiment with homology arms of 0.5 kb, 1.0 kb or 1.5 kb by amplification with primers FW_pheAHA1.5_out and RV_pheAHA1.5_out. The wild type amplicon is expected at 3,575 bp and the KO amplicon at 3,121 bp. When possible up to 6 colonies were analysed. **(D)** Phenotypic analysis of the *A. woodii* Δ*pheA* strains obtained. 1 and 2 are *A. woodii* Δ*pheA* strains obtained with 1.0 kb HA with 5 h or 7.5 h recoveries respectively; 3 and 4 are *A. woodii ΔpheA* strains obtained with 1.5 kb HA with 5 h or 7.5 h recoveries respectively; WT is the wild type *A. woodii* strain. **(i)** Minimal media supplemented with 20 µg/mL of phenylalanine; **(ii)** Minimal media without supplementation.

The knockout of *pheA* followed by plasmid loss and confirmation of phenotype confirms the successful use of the Type I-B CRISPR/Cas system of *A. woodii* for genome editing and highlights how straightforward and rapid the technique is.

3.2.3.6 Validation of the system with 3.2 kb *hsdR1* knock-out and impact of cell densities and DNA concentration

3.2.3.6.1 Knock-out generation

To test the robustness of the use of the endogenous system for gene KO in *A. woodii*, the restriction enzyme gene *hsdR1* hypothesized to be responsible for cleavage of the *cas9*-containing plasmids was targeted. This gene has never been targeted before in *A. woodii* and the deletion of the whole gene (3.2 kb) represents a 10-fold increase in deletion size compared to the previous successful deletions.

Homology arms of 1,455 bp and 1,489 bp and a 36 bp spacer starting 40 bp from the LHA were designed and inserted in a pMTL84151 vector already containing the native *A. woodii* leader sequence. *A. woodii* WT was transformed with the editing vector for *hsdR1* KO, pMTL8415_hsdR1 KO (Figure 3.13A). Cell densities and amount of DNA were investigated as potential factors to increase transformation efficiency of editing vectors as follows: 100 µl of cells and 1 µg of DNA (T1); 100 µl of high density cells and 1 µg of DNA (T2); 200 µl of high density cells and 4 µg of DNA (T3).

When possible, the *hsdR1* locus of up to six colonies for each of the three methods of transformation and each recovery times were analysed by colony PCRs. While plating at 7.5 h of recovery time yielded only wild type colonies (data not shown), successful knock-outs were obtained with all three methods of transformation plated after 5 h of recovery (Figure 3.13B).

Methods T1, T2 and T3 yielded 16.7%, 50%, and 75% efficiency, respectively. The high efficiency protocol significantly increased the editing efficiency for the knock-out of the *hsdR1* gene. Sanger sequencing of the PCR-amplified plasmid sequences for T1 colonies 1 and 2, all T2 colonies and T3 colonies 1, 2 and 4 revealed the absence of the spacer in the plasmid of colonies T1 colony 2 and T3 colony 4. As previously observed, recombination between the direct repeats flanking the spacer facilitated the excision of the spacer from the plasmid, rendering it non-functional. All Δ*hsdR1* colony PCR amplicons were confirmed by Sanger sequencing. After performing the plasmid loss protocol for the six confirmed Δ*hsdR1* colonies, 80.6% of colonies tested (36 in total) had successfully lost the plasmid.

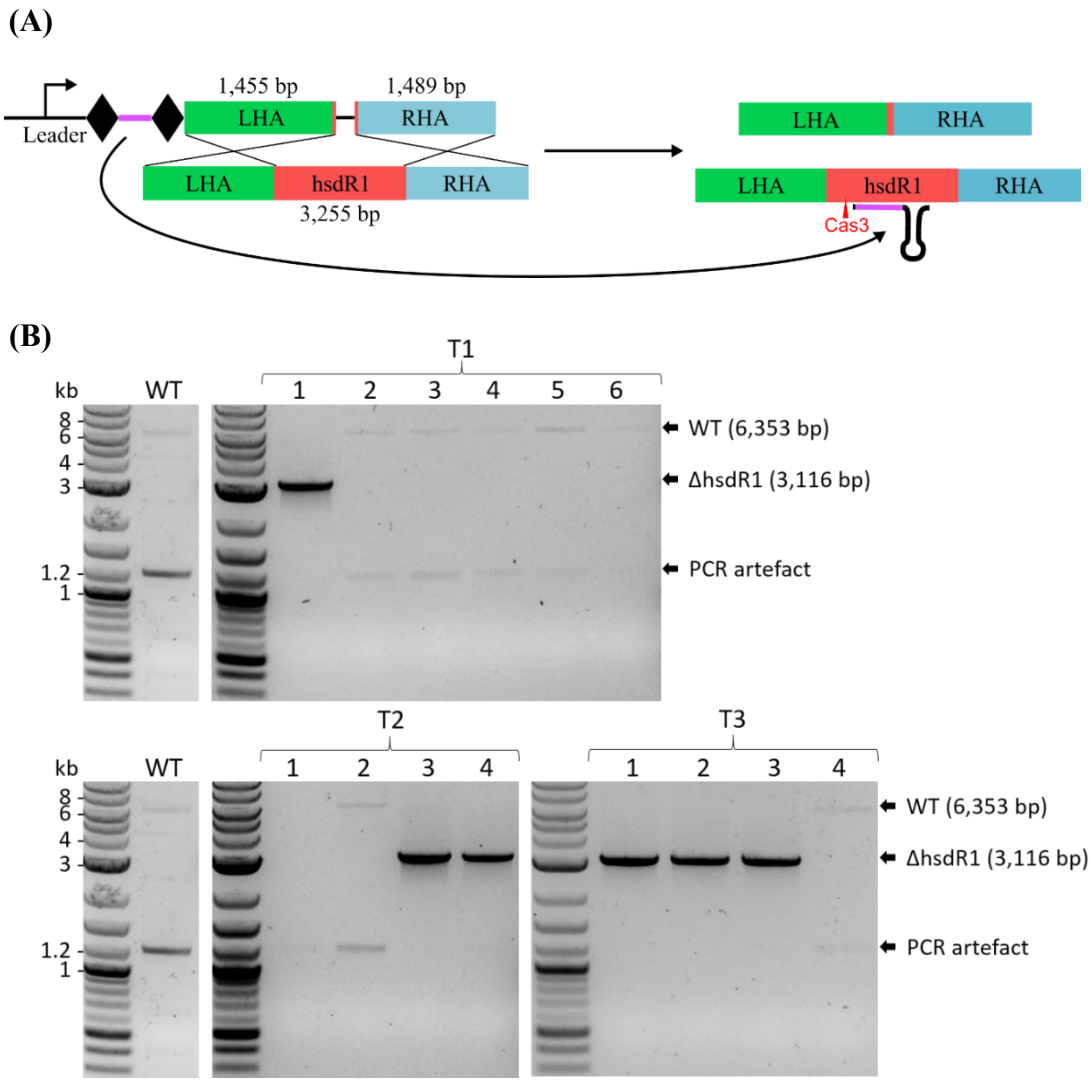


Figure 3-13: Endogenous CRISPR based deletion of the *A. woodii hsdR1* gene **(A)** Representation of the application-specific module of the editing plasmid for *hsdR1* (red) KO in *A. woodii*, pMTL8415_hsdR1_KO. The 36 bp spacer (purple) is flanked by the native DR (diamond) and expressed under the control of the native leader sequence. The editing template with a 1,455 bp LHA (green) and a 1,489 bp RHA (blue) allow for homologous recombination and deletion of a 3,237 bp region. **(B)** Colony PCR results with primers FW_hsdR1_screen_1.5 and RV_hsdR1_screen_1.5 of the *hsdR1* locus of colonies transformed with the editing vector for *hsdR1* KO. WT amplicon is expected at 6,353 bp and KO amplicon is expected at 3,116 bp. T1: standard *A. woodii* electroporation; T2: high efficiency protocol with 100µl cells and 1 µg of DNA; T3: high efficiency protocol with 200 µl cells and 4 µg of DNA.

This successful 3.2 kb deletion of *hsdR1* in *A. woodii* using its Type I-B CRISPR/Cas system confirms that the system is functional for relevant targets. It also shows that cell densities and DNA amount are limiting factors, with their increase improving the genetic engineering efficiency of this method.

3.2.3.6.2 Phenotypic analysis of the knock-out strains

In section 3.2.1.1.2, WT *A. woodii* transformation efficiencies were shown to be lowered by the presence of a 5′-TAAGN$_5$TCC-3′ motif on the plasmids. The *hsdR1* gene is the restriction subunit of the R-M system hypothesized to be responsible for recognition of this motif. To confirm this hypothesis, the R-M site 5′-TAAGN$_5$TCC-3′ was inserted in one, two or three copies in the MCS of the pMTL82151 backbone.

WT and Δ*hsdR1 A. woodii* strains were transformed with the three plasmids obtained along with the empty backbone as a positive control. Three separate colonies of Δ*hsdR1* were transformed while transformations were done in technical triplicates for the WT strain, the transformation efficiencies obtained were recorded and normalized to the ones obtained with the positive control (Figure 3.14). For two of the replicates, transformation efficiency of the Δ*hsdR1* strain with the empty backbone was lower than the transformation efficiency of the WT, representing 20.59% and 52.94% of the WT transformation efficiency.

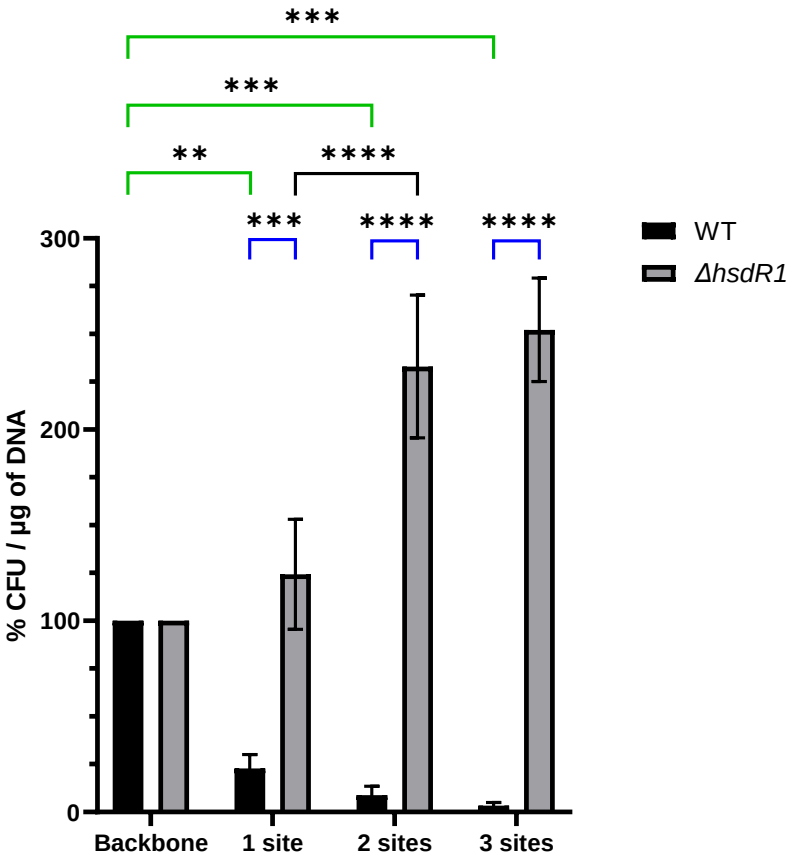


Figure 3-14: *A. woodii* WT and Δ*hsdR1* transformation efficiencies obtained with plasmids containing one to three R-M sites into normalized to the transformation efficiency obtained with the empty backbone pMTL82151. Data represent technical triplicates for WT and biological triplicates for Δ*hsdR1*. ANOVA test followed by multiple comparisons to test for significant differences in transformation efficiency; in green the transformation efficiencies of *A woodii* WT with plasmids containing 1 to 3 R-M sites compared to the empty backbone ; in blue transformation efficiencies of *A. woodii* Δ*hsdR1* compared to WT for each plasmid; in black the comparison between transformation efficiencies of *A. woodii* Δ*hsdR1* with plasmids with 1 or 2 R-M sites . **: $p < 0.005$; ***: $p < 0.001$; ****: $p < 0.0001$.

As shown by the multiple comparisons test represented in green in Figure 3.14, the presence of the R-M site on plasmids had a significant impact on transformation efficiency of *A. woodii* WT and increasing numbers of sites negatively impacted transformation efficiency.

The deletion of the HsdR1 subunit of the R-M system significantly increased relative transformation efficiencies obtained with plasmids containing the R-M recognition site compared to WT, as shown by the multiple comparisons test represented in blue in Figure 3.14. The deletion restored the transformation efficiency obtained with the plasmid containing one R-M site to the level obtained when transforming WT with the empty backbone. This data confirmed the hypothesis that the R-M system discussed recognizes the 5′-TAAGN$_5$TCC-3′ motif, they also show that the deletion of the restriction subunit alone can restore transformation efficiency.

Surprisingly, the presence of two or three R-M sites on the plasmids resulted in an increased relative transformation efficiency of *A. woodii* Δ*hsdR1* compared to plasmids with none or one site.

3.2.4 Harnessing *C. autoethanogenum* Type I-B system

3.2.4.1 Description of CRISPR locus in *C. autoethanogenum* and PAM prediction with Python script

In *C. autoethanogenum*, the CRISPR-Cas++ database identified a Type I-B system between position 1,491,001 and 1,499,580 in reverse orientation. The database does not predict a Cas8 gene, which is essential for a functional CRISPR system. A protein BLAST of the gene situated between the Cas7 and the Cas6 revealed homologies to proteins labelled Cas8b in other *Clostridium* spp. For instance, an alignment with the Cas8b protein of the Type I-B system of *Clostridium tetani* has an E-value of 1e-148. Overall, the gene organisation is the same as the one described for *A. woodii*.

Four CRISPR arrays were identified, including three with an evidence level of 4; array 2 is directly downstream of the Cas locus with 21 spacers while array 3 (42 spacers) and array 4 (33 spacers) are immediately upstream of the Cas locus. All those arrays have a

similar direct repeat with only one nucleotide varying: "ATTTAAATACATCT(C/T)ATGTTGAGGTTCAAC". This suggests that they are related and should give coherent results when run in the Python script. Hits with less than 15% mismatch are presented in Supplementary Table 12 (hits with mismatch between 15% and 20% are presented in Supplementary Table 13 for further information). Array 4 had the most hits with 15 hits in phage regions or phage related genes.

Overall, the Python script predicted a quite clear TCH / TTR putative 5′ PAM. Most notably, the spacer in position 27 of this array is a perfect match to a phage related gene in *C. ljungdahlii* with, in position 5′, the sequence "CTTCA". The TCA PAM was selected for harnessing the endogenous system as a genetic engineering tool in *C. autoethanogenum*.

3.2.4.2 Genome engineering capacity validation by *pyrE* knockout in *C. autoethanogenum*

3.2.4.2.1 Generation of knock-out strain

Design of the *pyrE* KO editing vector was based both on a 561 bp in-frame *pyrE* deletion already performed in *C. autoethanogenum* in our laboratory with CRISPR/Cas9 (Dr. Francois Seys, personal communications) and the native CRISPR array 4. Homology arms of 868 bp and 797 bp used in the CRISPR/Cas9 method were used and the synthetic array was designed to mimic the native array 4, with the native leader sequence controlling the expression of a 35 bp spacer flanked by direct repeats.

WT *C. autoethanogenum* was conjugated with three plasmids, with HAs alone, expressed spacer alone, and both HAs and expressed spacer (Figure 3.15A). Analysis of the *pyrE* locus of the colonies obtained by colony PCRs (Figure 3.15B) demonstrated that the editing vector containing both the HAs and the expressed spacer had an editing

efficiency of 100%. In contrast, vectors harbouring only one part of the editing cassette did not produce any Δ*pyrE* mutants.

To establish a plasmid loss protocol adequate to the application of the use of *C. autoethanogenum* endogenous CRISPR system, two protocols were tested, as described in Materials and Methods (section 2.7.3.2). An additional subculture without selection increased plasmid loss rate from 33.3% to 58.3% (out of 12 colonies tested in both conditions).

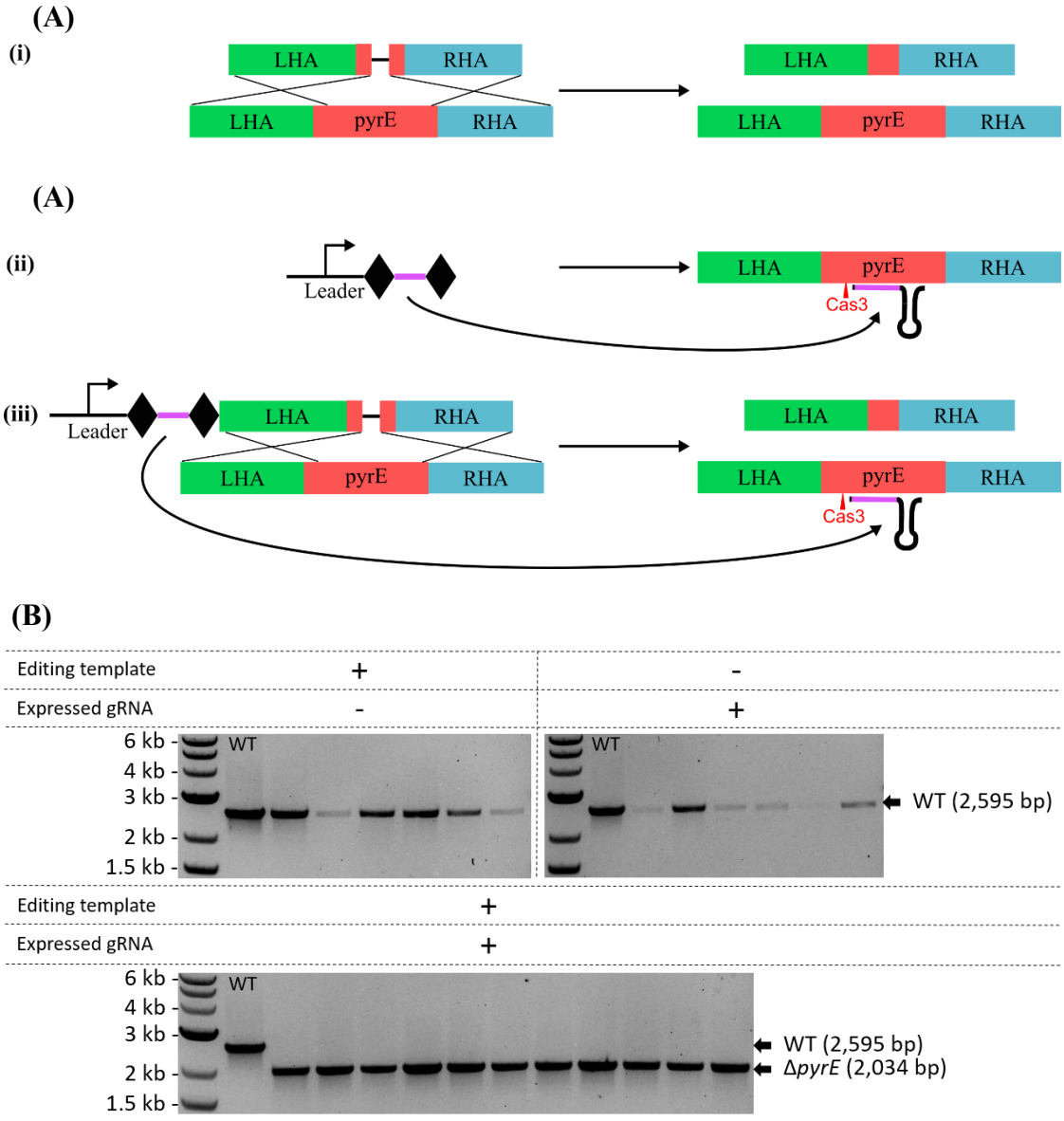


Figure 3-15: Endogenous CRISPR based deletion of the *C. autoethanogenum pyrE* gene. (A) *pyrE* (red) deletion strategy in *C. autoethanogenum* using the native Type I-B system. Representation of the application-specific modules of the following vectors: **(i)** pMTl8315_CLAU_pyrE_HA, HAs alone (868 bp LHA in green, 797 bp RHA in blue) allow for homologous recombination in some of the cells of the population; **(ii)** pMTL8315_CLAU_leader3_pyrE gRNA, the expression of the 35 bp spacer (purple) flanked by DRs (diamonds) alone leads to self-cleavage by *Cas3* (red) at the *pyrE* gene and cell death; **(iii)** pMTL8315_CLAU_pyrE KO, the full editing cassette allows for homologous recombination, deletion of the 561 bp region and selection of the successful KO. **(B)** Colony PCR results for the pyrE locus of *C. autoethanogenum* with primers FW_pyrECLAU_screening and RV_pyrECLAU_screening, the wild type amplicon is expected at 2,595 bp and the KO amplicon at 2,034 bp. + and – indicate respectively the presence or absence of the corresponding element.

3.2.4.2.2 Phenotype validation

The phenotype of the strains obtained after plasmid loss was confirmed by plating on minimal media with or without uracil supplementation (Figure 3.16).

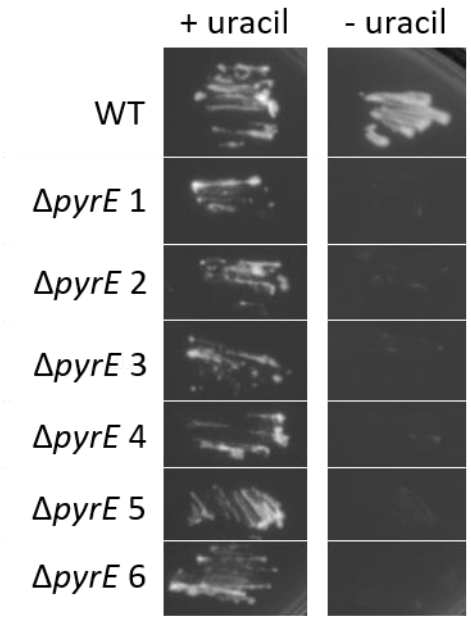


Figure 3-16: Phenotypic analysis of the *C. autoethanogenum* Δ*pyrE* strains obtained. WT and 6 colonies of Δ*pyrE* *C. autoethanogenum* strains obtained after plasmid loss were plated on minimal media supplemented (+ uracil) or not (- uracil) with 20 µg/mL of uracil.

This is the first report of successful use of *C. autoethanogenum* Type I-B endogenous system for genome editing. It demonstrates the robustness of the developed workflow from automated PAM identification to gene knock-out.

3.3 Discussion

3.3.1 *S. pyogenes* CRISPR/Cas9 system in *A. woodii*

Analysis of the R-M systems of *A. woodii* revealed a Type I R-M system encoded by Awo_c08800, Awo_c08810 and Awo_c08870 and recognizing site 5′-TAAGN$_5$TCC-3′, this site is rarely present in modules and vectors commonly used for genome engineering *A. woodii* but present in the *cas9* gene. It was additionally found in the low copy number Gram-negative replicon p15a and riboswitch G. Its presence in the MCS of a backbone significantly reduced the *A. woodii* transformation efficiency obtained compared to an empty backbone, confirming the activity of the restriction subunit of the R-M system. To alleviate this barrier when transferring plasmids containing *cas9* in *A. woodii*, the impact of the deletion of the recognition site in the *cas9* sequence was studied. Surprisingly, a silent mutation abolishing the recognition site did not lead to a significant increase in transformation efficiency (Figure 3.6). The transformation efficiencies obtained with *cas9*-containing plasmids with mutated R-M sites were very variable. This observed variation of effectiveness could be due to the slight variations of transformation or competent cell conditions and cell state in biological replicates. In Type I R-M systems, the expression of each of the three genes is needed for a functional system and is finely balanced (Mruk & Kobayashi, 2014; Roer et al., 2015). A slight variation in cell state could potentially cause differences in the level of protection against exogenous DNA.

It could also be hypothesized that the silent mutation was not sufficient for complete abolishment of plasmid targeting, leading to DNA transfer hindrance. The Cas9 toxicity might also be the main barrier for DNA transfer of *cas9*-containing plasmid, resulting in a comparably negligible effect of the R-M recognition site presence.

In order to prevent future barriers for DNA transformation, plasmid sequences were mined for the recognition site 5′-TAAGN$_5$TCC-3′. It was found in the sequence of the *ltrA*, encoding the Intron-Encoded Protein (IEP) in the ClosTron system (Kuehne & Minton, 2012). To validate the impact of the R-M site on the cas9-containing plasmids, a methylation strain expression the methylation gene *hsdM2* could be built; transformation efficiencies obtained with methylated plasmids compared to unmethylated plasmids would reflect the impact of the R-M system. This methylation strain could then be employed when 5′-TAAGN$_5$TCC-3′ containing modules are needed, circumventing the need for mutation of the R-M site(s).

None of the sequences analysed contained the recognition site 5′- GATN$_6$TGC -3′ of the second Type I R-M system identified, encoded by Awo_c17130 and Awo_c17140. For future genome engineering experiments in *A. woodii*, particular care should be taken to identify all R-M systems potentially targeting the exogenous DNA designed and strategies put in place to evade them.

In the RiboCas system tested, *cas9* is under the control of the P$_{fdx\text{-}RbE}$ inducible promoter, a ferredoxin promoter linked to a theophylline inducible riboswitch (Canadas et al., 2019). Nonetheless, the presence of *cas9*, under non-induced promoter control and without sgRNA present, was shown to lead to a significant decrease in transformation efficiency. The presence of the Cas9 encoding gene on the plasmid decreased transformation efficiency 6-fold, compared to plasmids encoding dCas9 with no nucleolytic activity. It can be concluded that this nucleolytic activity is responsible for the decrease in transformation efficiency and that the expression control exerted by P$_{fdx\text{-}RbE}$ is not sufficient to repress *cas9* expression. This is consistent with previous studies

showing that Cas9 toxicity decreases transformation efficiencies and hinders genome engineering (Huang et al., 2016; Ingle et al., 2019; Jinek et al., 2012; Q. Li et al., 2016; Xu et al., 2015).

In the literature, modulation of the nucleolytic activity by mutation of the *cas9* to generate a nickase led to successful genome engineering (Huang et al., 2016; Ingle et al., 2019; Y. Li et al., 2016; Xu et al., 2015). However, in *A. woodii*, the complete abolishment of nucleolytic activity and mutation of the R-M system recognition site as described above was not sufficient to restore transformation efficiency to the efficiency achieved with the empty vector. In *E. coli*, it has been shown that dCas9 binds non-specifically to DNA, especially when not bound to sgRNA (Jones et al., 2017; Zhang & Voigt, 2018). dCas9 operates by opening DNA double strand to identify PAM sites (Jinek et al., 2012; Sternberg et al., 2014). At the genome scale, these openings of chromosome DNA could explain dCas9 toxicity in bacteria (Zhang & Voigt, 2018). It is also likely that the increase in plasmid size linked to the addition of the Cas9 gene resulted in a decrease in plasmid transformation efficiency.

Overall, the use of a Cas9 with no nucleolytic activity combined with the alleviation of the R-M barrier by silent mutation in the recognition site still led to a significant decrease of transformation efficiencies compared to the empty backbone. Similar to previous studies in acetogens, these results show the use of *cas9* to be unsuitable for genome engineering in *A. woodii*. Accordingly, the use of *A. woodii* endogenous CRISPR system was explored, starting with the development of an *in silico* pipeline to facilitate PAM prediction.

3.3.2 Harnessing endogenous CRISPR/Cas systems

3.3.2.1 Python script development and validation for PAM prediction

Endogenous CRISPR/Cas systems are starting to offer alternative genetic engineering tools for microorganisms which contain them. PAM identification is the first necessary step and is also recognized as one of the challenges for the use of endogenous CRISPR systems for genetic engineering (Zhou et al., 2021).

In this study, a dedicated bioinformatics tool was developed to allow for fully automated predictions based on the data obtained from the CRISPRFinder database. It was successfully validated against published literature for validated PAMs in *Clostridium* spp. (Table 3.1). It was then further used to predict PAM candidates in other *Clostridium* spp., methanotrophs and *M. thermoacetica*. The nucleotide pairs TC and TT were found to be prevalent in the PAM candidates of both Type I-B and Type I-C systems of these microorganisms (Table 3.2, Table 3.3). This observation is coherent with the Type I-B (Boudry et al., 2015; Maikova et al., 2019; Pyne et al., 2016; Walker et al., 2020; Zhang, Zong, et al., 2018) and Type I-C (Leenay et al., 2016) systems characterized in the literature.

During the development of this study, other tools to determine functional PAMs were published. In 2021 Vink and co-workers analysed more than 200,000 unique spacers (Vink et al., 2021) and in 2022 Rybnicky and co-workers developed a tool called Spacer2PAM to predict functional PAMs (Rybnicky et al., 2022). The concept of the three tools, including the Python script developed here, are very similar: spacers are run through BLAST and flanking sequences of the matches are retrieved to compile a prediction of the PAM. Interestingly, BLAST and subsequent filtering parameters vary slightly between the different methods. In Rybnicky et al. and in the Python script

developed here, the BLAST is run in the NCBI nucleotides database excluding eukaryotes (Rybnicky et al., 2022). Vink et al., in contrast, use the NCBI nucleotide database along with the metagenomics databases and did not exclude eukaryotic sequences. They observed that only 1% of matches were found in eukaryotic or eukaryotic viral sequences, confirming a low rate of false-positive (Vink et al., 2021). Spacer2PAM, contrary to the Python script described here, does not predict array direction but instead requires the user to input the information, which could lead to mistakes

Filtering of false-positives is done only based on global mismatch (Vink et al., 2021) or a combination of parameters, including gap number, e-value, length of the hit, starting position of the hit (Rybnicky et al., 2022). However, interference relies on the recognition of a seed region (Semenova et al., 2011) with up to two mismatches in the seed tolerated for functional interference (Xue et al., 2015). Accordingly, the more stringent mismatch parameters for the seed region adjacent to the PAM in the Python script developed here were relevant.

CRISPR/Cas systems are prokaryotic defence mechanisms against invading elements (Makarova & Koonin, 2015), biologically relevant hits are thus likely found in Mobile Genetic Elements (MGEs), including plasmids, phages, prophages, and any related proteins. The Spacer2PAM tool does not automatically filter based on MGEs and only has the option to filter matches found in phages by the PHASTER platform. The Python script described here has more in-depth filtering capabilities for MGEs sequences but Rybnicky and co-workers conveniently created a filtering mode to display all the filtering parameter combinations and their corresponding PAM score to determine the most appropriate combination for each microorganism (Rybnicky et al., 2022). The similar performances between Spacer2PAM and the Python script, albeit a more

comprehensive filtering parameters selection in Spacer2PAM, could be explained with the refined selection for biologically relevant hit sequences presented here.

Because *C. autoethanogenum* is an important industrial strain, its CRISPR/Cas system has been of interest in recent years (Pyne et al., 2016; Rybnicky et al., 2022). Consequently, *C. autoethanogenum* can serve as a case study to compare the different methods of PAM elucidation. In 2016, the analysis was performed manually and the array direction was not taken into account, leading to inaccurate PAM prediction (Pyne et al., 2016). The Python script output two putative 5′ PAMs TCH and TTR; a TCA PAM was confirmed *in vivo* by successful genome editing (this thesis). Spacer2PAM predicted a TTNN PAM consensus while a NYCN consensus was experimentally confirmed (Rybnicky et al., 2022). It was interesting to note that in Rybnicky et al., the pooled library experiment and the unpooled interference assay led to slightly different conclusions (Rybnicky et al., 2022). CRISPR systems have been shown to recognise a variety of PAMs and the interference mechanism tolerates mismatches in the PAM (Xue et al., 2015); mismatched PAM could lead to interference or primed adaptation (Xue et al., 2015). It was similarly hypothesized that a wider variety of PAMs can be recognised for interference compared to adaptation. Both of these hypotheses would result in a better defence system for the host microorganism, less likely to suffer from invasion from mutated MGEs or closely related MGEs (Rybnicky et al., 2022; Xue et al., 2015). Finally, the computational tool developed by Vink and co-workers identified matches for single spacers but was not able to identify potential PAM candidates *C. autoethanogenum* (Vink et al., 2021). However, they showed a direct link between the direct repeats sequence and the identified PAMs. This expanded on a previous study about the role of the three nucleotides from the direct repeat directly adjacent to the spacer (Xue et al., 2015); they were found to be essential to prevent self-targeting of the

CRISPR array and as such could already be hypothesized to be linked to the functional PAM. In Vink et al. (2021), spacers from different microorganisms with identical direct repeats were clustered and PAM prediction achieved for each direct repeat sequence cluster. *C. autoethanogenum* direct repeats were found to have high homology to a variety of *Clostridium botulinum* strains; the corresponding predicted PAM 5′ TCNX corroborates *in vivo* experiments performed in this study and by Rybnicky and co-workers (Rybnicky et al., 2022).

For *A. woodii* Type I-B CRISPR system, the Python script did not manage to compute a clear consensus, with most matches having a high level of mismatch with the spacers from the native CRISPR array. However, in total, 3 out of 8 matches contained two consecutive cytosines. It also revealed an interesting match with a phage protein and its corresponding 5′ CCA PAM, PAM which was shown to be functional in the Type I-B system of *C. difficile* (Boudry et al., 2015; Maikova et al., 2019). Overall, even if no clear consensus was found for *A. woodii*, the script allowed for rapid identification of matches with MGEs and identification of functional PAM candidates.

This diversity of functional PAMs is proving challenging for both *in silico* prediction and *in vivo* validation. One of the limitations of the bioinformatics tools described, including the Python script developed here, is the fact that PAM prediction relies solely on consensus of the DNA sequence flanking the protospacer, regardless of the spacer analysed. Xue and co-workers showed in 2015 that CRISPR/Cas systems are actually quite tolerant to mismatches in PAM or seed region (Xue et al., 2015). The authors demonstrated that, depending on the spacer, mismatches could either have no impact on interference, block interference or trigger primed adaptation (Xue et al., 2015) and found PAMs that could result in priming and/or interference. This implies that compiling the sequences that flank the protospacers might lead to identification of PAMs that could

result only in primed adaptation and not interference, this could explain the rather complex results obtained from the Python script for some microorganisms. Conveniently, Spacer2PAM from Rybnicky et al. (Rybnicky et al., 2022) provides a PAM score to help the user judge of the quality of each PAM within one CRISPR/Cas system. They have also shown that even low number of alignments can lead to the successful elucidation a functional PAM. Future tools could use that score but also focus on optimizing the user output to reflect the functional PAM diversity, without compromising user-friendliness and clarity.

A better understanding of the adaptation and the interference mechanisms is necessary to improve the parameters for PAM identification, but all of those recently developed tools and their output pave the way for a PAM prediction pipeline combining (i) spacer BLAST to find MGE matches and (ii) direct repeat homology to find candidate PAMs based on both *de novo* identification and mining of published PAMs.

3.3.2.2 Use of endogenous CRISPR/Cas system for genome engineering in acetogens

3.3.2.2.1 Optimization of the workflow in *A. woodii* and *pyrE* KO in *C. autoethanogenum*

In this study, proof-of-concept for the complete workflow of genome editing with the endogenous CRISPR/Cas system was established in *A. woodii*. After *in silico* analysis, the functional PAM was confirmed *in vivo,* and the leader sequence was characterised. Similar to the interference assay completed by Rybnicky et al. (Rybnicky et al., 2022), the transformation efficiencies observed were variable between the five replicates. However, the assay showed that the double cytosine is crucial for interference in *A. woodii.*

When using the endogenous CRISPR system, providing homology arms for homologous recombination allows evasion from self-targeting and cell death. Other pathways exist for CRISPR-based targeting evasion, including disruption of targeting by mutation of either the target site or the targeting mechanisms (Wimmer & Beisel, 2019). Analysis of the escape mutants obtained here predominantly revealed inactivation of the CRISPR-based targeting by loss of the plasmid-encoded spacer, likely through recombination of the direct repeats as previously hypothesized in the literature (Canez et al., 2019; Gomaa et al., 2014). Different parameters are reported to have an impact on editing efficiency, including spacer design and expression, Gram-positive replicon, editing template length, and cell density; they were each investigated with successive deletion targets and are discussed below.

In *A. woodii* and *C. autoethanogenum*, the spacers were chosen to have lengths corresponding to the average length of the protospacers in the native arrays. The native CRISPR arrays of *A. woodii* have a GC content comprised between 18.92% and 61.11% with an average of 38.39%, the spacers designed here had a GC content comprised between 40% and 60% (Javaid & Choi, 2021; Zhou et al., 2021). Other parameters for spacer design appear to have an impact on interference efficiency but remain elusive. In bacteria, spacer sequence was shown to have an impact on interference (Bourgeois et al., 2020; Xue et al., 2015). Studies performed in mammalian cells with CRISPR/Cas9 have demonstrated that sgRNA sequence impact interference, likely through sgRNA transcription and stability (Doench et al., 2014; Wu et al., 2014). Target sequence accessibility also modulates interference efficiency (Chari et al., 2015; Wu et al., 2014).

Spacer expression replicating the native arrays was effective in *A. woodii*, *C. autoethanogenum,* and in previous studies (Hidalgo-Cantabrana et al., 2019; Pyne et al., 2016; Zheng et al., 2019).

The initial demonstration of *pyrE* KO in *A. woodii* was obtained using both pCD6 and pBP1. Subsequent editing experiments were confined to plasmids based on the former replicon in view of its comparative instability (Baker et al., 2022), a desirable property when plasmid loss is required.

Targeting of *pheA* confirmed that the use of editing templates comprising larger HAs (1.5 kb compared to 0.5 kb and 1.0 kb) has a positive impact on genome editing efficiency (Zheng et al., 2019). Based on this study and the literature, it can be concluded that the minimal sizes for HAs are microorganism- and condition-dependent. It can be hypothesized that this is due the different levels of expression of the spacer as well as the host intrinsic homologous recombination capabilities. It would be interesting to correlate the abundance level of the spacer in the cells with the minimal HAs length required for successful genome editing. Lack of correlation would indicate that the homologous recombination capabilities of the strains is the main element determining the strategy to adopt for spacer expression and editing template length. The study of these capabilities in naïve microorganisms could, thus, help guide parameter selection for genome engineering experimental design.

Not unexpectedly, editing efficiencies were also enhanced through the use of more effective transformation procedures. In *C. autoethanogenum*, replicon, homology arms, and transformation protocol previously used for CRISPR/Cas9 genome editing led to successful editing with its endogenous CRISPR system. However, only one knock-out was attempted, further experiments are necessary to confirm the robustness of the method in *C. autoethanogenum.*

3.3.2.2.2 *A. woodii ΔhsdR1* strain phenotype

Increasing the number of R-M recognition sites on the plasmids led to decreasing transformation efficiencies of *A. woodii*. In Type I systems, the $R_2M_2S_1$ binds and translocates DNA, with cleavage happening at a non-specific site, distant to the binding site (Murray, 2000). Translocation happens in a bi-directional manner from the binding site (Studier & Bandyopadhyay, 1988) and cleavage is stimulated when translocation is paused by a physical barrier (Janscak et al., 1999; Studier & Bandyopadhyay, 1988). For circular DNA containing only one recognition site, DNA cleavage occurs when all the DNA has been translocated and the process is stalled (Janscak et al., 1999). When multiple Type I R-M recognition sites are present, cleavage occurs in the region between the two sites (Janscak et al., 1999; Studier & Bandyopadhyay, 1988). Collision of two translocating enzymes, even from different families, stimulates cleavage (Janscak et al., 1999; Murray, 2000). It was also hypothesized that dimerization of enzymes bound to distant sites could improve restriction (Ellis et al., 1999; Murray, 2000). Both hypotheses could explain the observe decrease in transformation efficiency when an increasing number of R-M sites are present in plasmids to be transformed in WT r^+m^+ *A. woodii*.

As expected, deletion of the *hsdR1* gene resulted in increased relative frequency of transfer of plasmids carrying its corresponding restriction recognition site(s). Interestingly, plasmids containing two or three R-M sites resulted in slightly higher transformation efficiencies of *A. woodii ΔhsdR1* strain compared to the backbone alone or with just one R-M site. This suggests that the methylation and specificity subunits of the system confer an advantage for transformation of *A. woodii* with plasmids containing the R-M recognition site in the absence of the restriction subunit. In *E. coli* K-12, alleviation of restriction from the restriction alleviation phage protein (Ral) was

shown to enhance methylation (m*) (Loenen & Murray, 1986). Additionally, a point mutation in the restriction subunit, resulted in a r⁻m* strain with enhanced modification and converted the R-M system into a *de novo* methylase system (Kelleher et al., 1991), methylating unmethylated DNA (Razin & Riggs, 1980). Taken together, this could explain the increased transformation efficiency of plasmids containing multiple R-M sites in the *A. woodii* Δ*hsdR1* strain.

For two of the three biological replicates, transformation efficiencies were lower with the Δ*hsdR1* strain than in the WT when transforming the empty backbone. Adenine-specific methylation systems have complex and diverse functions in bacterial cells, including but not limited to regulation of conjugation, DNA replication and various transcriptional and posttranscriptional gene regulations (Adhikari & Curtis, 2016; Low & Casadesus, 2008; Vasu & Nagaraja, 2013; Wion & Casadesus, 2006). Further study is necessary for a full understanding of the impact of the *hsdR1* deletion in *A. woodii* and to determine if the lower transformation efficiencies obtained with the Δ*hsdR1* strain are caused by the KO or by competent cell preparation variations.

3.4 Key outcomes

This study reports the characterisation of the barriers to the use of the *S. pyogenes* Type II CRISPR/Cas9 system in *A. woodii*. The toxicity of the Cas9 nuclease was shown to be a predominant hindrance for plasmid transfer. A Type I N^6-adenine methyltransferase system was shown to be an additional barrier for *A. woodii* transformation with exogenous DNA.

As an alternative, the use of endogenous CRISPR/Cas systems was investigated. A Python script was developed for fast and efficient PAM elucidation of endogenous CRISPR/Cas systems to enable their use as genome engineering tools. After validation

against published literature, it allowed for characterisation of Type I-B endogenous systems of two industrially-relevant microorganisms before their use for genome engineering in only a few steps. Editing cassettes were designed to be synthetic CRISPR arrays, replicating the native arrays of each host microorganism. They consisted of the leader sequence expressing an adequate spacer, flanked by direct repeats. The editing efficiencies varied between targets and between microorganisms but this work is the first report of successful use of the endogenous CRISPR/Cas systems of *A. woodii* and *C. autoethanogenum* for genome engineering.

Compared to most CRISPR/Cas9 based techniques, the workflow presented here removes the need for sophisticated inducible promoter systems necessary to mitigate Cas9 toxicity (Canadas et al., 2019). When compared to established techniques, such as ACE, it significantly reduces the number of restreaking steps required, lowering the risk of SNP emergence. These results are promising for the more widespread leveraging of *A. woodii, C. autoethanogenum* and more generally recalcitrant bacteria containing endogenous CRISPR systems.

CHAPTER 4: Elucidation of essential and detrimental genes during heterotrophic and autotrophic fermentations

4.1 Introduction

4.1.1 Transposon mutagenesis in *Clostridium*

The ability to create mutations is a crucial requirement for determination of gene function. Two fundamentally different strategies may be applied. On the one hand, reverse genetic approaches use gene editing tools to deliberately mutate a known gene, and then assess the phenotypic consequences. On the other hand, forward genetic approaches use random mutagenesis and selection of a specific phenotype, subsequent determination of genotype enable association of phenotype with genotype. Contrary to reverse genetics, forward genetics does not require any prerequisite knowledge of gene function.

The most common mutagenesis strategy used in forward genetics is the utilization of a transposase and a transposon carrying an antibiotic resistance marker which is used to demonstrate its insertion in the genome. Transposon mutagenesis is used to generate of large-scale insertional mutant libraries. Transposition can occur by different mechanisms, depending on the transposon, but is commonly mediated by a cut-and-paste reaction in which a transposase recognizes the inverted direct repeats (ITR) flanking the transposon sequence and enables the random transposition of the transposon element into a different locus (Lampe et al., 1998). Subsequent mutant selection steps can help elucidate genetic-phenotypic relationships either by isolation of single mutants or whole libraries. Selection of specific mutants based on a specific phenotype allows for identification of genes involved in said phenotype. For example, in *C. perfringens* the screening of 10,000 mutants for impaired gliding motility (Liu et al., 2013) and subsequent sequencing of the mutants displaying this phenotype enabled the generation of a list of genes which impacted the gliding motility function.

More high-throughput methods can also be used by exposing the pooled library to specific conditions, followed by simultaneous sequencing of the insertion sites of every single mutant in the pool by transposon insertion sequencing (TIS) enabled by NGS such as Illumina. In 2009, four separate TIS methods were published describing the high-throughput sequencing of transposon library samples (Gawronski et al., 2009; Goodman et al., 2009; Langridge et al., 2009; van Opijnen et al., 2009). Despite some differences, they rely on the same principles (van Opijnen & Camilli, 2013). In the case of Transposon Directed Insertion Sequencing, or TraDIS (Langridge et al., 2009), fragmentation of the DNA is performed by sonication, followed by transposon-genome junctions enrichment by PCR and fragment size selection on agarose gel.

To analyse the data, an insertion index is calculated from the number of unique insertion site (UIS) in each gene. Plotting of the distribution of insertion index for every gene revealed a bi-modal distribution, with one population containing genes with low insertion index and another with genes with higher insertion index (Langridge et al., 2009). Genes with low insertion index are defined as essential, transposon insertion in their encoding sequence results in a lower strain fitness and the mutants get eliminated from the population. Genes with a higher insertion index have no apparent impact on strain fitness. The library is saturated when a clear separation of the two populations is achieved and the number of genes with an ambiguous status is reduced (Langridge et al., 2009). TraDIS represents a high-throughput method for the elucidation of complex gene networks and metabolic pathways.

Various transposon systems have been characterised and employed in different microorganisms. Among them, the *Himar* C9 transposase originates from the horn fly and has been of particular interest. It was first studied in vitro (Lampe et al., 1998), transposition efficiency was sensitive to temperature, with an optimal at 28ºC, size of

the transposon, and preferentially transposed into TA target sites. The latter made the *Himar*1 transposase an excellent candidate for use into organisms such as *Clostridium* spp, which have a low GC content.

Clostridium are of particular interest for health reasons as a number of them are responsible for lethal diseases. Spores are often crucial for disease transmission, and as such, the TraDIS method has been used mainly to determine the genes required for spore formation and germination in *Clostridium* (Cartman & Minton, 2010; Dembek et al., 2015; Zhang et al., 2015). In most instances, the libraries created achieved a saturation level enabling significant conclusions to be drawn for gene essentiality conclusions. However, to generate those results accommodations had to be made to ensure the stability of the plasmid during the *E. coli* cloning steps, saturation of the library, and plasmid loss after the transposition event occurred.

To ensure the stability of the plasmid during the *E. coli* cloning steps, inducible promoters have been used to precisely control the timing of expression of the transposase. The *bgaL* lactose inducible system was used in *C. perfringens* (Liu et al., 2013) and the tetracycline inducible promoter in *C. difficile* (Dembek et al., 2015). In other instances, only repression of the transposase activity in the *E. coli* host was necessary rather than a precise control of the transposase expression. For use in *C. difficile*, the transposase was placed under the control of the *C. difficile* toxin B promoter (P_{tcdB}), only recognized by the *C. difficile* alternate sigma factor TcdR (gene *tcdR*). As a result, the transposase expression is activated only upon transformation of cells containing *tcdR* into their genome. Transfer of the method was achieved by insertion of *tcdR* into the genomes of *C. acetobutylicum*, *C. sporogenes* (Zhang et al., 2015), and *C. autoethanogenum* (Woods et al., 2022) using ACE.

An additional system was put in place to prevent premature transposition events to occur in the *E. coli* shuttle host. The transposon sequence was flanked with a copy of the *codA* gene and the strong promoter of the *C. acetobutylicum* thiolase gene (P_{thl}). CodA catalyzes the conversion of 5-fluorocytosine (5-FC) into the toxic 5-fluorouracil. Transposition places *codA* under the control of P_{thl} and production of CodA, a lethal event in the presence of 5-FC.(Woods et al., 2022).

The transformation efficiency obtained with specific Gram-positive replicons and their stability inside the cells remains a challenge. While low transposition events have been a problem, resulting in low saturation and inability to elucidate essential genes (Dembek et al., 2015), remaining plasmid can reduce the diversity of the library and prevent significant conclusions concerning gene essentiality. Plasmid loss can be achieved by utilization of a suicide vector lacking a Gram-positive replicon. In *Clostridium* spp, however, the use of suicide vectors for effective transposon delivery is limited by the low frequency of gene transfer. In *C. difficile*, replicon pBP1 was found to be suitable for transposon library creation, with high transformation efficiency and relatively high instability, representing a "pseudo-suicide" replicon (Cartman & Minton, 2010). To circumvent the use of suicide vectors altogether, application of a galactokinase (*galK*) negative selection marker was established (Liu et al., 2013). Addition of galactose, in the presence of *galK* but absence of *galT*, results in accumulation of toxic intermediates. The library was created in a Δ*galKT* strain and gene *galK* was introduced on the transposon plasmid. After plasmid transformation, selection on galactose ensured negative selection of the plasmid and plasmid loss. More recently, a conditional replicon was developed based on the observation that the P_{fac} promoter could be repressed in the presence of the *E. coli lacI* gene (itself under the control of the *C. beijerinckii* phosphotransbutyrylase promoter (P_{ptb})) (Zhang et al., 2015). Addition of IPTG resulted

in relief of the repression and readthrough into the pCB102 Gram-positive replicon, in turn, causing plasmid instability. However, this conditional replicon relies on IPTG, which has been shown to be a limiting factor in several *Clostridium*, including C. *beijerinckii*, *C. botulinum*, and *C. autoethanogenum*. In *C. difficile*, replicon readthrough was also designed to occur (Dembek et al., 2015); the tetracycline promoter used to control expression of the transposase is constituted of two divergent promoters, addition of tetracycline relieves repression and allows for both the expression of the transposase and readthrough of the pCD6 replicon, resulting in plasmid instability.

As a further measure to prevent contamination of data by plasmid reads, the I-SceI restriction site is often placed on library plasmid. This site is not present on the genome and allows for restriction cleavage of plasmid-derived PCR amplicons during sequencing sample preparation.

4.1.2 Gene overexpression for improved fitness in *A. woodii*

Acetogens live at the thermodynamic edge of life, they are able to fix CO_2 while conserving energy (Schuchmann & Müller, 2014). However, ATP limitations render them difficult to engineer for growth or yields improvements of metabolites. In *A. woodii*, attempts at expressing pathways for increased energy conservations or CO_2 fixation have led to disappointing results for strain improvements. This can often be explained by the intricate balance between the different cellular processes and overexpression leading to metabolic imbalance.

In the WLP, two bottlenecks were hypothesized: (i) formate processing by the THF-dependent enzymes (Peters et al., 1999); and (ii) ATP generation by conversion of acetyl-CoA to acetate by the combined action of a phosphotransacetylase (*pta*) and an acetate kinase (*ack*) (Ragsdale & Pierce, 2008). Separate plasmid-based overexpression

(Straub et al., 2014) of either the THF-dependent enzymes (formyl-THF-synthase, methyl-THF-cyclohydrolase, methyl-THF-dehydrogenase, and methyl-THF-reductase) or the *pta-ack* genes from *C. ljungdahlii* was achieved in *A. woodii*. Overexpression of the THF-dependent enzymes resulted in a higher increase in acetate kinase activity than with *pta-ack* overexpression, demonstrating an increase of carbon flow through the WLP in both cases (Straub et al., 2014). However, both strategies resulted in limited improvements in acetate yields with an increase from 8.16 g/L to 9.61 g/L or 9.52 g/L for the *pta-ack* and THF-dependent enzymes, respectively, during a 216 h batch culture on H_2/CO_2. During autotrophic batch cultivation in a stirred tank bioreactor, the acetate concentrations increased from 44.7 g/L to 50.5 g/L and 51 g/L with *pta-ack* and THF-dependent enzymes overexpression, respectively (Straub et al., 2014). While the cell-specific acetate productivity was not changed by THF-dependent enzymes overexpression, the increase in acetate observed with *pta-ack* overexpression corresponded to an increase from 20.54 $g.g^{-1}.d^{-1}$ to 21.7 $g.g^{-1}.d^{-1}$. The authors hypothesized that higher acetate concentrations might inhibit the acetate kinase, preventing any further gain of yield.

Increasing hydrogen partial pressure was shown to result in increased formate production, lowered acetate production, and decreased optical density of *A. woodii* WT (Kantzow et al., 2015). To evaluate the effect of overexpression of the THF-dependent enzymes and *pta-ack*, the same strains (Straub et al., 2014) were grown in autotrophic conditions with increased hydrogen partial pressure. Both strategies did result in a decrease of formate concentrations and a concomitant increase in acetate concentrations and dry cell mass. Acetate concentrations were increased from 45.3 g/L to 50.3 g/L and 47.3 g/L with overexpression the THF enzymes or pta-ack, respectively (Kantzow et al., 2015). Both strains also increased the hydrogen uptake by 11% and 4%, respectively.

Overall, plasmid-based overexpression of either the THF-dependent enzymes or the phosphostranacetylase-acetate kinase resulted in increased carbon flow through the WLP (Kantzow et al., 2015; Straub et al., 2014). However, a maximum of 50 g/L of acetate was reached.

In *C. autoethanogenum*, supplementation with arginine was found to increase ATP yields caused by the expression of an arginine deaminase pathway (Valgepea et al., 2017). The pathway from *C. autoethanogenum* was expressed in *A. woodii* to increase ATP generation and improve its growth (Beck et al., 2020). Strong constitutive expression with the $P_{pta\text{-}ack}$ promoter from *C. ljungdahlii* resulted in poor or no growth in mixotrophic or autotrophic grown cells, potentially because of a metabolic flux imbalance (Beck et al., 2020). On the other hand, inducible expression with the tetracycline promoter resulted in 36% higher optical densities, 52% lower acetate production and more fumarate (1.4 mM compared to 0.1 mM), when grown mixotrophically on fructose and H_2/CO_2 and with arginine as a nitrogen source (Beck et al., 2020). The decrease in acetate linked to the increase in biomass showed that the overexpression of the arginine deaminase pathway is not a realistic avenue for increased yields in *A. woodii*.

These examples demonstrated that while a pathway or a gene is essential for ATP synthesis, their overexpression seldom leads to straightforward growth improvements.

4.1.3 Mutant abundance studies

In 2009, Langridge and co-workers reported that successive passages of a transposon library under a particular condition resulted in a decrease of reads in a set of genes, confirming the essentiality of genes originally identified as ambiguously essential (Langridge et al., 2009). More interestingly, they also reported that the number of reads

increased in another set of genes. As disruption of those genes resulted in better strain fitness, they concluded that those genes are detrimental to growth under the conditions selected. They hypothesized that those genes, as well as not being required for growth, are likely metabolically costly.

A later study in *S. pneumoniae*, demonstrated that 6% of the genotype-phenotype interactions identified with the transposon library were found to be detrimental for growth (van Opijnen & Camilli, 2012). More recently, a study of *E. limosum* growth on methanol revealed that mutants with transposon insertion in one of five genes dominated the library at after four subcultures and they represented 51.58% of reads after the third subculture and 79.59% after the fourth subculture (Redfern, 2021). Study of those detrimental genes can provide new insights on strain evolution and gene loss.

There is potentially a wealth of information underutilised in the transposon library experiments traditionally performed. A more systematic study of the abundance of reads overtime could provide complementary information to essentiality studies by identifying genes detrimental to the growth conditions. Abundance studies could also reinforce the essentiality studies by identifying genes that do not reach the essential threshold but in which insertion mutants are shown to drastically decrease.

Indeed, overrepresented reads identified by quality control (with FastQC) and gene-level studies of abundance cannot provide a full picture of the population composition. While studies of the abundance of reads at the gene level can identify candidate genes, studies at the single mutant level are required, as a single mutant increasing in abundance could have also benefited from other mutations in its genome. Only a consistent behavior of the whole mutant pool containing an insertion in a specified gene can increase confidence on the detrimental effect of that gene. The method could be

applied retrospectively to datasets already generated, it could for example provide candidate genes detrimental to CO growth of *C. autoethanogenum* (Woods et al., 2022) and enable identification of genetic engineering targets for improved fermenter growth.

Discovery of detrimental genes in specific growth conditions represent an exciting new avenue for transposon libraries, as their deletion is more straightforward than the overexpression of essential genes. In *A. woodii*, deletion of identified genes can be achieved with the endogenous CRISPR/Cas genetic tool described in Chapter 3.

4.1.4 Transposon mutagenesis in *A. woodii* (Dr. Jonathan Baker)

The system designed for *A. woodii* was relatively simple, with the *Himar1* C9 transposase under the expression control of the *C. difficile* toxin B promoter (P_{tcdB}) and a chloramphenicol resistance cassette flanked by inverted terminal repeats (ITR) (plasmid pMTL-GL15, Figure 4.1). The plasmid did not contain a Gram-positive replicon, making it segregationally unstable for easy plasmid loss after transposon library generation. For assembly in *E. coli*, the plasmid contained the ColE1 Gram-negative replicon and the erythromycin resistance gene *ermB*.

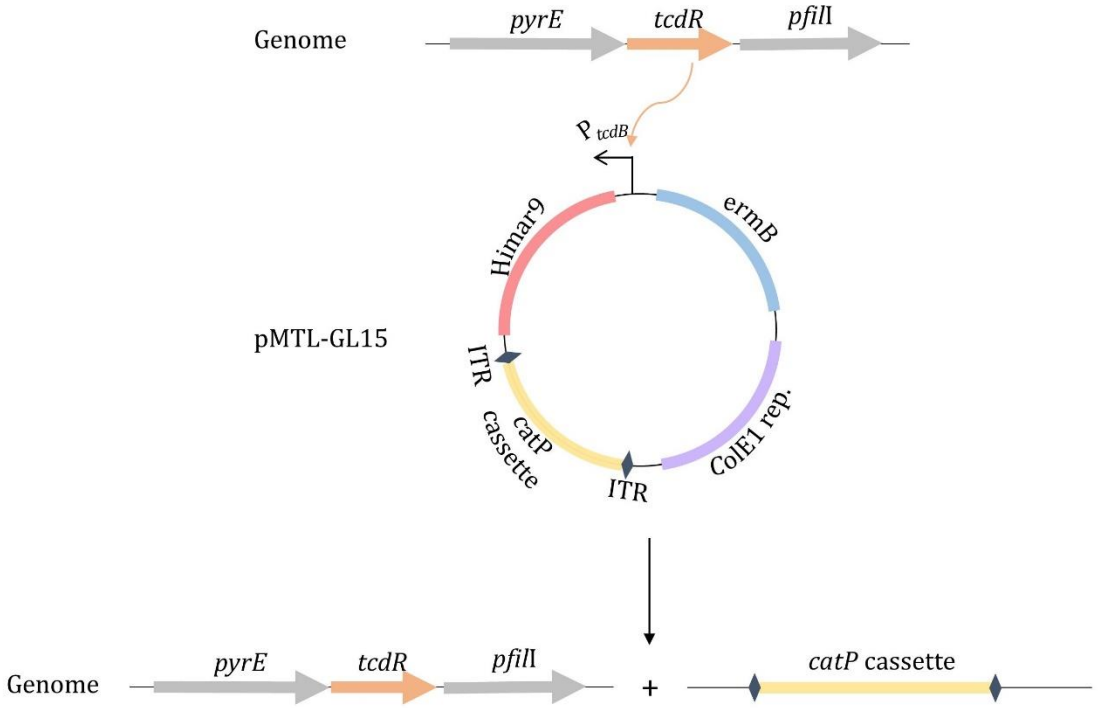


Figure 4-1: Schematic representation of the library construction in *A. woodii* with pMTL-GL15 plasmid.

Transformation of *A. woodii tcdR* with plasmid pMTL-GL15 resulted in a 265-fold increase in thiamphenicol resistant colonies compared to transformation of *A. woodii* WT. This demonstrated successful activation of the P_{tcdB} promoter and expression of the transposase in the presence of TcdR but it also showed a low level of leakage of P_{tcdB} in the absence of TcdR. This leakage was not an obstacle to library construction as high level of transposase expression is required as soon as the plasmid enters the cells.

Generation of $1.7x10^5$ thiamphenicol resistant colonies per µg of plasmid DNA was achieved for each transformation. Six transformations were performed to obtain a one million mutant library.

Inverse PCRs performed on the generated colonies followed by sequencing of the amplicons (14 in total) confirmed that each colony contained a unique and random insertion site. To test for plasmid retention, colonies were restreaked on clarithromycin.

Unsurprisingly, the 80 colonies tested were sensitive to clarithromycin, indicating the absence of the pMTL-GL15 plasmid.

While extensive conditional replication and plasmid loss strategies had to be put in place to achieve adequate mutant levels and library saturation in *Clostridium* (Liu et al., 2013; Zhang et al., 2015; Dembek et al., 2015; Woods et al., 2022), the transformation efficiencies obtained from suicide vectors were sufficient to produce an adequate library in *A. woodii*. This can be explained by better transformation efficiency but the temperature might have also played a role in this apparent difference. *A. woodii* is routinely cultured at 30ºC while *C. autoethanogenum* is cultured at 37ºC, this temperature difference resulted in approximately 50% decrease in transposase activity when studied *in vitro* (Lampe et al., 1998).

4.1.5 Fermentation

The stresses applied to the cells are entirely different during fermentation and static subcultures. Fermentation entails gas flow and high-speed agitation to maintain anaerobic conditions or provide carbon and energy sources. It can be expected that genes involved in motility or resistance to mechanical stress might have a different impact on strain fitness when grown in fermenter in comparison to static batch growth. Additionally, bioprocesses exclusively rely on fermentation and an effort to replicate those conditions would allow better characterisation of the industrial strains considered.

Fermentation has already been applied to transposon libraries in the literature. For example, *C. autoethanogenum* was grown on CO in batch mode for 360 h (Woods et al., 2022). The authors, however, did not study single mutants abundance over the course of the fermentation and did not compare the essentiality to a baseline performed with static subcultures.

4.1.6 Aims

With a transposon library already constructed in *A. woodii*, the aim of this study was to establish a new method of analysis by monitoring the evolution of mutant abundance over time in long-term continuous fermentation as a complementary strategy to traditional essentiality studies. Subsequent identification of genes with reads decreasing and increasing in abundance would validate the method, demonstrating either an essential or detrimental status of genes, respectively.

For this, fermentation of the *A. woodii* transposon library was performed in continuous mode with fructose as the carbon source. It was hypothesized that this mode of fermentation would allow for rapid positive and negative selection of strains with increased and decreased fitness, respectively. Additionally, samples were collected from static growth of the library on fructose to constitute a baseline to compare the two methods. They also allowed to potentially decipher genes essential specifically to fermenter growth.

Another fermentation was also performed in which the carbon and energy source was H_2/CO_2, a condition more appropriate to the industrial processes in operation, with LanzaTech for instance. Sequencing and comparison of the final composition of the mutant library after heterotrophic and autotrophic fermentations enabled the identification and comparison of genes identified as essential or detrimental under the two different fermentation conditions.

The ultimate aim of this study was to decipher the genetic basis of heterotrophic and autotrophic continuous fermentation of *A. woodii* and to provide genetic engineering candidates for improvement of *A. woodii* as an industrially relevant fermentation strain.

4.2 Results and discussion

4.2.1 Continuous fermentation on fructose or H_2/CO_2

Continuous fermentations were performed with either fructose or H_2/CO_2 as carbon source for 356.0 h and 448.8 h, respectively. Samples taken at regular intervals enabled growth to be monitored through determination of the optical density at 600 nm (OD_{600nm}) and metabolite analysis by HPLC (Figure 4.2).

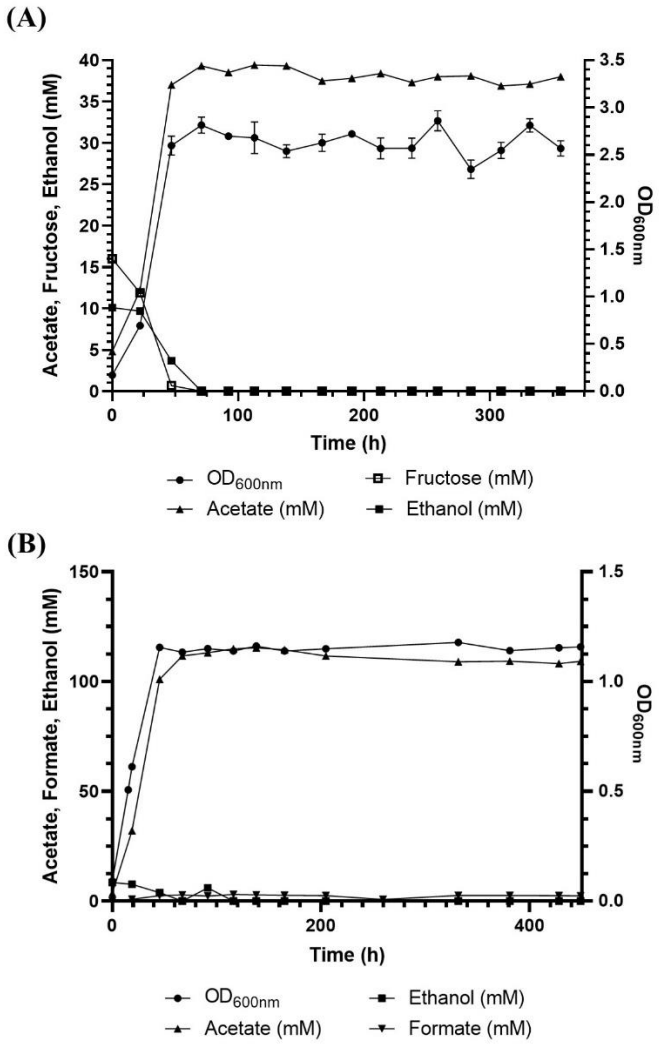


Figure 4-2: Growth and metabolites profiles during (A) heterotrophic and (B) autotrophic (H_2/CO_2) continuous fermentations. $OD_{600\ nm}$ (●), acetate (▲), fructose (□), ethanol (■), and formate (▼) are represented over the course of the fermentations.

In the fructose fermenter, continuous mode was initiated at 22.8 h. After 47.0 h, the amount of fructose present was 0.7 mM with none detected at 70.9 h, indicating complete consumption by the growing cells. From 47.0 h, the OD_{600nm} of the fermenter stabilized at 2.638 ± 0.131 at which point the acetate concentration was 38.04 ± 0.83 mM. The observed high OD_{600nm} variations were a consequence of foaming, which had to be controlled manually through addition of anti-foam. High levels of foam resulted in variations of the dilution, in turn resulting in variations of OD_{600nm}.

In the gas fermenter, continuous mode was started at 20.0 h. No fructose could be detected in the initial HPLC sample, indicating that the initial fructose-based batch mode allowed for complete transition between fructose preculture and gas fermentation. From 45.5 h, the OD_{600nm} of the fermenter stabilized at 1.151 ± 0.012 and acetate concentration stabilized from 67.2h at 110.72 ± 3.92 mM.

The heterotrophic culture achieved a higher OD_{600nm} than when the cells were grown autotrophically, but acetate levels were higher during autotrophic growth. For both fermenters, the acetate and OD_{600nm} remained stable after reaching steady-state, demonstrating that selection of mutants of greater fitness did not lead to an overall increase of growth rate or acetate production.

In both cases, ethanol was detected at the start of the fermentations at concentrations of 10.5 mM and 8.5 mM (for fructose and gas fermenters, respectively) but was not observed after 100 h. Thiamphenicol in 50% ethanol solution was added in the medium of both fermenters, this result showed ethanol consumption by the growing cells in both fermenters (Bertsch et al., 2016). Interestingly, low formate concentrations were also detected during H_2/CO_2 fermentation only. This corroborated the hypothesis that

formate conversion by the four THF-dependent enzymes is a bottleneck for carbon flow through the WLP (Peters et al., 1999).

During steady state, dilution rates of 0.040 ± 0.003 h^{-1} and 0.030 ± 0.002 h^{-1} were employed for the fructose and gas fermenters, respectively. Generation times were calculated based on the dilution rates. The final time points were equivalent to 40.8 and 40.7 generations for the fructose and gas fermenters, respectively. As such, the final time points can be used to compare the two fermenters, for both essentiality studies and evolution of reads counts of specific mutants over the course of the fermentations.

Along with an initial sample from the precultures, samples for Illumina sequencing were selected at regular time intervals. For the fructose fermenter, samples from 47.0 h, 112.9 h, 166.7 h, 258.4 h and 356.0 h of growth were used for sequencing. For the gas fermenter, samples from 45.5 h, 138.0 h, 204.7 h, 331.5 h and 448.8 h of growth were used for sequencing.

4.2.2 Read mapping and essentiality studies

4.2.2.1 Reads mapping and coverage

Four successive sequencing runs were performed to increase the number of reads generated. While the percentage of reads mapping to the *A. woodii* genome varied between samples (98.57% to 99.36% for the static batch cultures, 97.52% to 98.06% for the fructose fermentation and 96.24% to 98.68% for the gas fermentation), an increase in the number of reads resulted in a proportional increase of reads mapped to the genome.

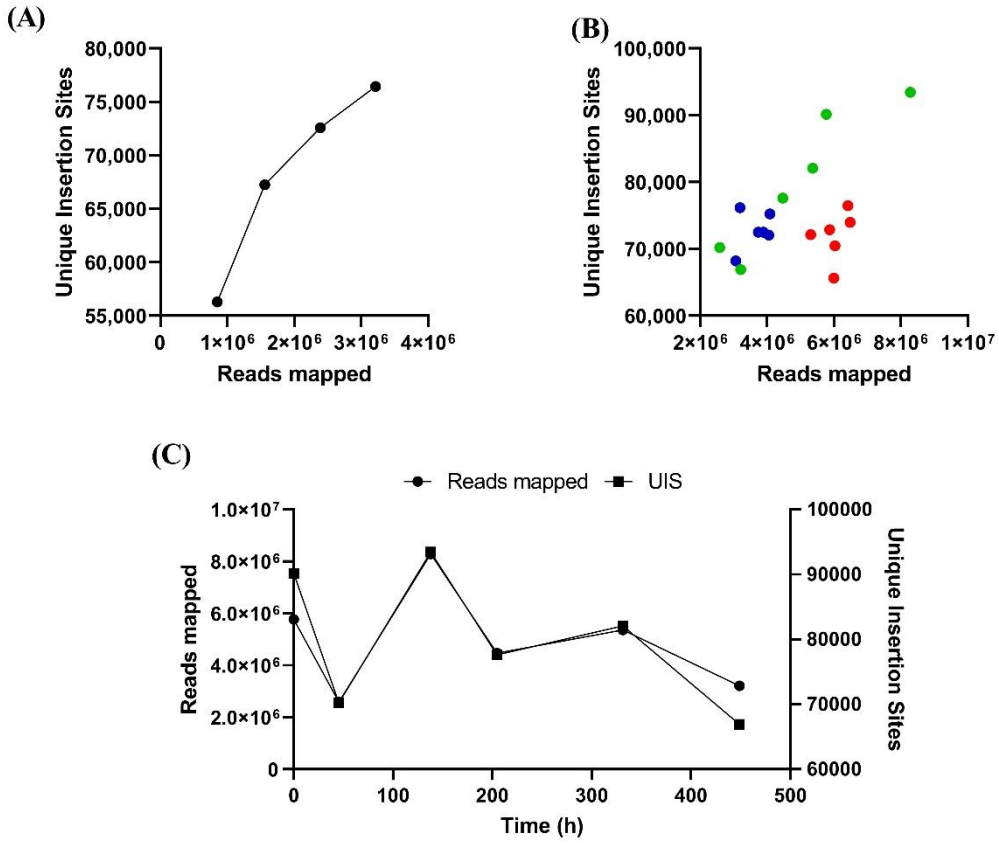


Figure 4-3: Analysis of reads mapping and discovery of unique insertion sites (UIS) for the library. (A) Rarefaction analysis of the TraDIS dataset, UIS number in relation to the number of mapped reads obtained by sequentially concatenating reads from the four runs for the first sample of the fructose fermenter. **(B)** UIS number in relation to the number of mapped reads for all the samples studied for the fructose batch culture (red), the fructose fermenter (blue), and the gas fermenter (green). **(C)** Number of reads mapped and UIS over time during autotrophic fermentation.

Since four successive Illumina sequencing runs were performed on each sample, a rarefaction study was performed to analyse the sequence coverage obtained. For this, the results for one sample were sequentially concatenated and the number of UIS plotted against the total number of reads mapped to the genome (Figure 4.3A). Consecutive sequencing runs resulted in diminishing returns with less new unique insertion sites discovered. The equation obtained ($y = 15{,}089\ln(x) - 149{,}071$) from the graph (Figure 4.3A) was used to predict that only an additional 3,819 UIS could have been discovered by a fifth sequencing run, making its implementation not justifiable.

However, this trend was not verified when comparing the different samples (Figure 4.3B). While an increase of reads mapped correlated with an increase of UIS in the gas fermenter, samples from static batch growth on fructose resulted in a higher number of mapped reads than in the fructose fermenter but a comparable number of UIS. The number of unique insertion sites varied between 65,636 and 76,479 for the static batch cultures samples and between 66,907 and 93,447 for the continuous fermentation samples. The mutant pool complexity was expected to decrease during fermentation but surprisingly, plotting of the number of reads and UIS over time for the gas fermenter (Figure 4.3C) revealed no correlation between the number of UIS and the progression of the fermenter. This demonstrated that the number of mapped reads and UIS was likely mainly dependent on library preparation and sequencing method.

Coverage was calculated by comparing the genome length to the number of UIS. There was an insertion every 52.89 to 61.62 bp for the batch fructose cultures, 53.08 to 59.28 bp for the fructose fermenter, and 43.28 to 60.45 bp for the gas fermenter.

This coverage was comparable to previous transposon libraries reported in *C. difficile* (Dembek et al., 2015) and *C. autoethanogenum* (Woods et al., 2022). After initial growth in rich medium, the number of mutants for transposon libraries in these microorganisms was equivalent to an insertion every 54 bp and 43 bp, respectively. In *E. limosum* the coverage obtained was higher, with one insertion every 25 bp in the initial library (Redfern, 2021). As expected, subsequent cultures in more restrictive conditions reduced the coverage in both *C. autoethanogenum* and *E. limosum*. In *C. autoethanogenum*, growth in autotrophic conditions reduced the coverage to one insertion every 65 bp. In *E. limosum*, growth subcultures in glucose reduced the coverage to one insertion every 49 bp in the third subculture and growth in methanol

reduced coverage to one insertion every 60 bp and 253 bp for the third and fourth subculture, respectively.

4.2.2.2 Selective pressure and mutant abundance

The diversity of the mutant pool did not appear to decrease based on unique insertion site numbers alone. To evaluate mutant selection between the two subcultures in fructose and during the course of the fermenters, the fold change of abundance of reads in each of *A. woodii* genes was calculated. The $\log_2$ of those values were calculated and their distribution for each of the three conditions is shown in Figure 4.4.

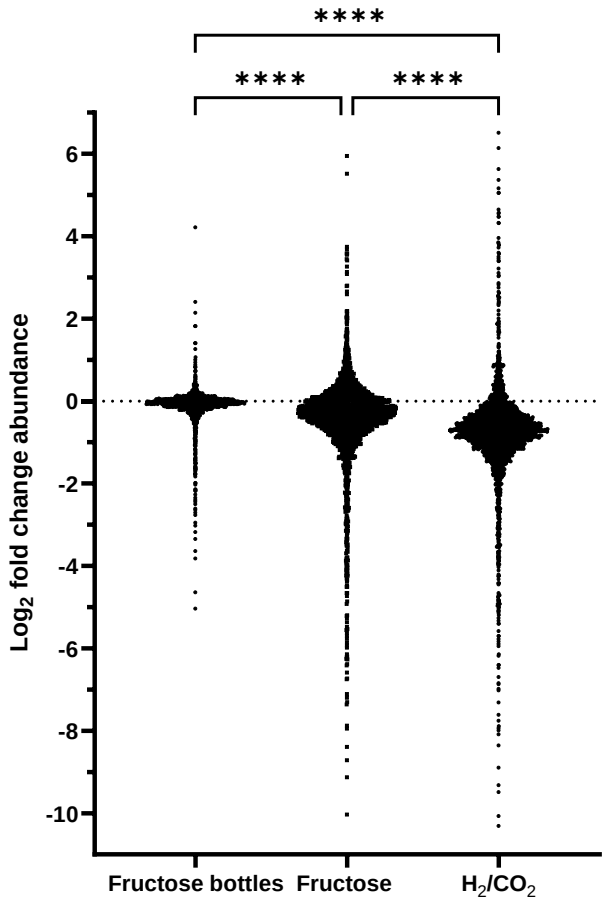


Figure 4-4: $\log_2$ of the fold change of abundance of reads for each *A. woodii* gene between the first and second subculture in fructose and the first and last sample of both the fructose and H_2/CO_2 fermenters. Each data point represents a gene. Statistical analyses to test for identical distributions were done with a Kruskal-Wallis test, ****: $p<0.0001$.

For all the conditions, negative and positive selection is observed. For both fermenters, the distribution is significantly different to the distribution from the static fructose growth. The distribution is also significantly different between the two fermenters. The

mean log_2 fold change is the lowest (mean log_2 fold change = -0.7830) when the library was grown in H_2/CO_2 fermenter, second lowest is when grown in fructose fermenter (mean log_2 fold change = -0.4378), and static bottles have the highest (mean log_2 fold change = -0.1002). Genes with extremely high or low log_2 fold change are more frequent and with more extreme values when *A. woodii* was grown in fermenters.

Despite the lack of a clear trend on the number of UIS between the different samples (section 4.2.2.1), the gene-level change of abundance of reads between the two fructose subcultures and during the fermentations demonstrated the selective pressure in the respective conditions tested. As expected, growth in fermenter increased the selective pressure compared to static batch cultures and autotrophic conditions increased the selective pressure further.

The genes with a negative log_2 fold change have a decreasing number of reads, they represent genes that are essential or important for the corresponding condition. Interestingly, a high number of genes have a positive log_2 fold change signifying an increasing abundance of reads. Reads in those genes become overrepresented in the library population, indicating that mutants with insertion in those genes have an increased fitness in the condition studied. As such, those genes can be hypothesized to be detrimental to growth in the corresponding condition.

As described in the introduction, detrimental genes have been previously reported in *E. limosum* (Redfern, 2021), *S. pneumoniae* (van Opijnen & Camilli, 2012), and *Salmonella* (Langridge et al., 2009). The list of genes with the highest fold change of abundance during heterotrophic and autotrophic fermentations can be found in Supplementary Table 14 and Supplementary Table 15. To mitigate the risk of single

mutants creating a bias at the gene-level, the following study combined analysis of gene-level to single mutant-level read abundance when necessary.

4.2.2.3 Residual library plasmid

Sequencing revealed a read mapping to the library construction plasmid, pMTL-GL15, in all but one sample sequenced. The percentage of this read represents the percentage of mutants in the library that still contained the plasmid (Figure 4.5).

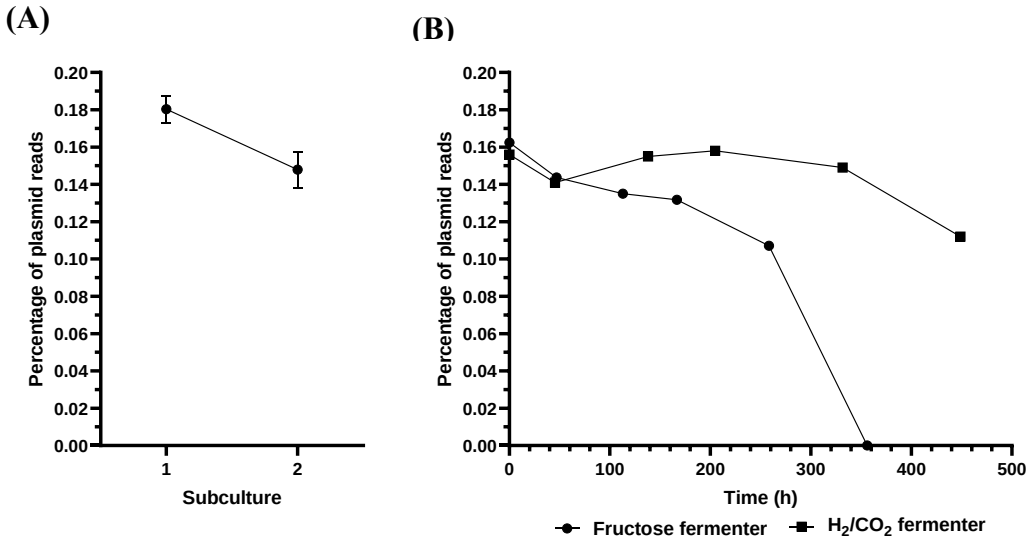


Figure 4-5: Percentage of plasmid-containing mutants during (A) heterotrophic batch fermentations, and (B) heterotrophic (●) or autotrophic (■) continuous fermentations. The mean percentage and standard deviation of plasmid-containing mutants for the first and second subculture in fructose are represented, n=3.

The initial percentage of plasmid-containing mutants was comparable under the different conditions with an average of 0.180% for fructose bottles, and 0.163% or 0.156% for heterotrophic and autotrophic fermenters, respectively. This did not contradict the data obtained during library construction as only 80 colonies were tested for plasmid retention. This is not unprecedented as a comparable proportion (0.2%) of plasmid reads was obtained in equivalent experiments with *C. autoethanogenum* (Woods et al., 2022). Antibiotic selection was applied in all the conditions studied to ensure selection of *A. woodii* cells containing the transposon sequence. Plasmid pMTL-

GL15 is a suicide vector, not able to replicate in Gram-positive bacteria, it is thus likely that the whole plasmid sequence was transposed into the genome. This event would enable expression of the *catP* gene at similar levels as when only the transposon cassette is inserted.

Under all the conditions, the proportion of mutants containing the plasmid sequence was not stable over the course of subcultures or during continuous fermentation.

During batch cultures, in bottles with fructose and in the initial phase of both fermenters, the abundance of mutants containing the plasmid appeared to decrease. At the start of the steady-state of the fructose fermenter, the abundance of mutants with the plasmid stabilized before decreasing again and dropping from 0.107% to 0% between 258.4 h and 356 h.

The results for the H_2/CO_2 fermenter were slightly different. At the start of the steady-state, the abundance of mutants with the plasmid slightly increased before decreasing again after 331.5 h, but only from 0.149% to 0.112%. Similar to the insertion of the transposon sequence, insertion of the whole plasmid in the genome is likely to have happened randomly and in a different locus in each instance. The data plotted (Figure 4.5) showed the evolution of abundance of the read mapping to the plasmid regardless of the insertion locus.

Two hypotheses can explain the decrease in abundance of mutants containing the whole plasmid, the low fitness of those mutants or the successful expression of the transposase from the genome. The former would lead to slower growth of plasmid-bearing cells and their faster disappearance from the fermenter compared to cells with the transposon alone. Lower fitness could be caused by the higher burden imposed on the cells with the expression of the additional genes present on the pMTL-GL15 plasmid. The latter would

lead to transposition events of the transposon element into new genomic loci during fermentation. This would in turn lead to the disappearance, not of the plasmid-bearing cells, but of the reads mapping to the plasmid. While the former would have no impact on the essentiality and read abundance analyses, it is important to consider the consequences of the latter. Residual transposition activity during the fermentations would be linked to the emergence of new mutants in the population. This should not interfere with any of the essentiality studies to be carried out as transposition into essential genes cannot occur. However, it would enable the emergence of new insertion in other genes, a factor that needs to be considered when analysing read abundance for single mutants. While low number of reads in the initial sample should be highlighted as it can result in high errors when calculating the corresponding fold change of read abundance, single mutants with no reads detected in the first sample should not be excluded from the analysis. On the contrary, they should be analysed separately to enable quantification of the potential transient transposition phenomenon.

4.2.2.4 Essentiality studies

The insertion index distributions of the second fructose subcultures, as well as the final time point of both the heterotrophic and autotrophic fermenters, followed a bi-modal distribution (Langridge et al., 2009) and allowed for the determination of essentiality cut-offs (Figure 4.6).

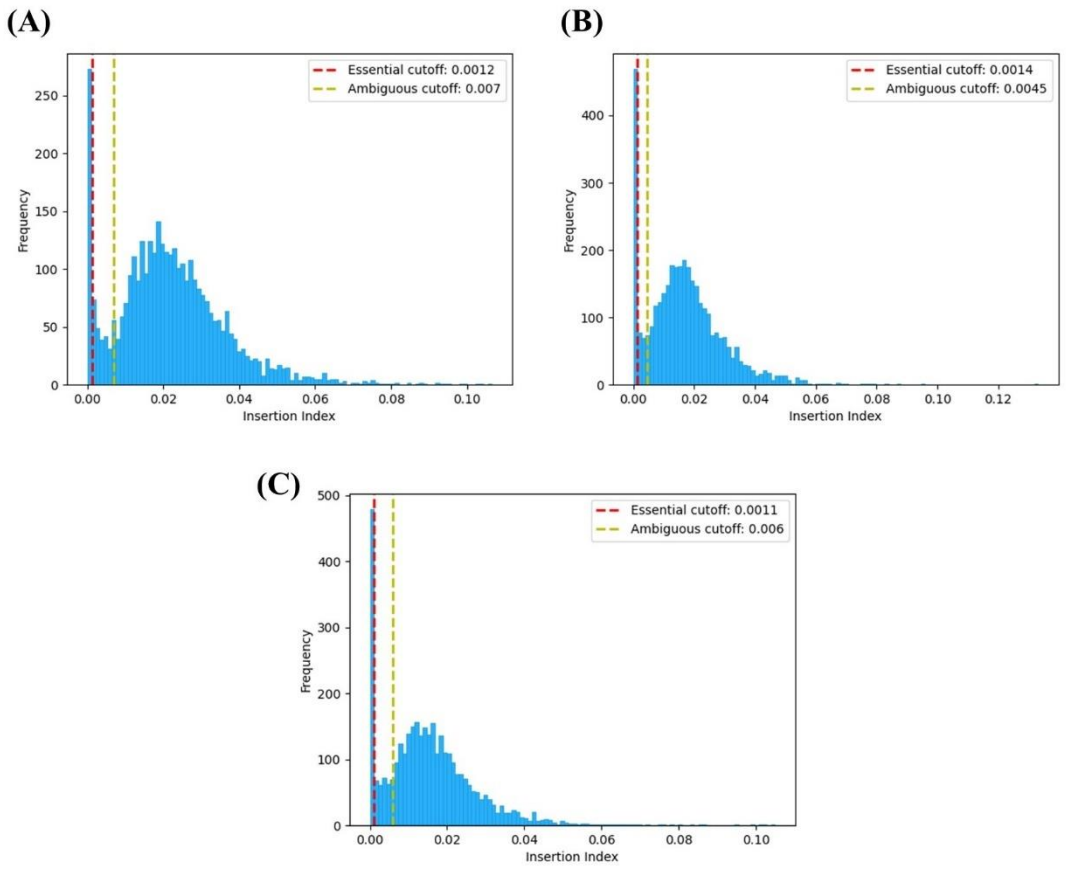


Figure 4-6: Distribution of insertion index for (A) the second fructose subculture, and the final time point of both (B) heterotrophic and (C) autotrophic fermenters. The cut-offs used for determination of essential and ambiguous genes are represented in red and green, respectively.

Core gene essentiality predictions suggested that 282 genes were essential for static batch growth in rich medium, with a further 263 ambiguous under this condition. Continuous heterotrophic fermentation expanded the number of essential genes to 473 and a further 167 were found to be ambiguous. For continuous autotrophic fermentation, the number of essential genes was comparable (487 genes) but the number of ambiguous genes increased to 301.

Gene-level read abundance has shown that growth in fermenter led to greater selective pressure than in static batch bottles and that autotrophic conditions, in turn, exerted even greater selective pressure (section 4.2.2.2). This was in part confirmed by the essentiality study, as the core gene essentiality was mostly shared between all the conditions and

growth in chemostat led to an increase in essential genes (shared between the two fermenters) (Figure 4.7). Nonetheless, it demonstrated that both heterotrophic and autotrophic fermentation conditions led to additional distinct essential gene sets (Figure 4.7). In total, 115 genes were found essential or ambiguous for fermentation compared to static batch culture. An additional 25 essential or ambiguous genes were discovered specifically for fructose fermentation and 150 specifically for H_2/CO_2 fermentation.

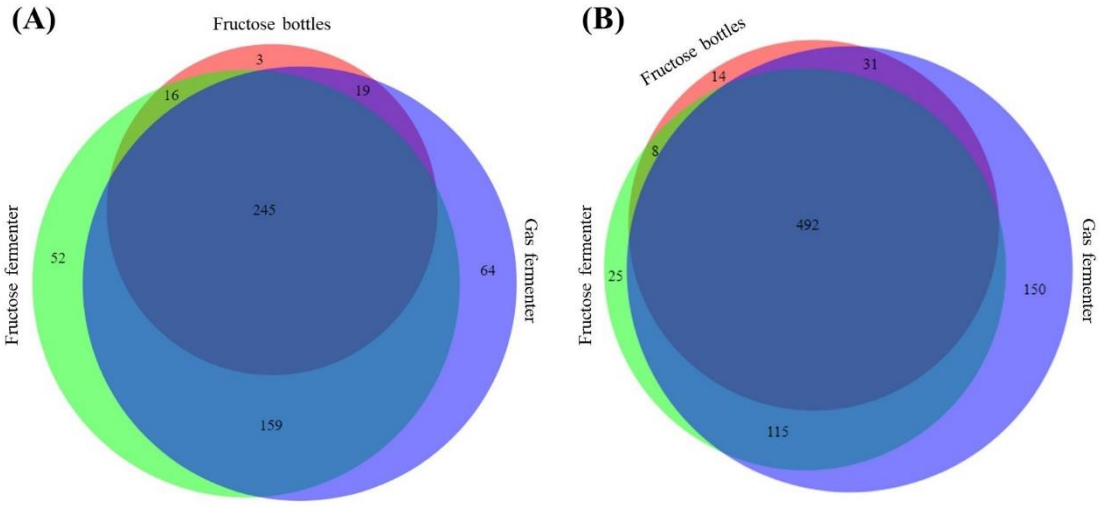


Figure 4-7: Venn diagram of essential genes for the fructose subcultures, fructose fermenter, and gas fermenter, (A) excluding or (B) including ambiguous genes.

Overall, the core gene essentiality for *A. woodii* represented 15.3% of its genes, including those scored as ambiguous. Following the heterotrophic and autotrophic growth in reactors, this number increased to 18.0% and 22.1%, respectively.

The core essentiality of *C. autoethanogenum* (Woods et al., 2022) and *C. difficile* (Dembek et al., 2015) represented 11% of their genome when grown in rich medium. For *C. autoethanogenum*, gene essentiality increased to 19% in autotrophic conditions (Woods et al., 2022). In the case of *E. limosum*, growth was undertaken in a defined medium without added amino acids and only three vitamins. As a result, its elucidated core gene essentiality was higher than for both *C. difficile* and *C. autoethanogenum* at

12.47% of the genome (Redfern, 2021). As expected, gene essentiality increased to 16% after growth on methanol (Redfern, 2021).

4.2.3 Role of *A. woodii* genes under different conditions

The essentiality study was combined to the fold change of abundance of reads at the gene-level to analyse the role of different known genome features of *A. woodii* such as carbon metabolism and energy conservation in the fructose static bottle cultures and both heterotrophic and autotrophic fermenters.

4.2.3.1 Glycolysis / gluconeogenesis, pyruvate metabolism, and tricarboxylic acid (TCA) cycle

As previously mentioned, construction of the library on fructose resulted in a bias of the single mutants. Mutants with insertions in genes essential for fructose growth in static batch conditions were eliminated at an early stage of library construction and their status remained essential throughout the different conditions tested. This limitation needs to be considered, specifically when studying glycolysis pathways.

During heterotrophic growth on fructose, fructose uptake is achieved through a phosphotransferase system (PTS) which transfers a phosphoryl group from phosphoenolpyruvate (PEP) to fructose through the action of enzyme I (Awo_c25710), a phosphocarrier HPr (Awo_c25700), and enzyme II (specific to fructose) encoded by *fruA (*Awo_c03330), resulting in fructose-1-phosphate. The 1-phosphofructokinase *fruK* (Awo_c03340) further converts fructose-1-phosphate into fructose-1,6-phosphate. Substrate-level phosphorylation is subsequently accomplished by the Embden Meyerhof Parnas pathway, which allows production of pyruvate from hexoses with generation of ATP (Figure 4.8).

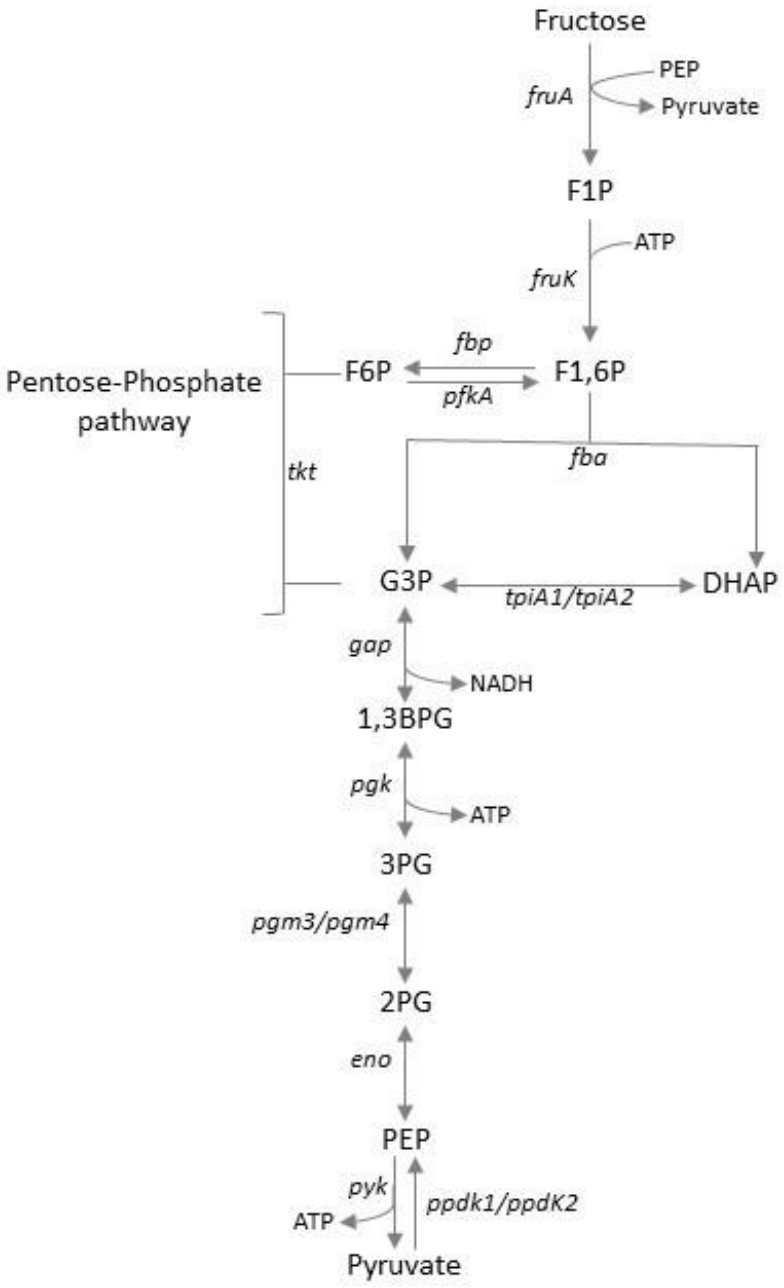


Figure 4-8: Fructose metabolism in *A. woodii* and the genes involved. F1P = fructose-1-phosphate; F1,6P = fructose-1,6-biphosphate; F6P = fructose-6-phosphate; G3P = glyceraldehyde-3-phosphate; DHAP = dihydroxyacetone phosphate; 1,3BPG= 1,3-bisphosphoglycerate; 3PG= 3-phosphoglycerate; 2PG = 2-phosphoglycerate ; PEP = phosphoenolpyruvate.

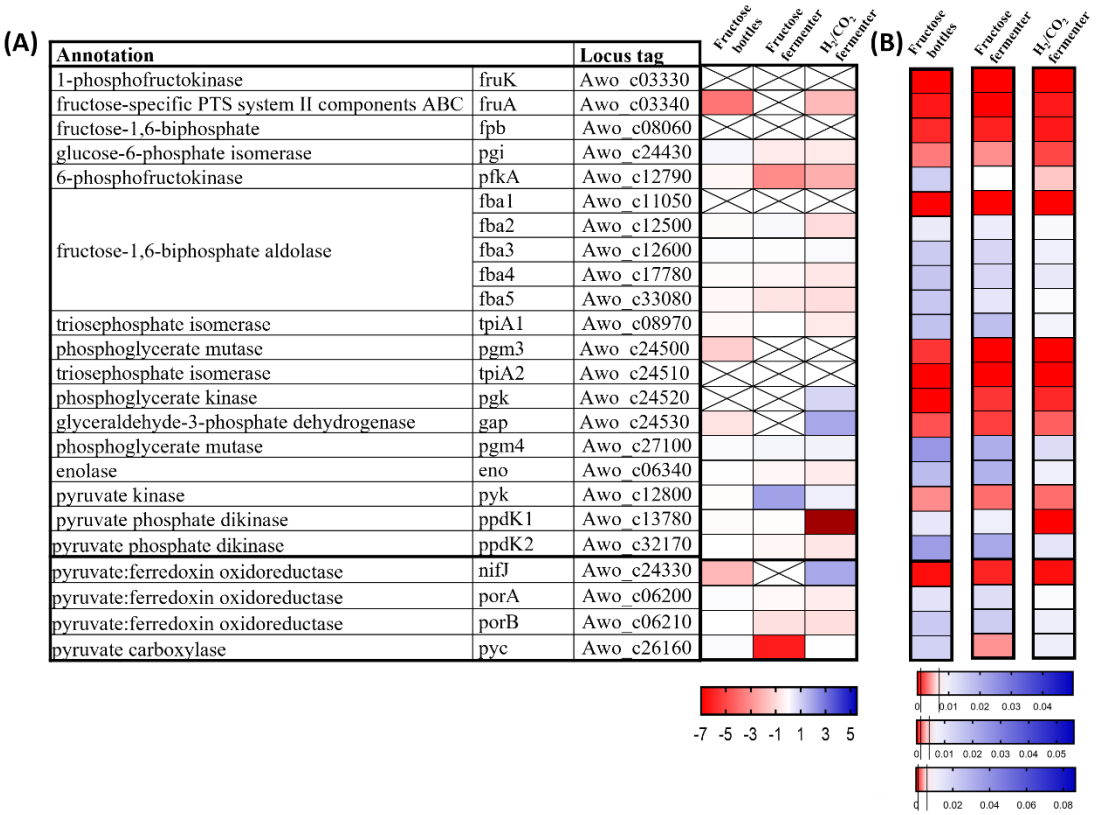


Figure 4-9: Gene-level abundance and essentiality studies of the genes involved in glycolysis/gluconeogenesis and pyruvate metabolism. (A) Gene-level reads abundance, the colour gradient represents the $\log_2$ of the fold change of reads abundance at the gene-level; crosses indicate instances where no fold change could be computed because of a null value; brown represents a value outside of the gradient range. **(B)** Essentiality study with independent colour gradient for each condition, lowest value in red, central value is the ambiguous cut-off determined for each condition. Both the essential and ambiguous cut-offs are represented on the colour gradients.

As expected, genes of the glycolysis pathway were found to be essential during growth on fructose (Figure 4.9). Some genes displayed redundancy and this study enabled the elucidation of the genes involved in glycolysis, as discussed hereafter. Only the fructose-1,6-biphosphate aldolase encoded by *fba1* (Awo_c11050) was shown to be essential, suggesting that the other encoded aldolases might be specific to other sugars. Similarly, only the triosephosphate isomerase *tpiA2* (Awo_c24510) and the phosphoglycerate mutase *pgm3* (Awo_c24500) appeared essential, showing their prominent role in glycolysis in comparison to their homologs *tpiA1* and *pgm4*. The genes *tpiA2* and *pgm3*

form a gene cluster with *pgk* and *gap*, corroborating their simultaneous expression during glycolysis. Additionally, the transketolase encoded by *tktA1* and *tktB1* was also found to be essential in all the conditions tested while gene pairs number 2, 3, and 4 were non-essential in all the conditions. All these redundancies are corroborated by transcription and translation studies, which showed that genes found to be essential here were relatively more highly expressed (Shin et al., 2021).

Unexpectedly, the enolase (Awo_c06340) appeared non-essential, its absence would disable ATP production from fructose. Study of the position of transposon insertions (Figure 4-10) revealed that a third of insertions were in the first 17 amino acids of the annotated gene and no insertion was detected between amino acid 18 and 245. In other *Acetobacterium* bacteria, the gene annotation starts at amino acid 18 (methionine). This implied that only the domain contained between amino acid 18 and 245 is essential for heterotrophic growth of *A. woodii*..

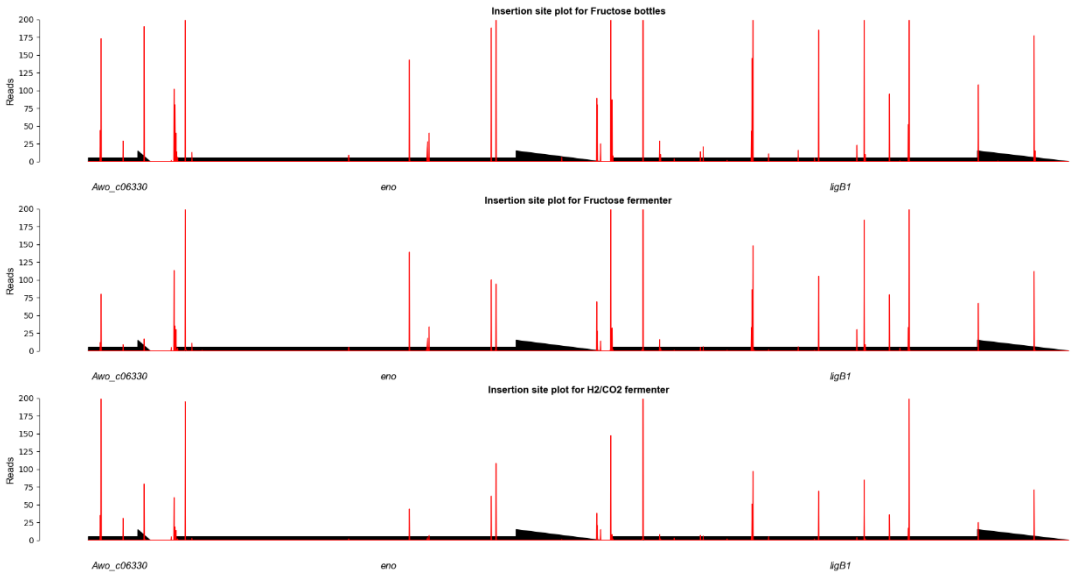


Figure 4-10: Insertion plots in the enolase gene for the second fructose subculture and final time point of both fermenters. Plots were capped at 200 reads to improve resolution of insertion sites with lower read counts.

Pyruvate is at the centre of the metabolism in acetogens. The decarboxylation of pyruvate into acetyl-CoA and CO_2 links substrate-level phosphorylation to the WLP for acetate and ATP generation (Figure 4.10). The reverse reaction, carboxylation of acetyl-CoA into pyruvate, links the WLP to gluconeogenesis for synthesis of cellular macromolecules (Furdui & Ragsdale, 2000). Acetyl-CoA carboxylation is possible through three different systems: pyruvate:ferredoxin oxidoreductases (PFOR, genes *nifJ*, *porA*, and *porB*), pyruvate formate lyases (PFL, genes *pflA, pflB1, pflB2, pflB3*, and *pflC*), and pyruvate dehydrogenases (PDH, genes *pdhC3*, *pdhB3*, and *pdhA3*). The PFOR route has been shown to be the active route for pyruvate synthesis in acetogens (Furdui & Ragsdale, 2000). In *A. woodii*, the PFOR (gene *nifJ*) displayed the highest transcription and translation levels across the three potential pyruvate synthesis routes in cells grown in fructose and H_2/CO_2 (Moon et al., 2021; Neuendorf et al., 2021; Shin et al., 2021).

This is corroborated by the observations made here where, of all the genes listed above, only *nifJ* was found to be essential in the fructose static bottles and in both fermenters. While it has been hypothesized that the knock-out of *nifJ* would allow for pyruvate formation via the PFL route (Neuendorf et al., 2021), the data presented here demonstrated that the other pyruvate synthesis routes cannot compensate for pyruvate synthesis in the absence of *nifJ.*

During autotrophic growth, gluconeogenesis, and the tricarboxylic acid cycle (TCA) cycle, have been shown to be crucial for synthesis of the key molecules in acetogens, including nucleic acids and amino acids (Berg et al., 2010; Song et al., 2018). This was demonstrated by a dramatic decrease in abundance during fermentation, with a $\log_2$ fold change of -8.90 of the total reads in gene *ppdK1*. Gene *ppdK1* was found to be essential in the gas fermenter only, it encodes a pyruvate phosphate dikinase responsible for the

interconversion of pyruvate into phosphoenolpyruvate (PEP) and constitutes the first step of gluconeogenesis from pyruvate. A transcriptional regulator (Awo_c13760) and a phosphotransferase (Awo_c13770) are encoded upstream of *ppdK1*, while the two genes overlap, only Awo_c13770 was found to be important in both fermenters. It was found to be essential or ambiguous during heterotrophic or autotrophic fermentation, respectively, as corroborated with fold change of abundance at the gene-level of -6.20 and-5.82. Awo_c13770 likely encodes a bifunctional protein involved in PpdK1 regulation by catalyzing its phosphorylation/dephosphorylation, leading to its inactivation or activation, respectively. These observations highlight the importance of activity regulation of PpdK1 to shift between glycolysis and gluconeogenesis during fermentation.

The TCA cycle is crucial for production of amino acids, in acetogens it is incomplete, forming a branched cycle (Song et al., 2018). In *A. woodii*, several metabolic steps do not correspond to any genes annotated in the genome (Figure 4.10). However, a recent modelling *A. woodii* metabolism identified candidate genes for the unassigned steps through automated enzyme detection with the *in silico* PRIAM tool (Claudel-Renard et al., 2003; Mesfin, 2020). Briefly, a citrate synthase (Awo_c19610) converts oxaloacetate and acetyl-CoA into CoA and citrate, an aconitate hydratase (*acnA*) citrate to isocitrate, and an isocitrate dehydratase (*icd*) isocitrate to 2-oxoglutarate (Figure 4.10). An aspartate aminotransferase (*aspB1*, *aspB2*) converts oxaloacetate and glutamate into aspartate and 2-oxoglutarate. On the other "branch", oxaloacetate is converted to malate by a malate dehydrogenase (*mdh*), and malate is further converted to fumarate by a fumarate hydratase (*fumA*, *fumB*). While genes Awo_c19610, *acnA*, and *icd* were found to be essential under all the conditions tested, *aspB2* was essential during fermentation

only and *fumB* was ambiguous in the gas fermenter. Genes *mdh*, *fumA*, and *aspB1* were non-essential under all conditions.

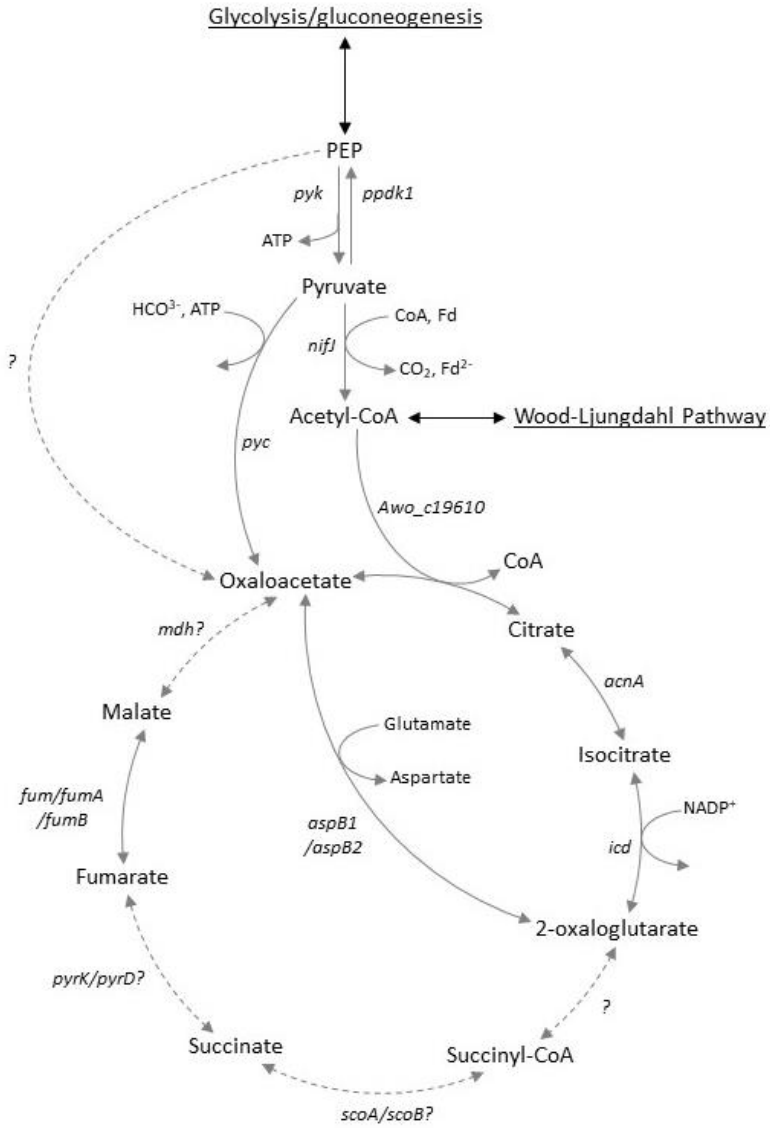


Figure 4-11: Classical tricarboxylic acid (TCA) cycle as identified in *A. woodii* and with the metabolic reactions linking it to glycolysis/gluconeogenesis and the WLP. Dashed metabolic reactions represent enzymatic reactions that are not verified in *A. woodii*, the locus tags indicated for these steps were obtained by *in silico* prediction (Mesfin, 2020). PEP = phosphoenolpyruvate.

Multiple pathways lead to the TCA cycle. Pyruvate can be directly converted to oxaloacetate though a pyruvate carboxylase (*pyc*) or it can be converted to acetyl-CoA by a pyruvate:ferredoxin oxidoreductase (*nifJ*), the acetyl-CoA is then used by the citrate

synthase (Awo_c19610) to convert oxaloacetate to citrate (Figure 4.10). While *nifJ* and Awo_c19610 were found to be essential under all the conditions tested, *pyc* was only found to be important during heterotrophic fermentation and the deletion of *pyk* was found to be favourable for strain fitness (abundance of mutants with insertion in *pyk* increased 4-fold during heterotrophic fermentation). Additionally, in *A. woodii*, the deletion of gene *ppdK1* was found to be highly detrimental during autotrophic fermentation. These observations led to the hypothesis that *A. woodii* encodes for an enzyme catalysing direct conversion of PEP into oxaloacetate. In theory, a PEP carboxylase (PEPC, EC 4.1.1.31, reaction (1)) or a PEP carboxykinase (PEPCK, reaction (2)) can catalyse the reaction, PEP carboxykinases are subdivided in 3 families depending on the phosphate donor, GTP (EC 4.1.1.32), ATP (EC 4.1.1.49), and PP_i (EC 4.1.1.38) (Sauer & Eikmanns, 2005).

(1) $PEP + HCO_3^- \rightarrow Oxaloacetate + P_i$

(2) $PEP + CO_2 \leftrightarrow Oxaloacetate + phosphate$

While an ATP-PEPCK has been identified in *C. autoethanogenum* and its deletion resulted in impaired growth in autotrophic conditions (Marcellin et al., 2016), an early study of *A. woodii* cell extracts obtained after growth on CO_2 revealed the enzymatic activity of a PP_i-PEPCK (PEP carboxytransphosphorylase) (Eden & Fuchs, 1983). PP_i-PEPCK have been described in a very restricted list of bacteria and have no dedicated InterPro protein family, making them difficult to identify in new microorganisms. Protein BLAST of the *C. autoethanogenum* ATP-PEPCK (U5S013) or of Gram-positive PP_i-PEPCK (*Acidipropionibacterium virtanenii*, A0A344UUE8; *Actinomyces israelii*, A0A5S8WF55; *Micropruina glycogenica*, A0A2N9JK44; *Propionibacterium freudenreichii*, A0A0C5URS1) failed to identify a homolog in *A. woodii*. Additionally, no PEPCK has been identified in *Acetobacterium* or

Clostridium on the InterPro database (family IPR021135). Further study is required for a more in-depth understanding of enzymes involved in the PEP-pyruvate-oxaloacetate interconversion in *A. woodii*.

Genes for glutamate to L-glutamine conversion (*glnA1*), L-glutamine to carbamoyl-phosphate conversion (*carA1, carB1*) and to D-glucosamine-6-phosphate (*glnS*) were ambiguous or essential under all the conditions tested. Additionally, the genes for UMP synthesis from carbamoyl-phosphate (*purC* was essential in the fructose fermenter only) and IMP synthesis from L-glutamine were found to be essential only during fermentation. In contrast, the uracil phosphoribosyltransferase (*upp*) for UMP synthesis from uracil was essential under all conditions tested. In *S. pneumoniae*, it was hypothesized that the *de novo* pyrimidine pathway was essential even in the presence of uracil in the medium because disruption of genes in the pathway leads to accumulation of a toxic intermediate (van Opijnen & Camilli, 2012). They also observed that the pyrimidine pathway was transcriptionally activated in specific growth conditions, rendering the genes of the pathway essential only in those conditions. In *A. woodii*, it is possible that a similar regulatory system exists as the growth medium contained yeast extract. The levels contained in the medium might have not been sufficient during fermentation.

4.2.3.2 Wood-Ljungdahl Pathway, energy conservation and electron bifurcation

The Wood-Ljungdahl pathway (WLP) allows production of acetyl-CoA by fixation of CO_2, acetyl-CoA is further converted into acetate by the combined action of a phosphotransacetylase and acetate kinase. However, it does not result in a net ATP gain. Energy conservation is achieved by an F_1F_o-ATPase which relies on a transmembrane ion gradient created by the Rnf complex (Rhodobacter nitrogen fixation). Figure 4.11

presents genes of the reductive acetyl-CoA pathway and associated energy conservation mechanisms and redox balancing elements in *A. woodii* metabolism and Figure 4.12 presents the results of the combined studies of essentiality and mutant abundance obtained for the genes involved. Analyses are detailed in the sections below.

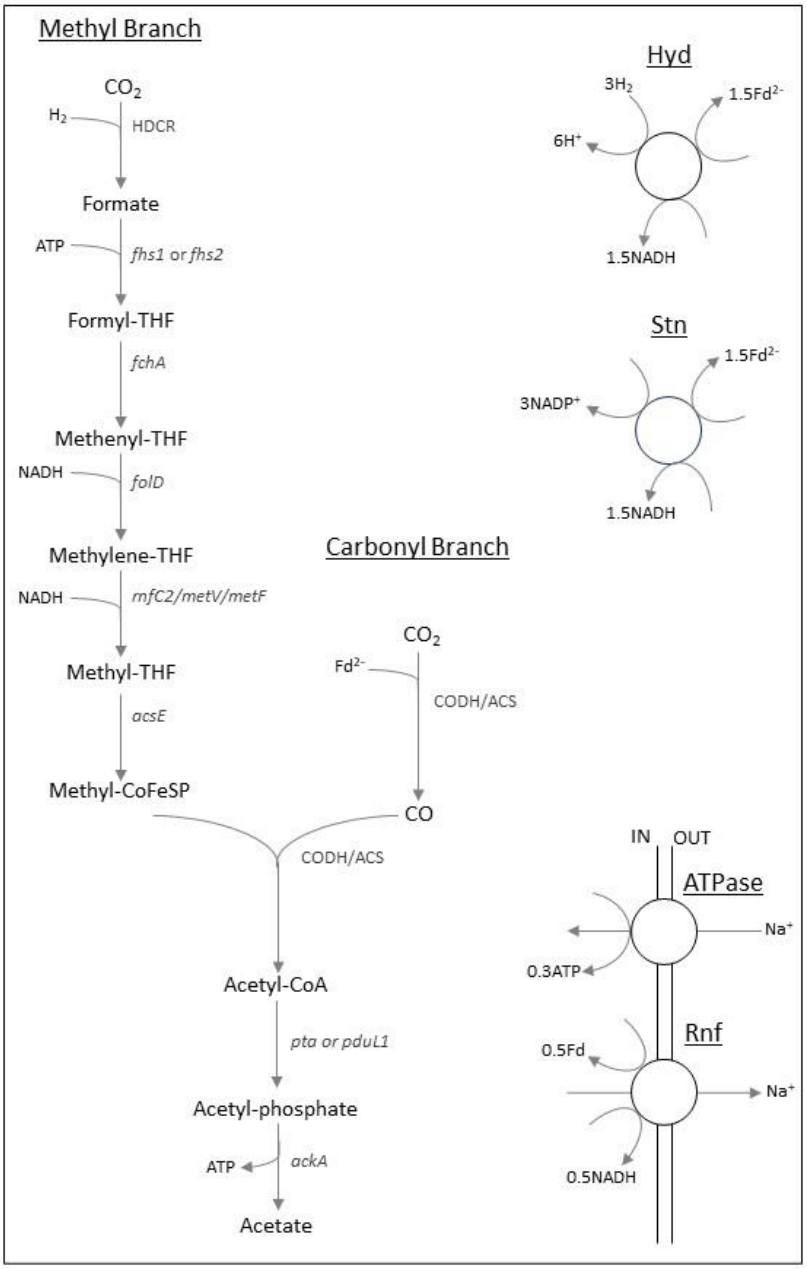


Figure 4-12: Bioenergetics of acetogenesis from H_2 + CO_2 in *A. woodii.* The Stn complex pictured here is based on the Nfn complex of *Sporomusa ovata* (Kremp et al., 2020). See the following sections for further explanations. HDCR: H_2-dependent CO_2 reductase, genes Awo_c08190 to Awo_c08260; CODH/ACS: carbon monoxide dehydrogenase/acetyl-CoA synthase, genes Awo_c10670 to Awo_c10760.

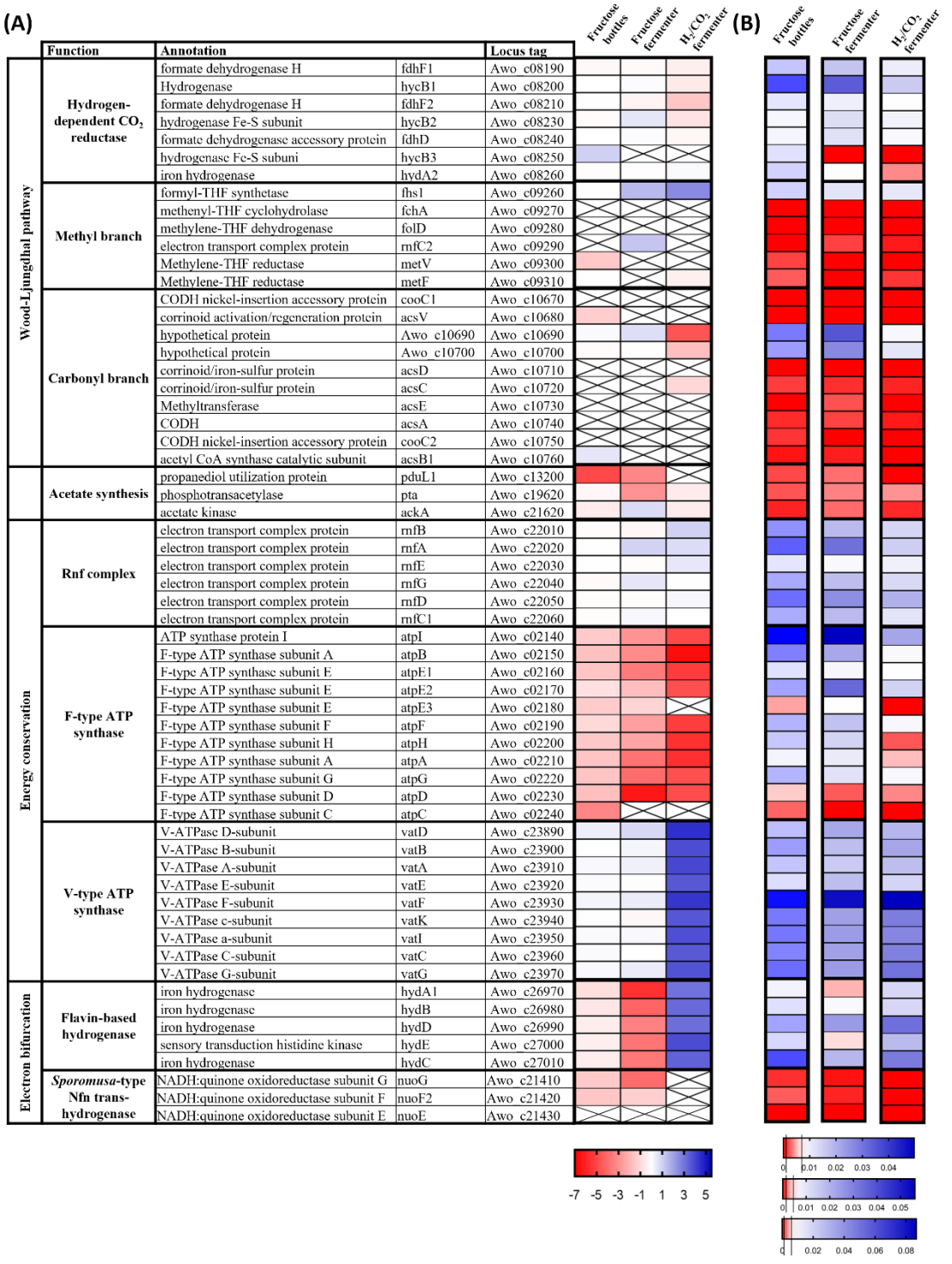


Figure 4-13: Gene-level abundance and essentiality studies of the genes involved in the WLP, energy conservation, and electron bifurcation. (A) Gene-level reads abundance, the colour gradient represents the $\log_2$ of the fold change of reads abundance at the gene-level crosses; indicate instances where no fold change could be computed because of a null value. **(B)** Essentiality study with independent colour gradient for each condition, lowest value in red, central value is the ambiguous cut-off determined for each condition. Both the essential and ambiguous cut-offs are represented on the colour gradients.

4.2.3.2.1 Wood-Ljungdahl Pathway

i) HDCR gene cluster

The HDCR gene cluster constitutes the first step in the methyl branch. It allows for conversion of CO_2 into formate with the concomitant oxidation of H_2 into H^+ (Ceccaldi et al., 2017) and comprises 7 genes (Awo_c08190 – Awo_c08250).

Only *hycB3* was essential during both fermentations and *hydA2* was ambiguous during autotrophic fermentation only (Figure 4.12; Figure 4.13) The *hycB1* gene had the highest insertion index, as corroborated by its absence in the proteome for both fructose and H_2/CO_2 growth (Neuendorf et al., 2021).

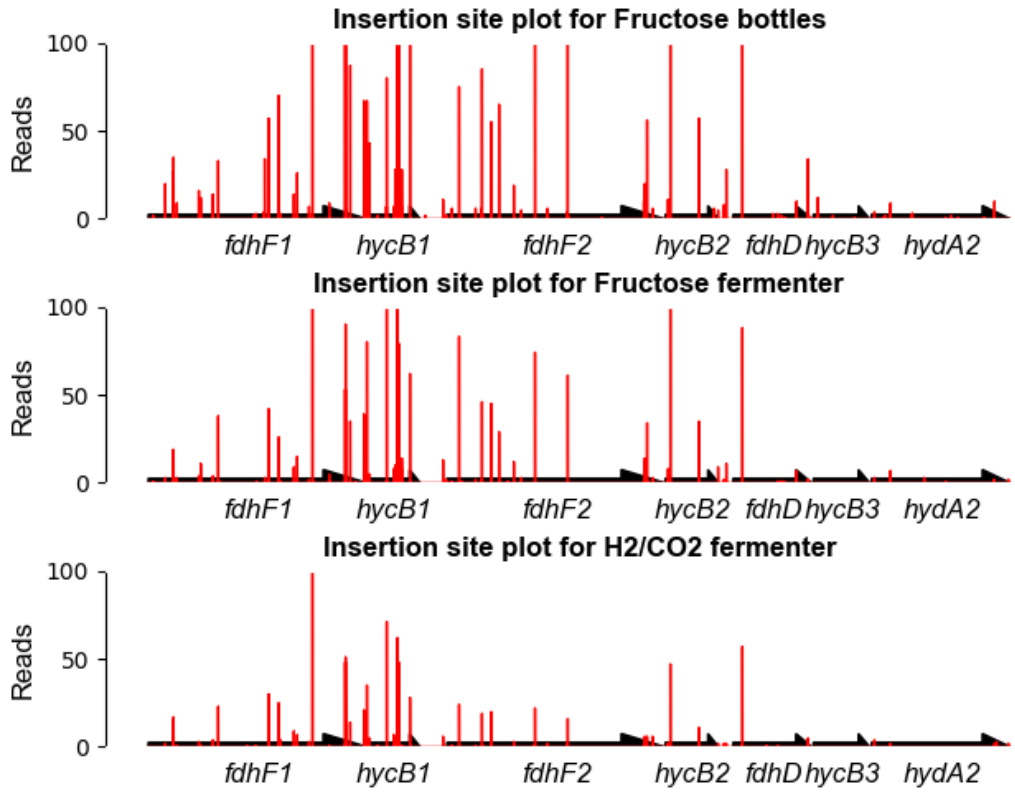


Figure 4-14: Insertion plots in genes of the HDCR for the second fructose subculture and final time point of both fermenters. Plots were capped at 100 reads to improve resolution of insertion sites with lower read counts.

The HDCR complex is formed of two connected catalytic subunits, HydA2 and FdhF2 functioning as a hydrogenase and a formate dehydrogenase, respectively (Ceccaldi et

al., 2017). Genes *fdhF2/hycB2* form the formate dehydrogenase domain (Ceccaldi et al., 2017). Gene *fdhF1* is also annotated as a formate dehydrogenase and the cluster contains a gene annotated *hycB1*. Although gene pair *fdhF2/hycB2* displayed high transcription in translation in both heterotrophic and autotrophic condition (Shin et al., 2021), confirming its primary role as formate dehydrogenase, it is likely that genes *fdhF1* and *hycB1* can compensate for the loss of function of *FdhF2* and *hycB2*, respectively. Explaining the non-essentiality of the formate dehydrogenase catalytic subunit. The accessory protein FdhD did not appear essential at the gene level (Figure 4-13) but study of single insertion mutants (Figure 4-14) suggested high prevalence of mutants with transposon insertions at the start of the encoding gene. This suggested the presence of a downstream start codon, conservation of the protein function, and its potential essentiality growth.

Addition of formate restored the growth of a Δ*hdcr A. woodii* strain on fructose or H_2/CO_2 (Moon et al., 2023), mutants with transposon insertion in the HDCR gene cluster might have benefited from formate produced by the rest of the transposon library population. However, formate was detected only in the gas fermenter (up to 4.7 mM) and the hydrogenase domain was shown to be important for autotrophic growth, supporting the hypothesis of gene redundancy to explain the non-essentiality of *fdhF2/hycB2*.

Previous studies of HDCR were performed by deletion of genes *fdhF1* to *hydA2* (Moon et al., 2023), deletion of *fdhF2* and *hycB2* only would confirm their redundancy. Comparison of transcription levels of *fdhF1* and *hycB1* in the WT and in a Δ*fdhF2/hycB2* strain would confirm their role.

ii) Methyl branch gene cluster

The methyl branch comprises 6 genes (Awo_c09260 – Awo_c09310) encoding enzymes that catalyse the conversion of formate into methyl-THF. The essentiality study revealed that all but one gene from this cluster were essential in both fermenters, only *metV* and *metF* remained ambiguous in the fructose batch growth. The exception is *fhs1* (Aw_c09260) which was found to be non-essential in all the conditions tested (Figure 4.12).

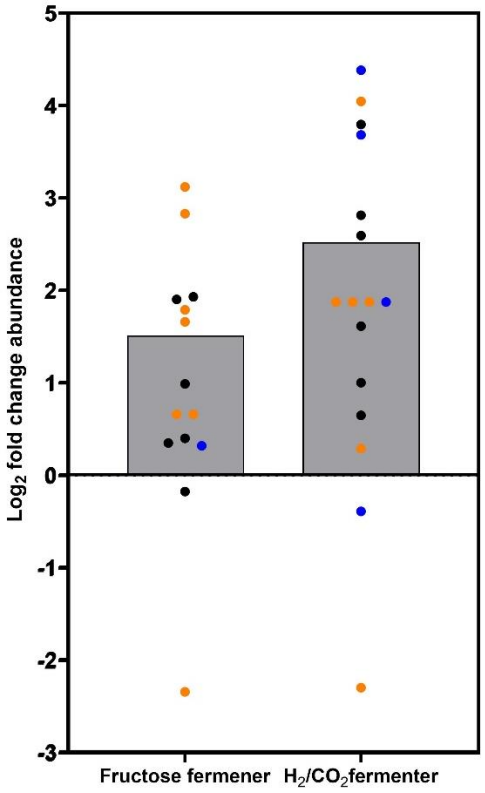


Figure 4-15: Single mutants abundance in gene *fhs1* in both fermenters. Bars represent the gene-level log_2 fold change of reads abundance. The log_2 of the fold change of abundance of individual mutants in the fhs1 gene are represented as individual points. Mutants with a starting number of reads ≤10 and ≤20 are represented in orange and blue respectively, mutants starting with a number of reads >20 are represented in black.

Mutants with insertion in the formyl THF synthase *fhs1* (Awo_c09260) presented a 2.85- and 5.72-fold change of abundance in the heterotrophic and autotrophic fermenter, respectively (Figure 4.14). This demonstrated better fitness of mutants with a disrupted *fhs1*. One study found that its transcription increased 6.2-times between heterotrophic and autotrophic growth (Moon et al., 2021), while another study only observed a 1.96-

times increase and a relatively stable (1.05-fold increase) translation abundance (Shin et al., 2021).

The genome of *A. woodii* contains another formyl-THF-synthetase, *fhs2* (Awo_c08040) commonly associated with the formate metabolism. Gene *fhs2* was found to be non-essential in all the conditions tested and mutants with insertions in its encoding sequence were found to slightly decrease during autotrophic fermentation with a $\log_2$ fold change of -0.90.

While *fhs1* is in the gene cluster of the methyl branch of the WLP it appeared likely that *fhs2* can compensate for its disruption and benefit autotrophic fermentation. Generation of knock-outs of *fhs1* and *fhs2* separately or together would confirm this conclusion. In single deletion mutants, change in expression of the other gene could be studied to confirm redundancy. Subsequent study of the phenotype of the single deletion strains compared to the double deletion strain on formate or H_2/CO_2 would further validate this result.

iii) Carbonyl branch gene cluster:

The carbonyl branch of the WLP comprises 8 genes (Awo_c10670 – Awo_c10680 and Awo_c10710 – Awo_c10760) and is responsible for the conversion of methyl-THF into acetyl-CoA. Unsurprisingly, this gene cluster was found to be essential under all the growth conditions tested. Genes Awo_c10690 and Awo_c10700 encode hypothetical proteins and are situated in middle of the carbonyl branch gene cluster. Both were found to be non-essential under the three conditions tested. Abundance of mutants with insertions in Awo_c10690 and Awo_c10700 varied very little with a fold change of 1.65 and 0.98, respectively, in the fructose fermenter (Figure 4.15). In contrast, mutants with

insertions in Awo_c10690 or Awo_c10700 had low fitness in the H_2/CO_2 fermenter, with a fold change of 0.04 (Awo_c10690) and 0.29 (Awo_c10700) (Figure 4.15).

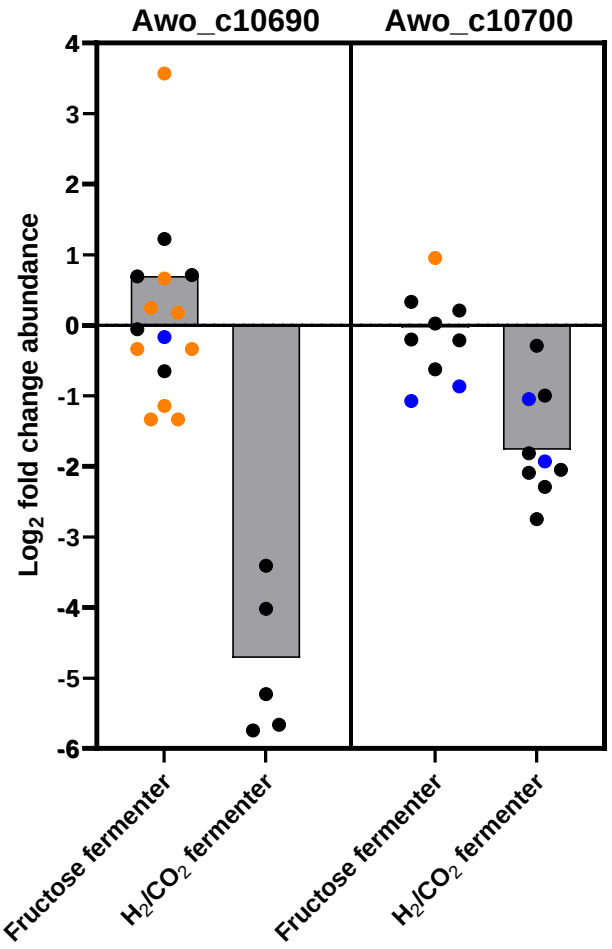


Figure 4-16: Single mutants abundance in genes Awo_c10690 and Awo_c10700 in both fermenters. Bars represent the gene-level log_2 fold change of reads abundance. The log_2 of the fold change of abundance of individual mutants in both genes are represented as individual values. Mutants with a starting number of reads ≤10 and ≤20 are represented in orange and blue respectively, mutants starting with a number of reads >20 are represented in black.

While Awo_c10690 and Awo_c1070 are not described as part of the carbonyl branch of the WLP, the two genes are conserved across the *Acetobacterium* genus (Ross et al., 2020). Additionally, previous studies have shown increase in their transcription between heterotrophic and autotrophic growth (Moon et al., 2021; Shin et al., 2021). Taken together, these observations and the results presented here confirm the likely role of Awo_c10690 or Awo_c10700 in autotrophic fermentation.

iv) Acetate synthesis

In *A. woodii*, ATP is generated by conversion of acetyl-CoA into acetate through a phosphate acetyltransferase (acetyl-CoA to acetyl-P) followed by an acetate kinase (acetyl-P to acetate with ATP generation). Both *pduL1* (Awo_c13200) and *pta* (Awo_c19620) are annotated as phosphate acetyltransferases, while gene *ackA* (Awo_c21620) is annotated as the acetate kinase. The three genes were found to be alternatively ambiguous or essential over the course of the fructose fermentation. The *ackA* gene was found to be essential in both second fructose subculture and the gas fermenter.

While the *pduL1* (Awo_c13200) gene is assigned as catalysing the acetyl-CoA to acetyl phosphate conversion. Gene *pta* (Awo_c19620) has no assigned function in the KEGG database. In the literature, however, either gene have been studied in transcriptomic and proteomic analyses of *A. woodii* metabolism (Moon et al., 2021; Neuendorf et al., 2021; Shin et al., 2021)(Neuendorf et al., 2021; Moon et al., 2021; Shin et al., 2021). In 2021, Neuendorf and co-workers studied the gene locus Awo_c13200 (*pduL1*) incorrectly annotated as "*pta*" (Neuendorf et al., 2021). In the same year, Shin and co-workers studied translational control of acetogenesis and Moon and co-workers studied formate metabolism, both focused on gene *pta* (Awo_c19620) (Moon et al., 2021; Shin et al., 2021).

Both genes were analysed in the present study (Figure 4.12). The gene *pduL1* was found essential in the gas fermenter and ambiguous in the fructose subculture, mutants with insertion in its encoding sequence were highly selected against in both the fructose static bottles and the fermenter, with fold changes of 0.03 and 0.10, respectively. In contrast, *pta* (Awo_c19620) was ambiguous under all the conditions tested with only changes of 0.68- and 0.84-fold changes in abundance in the gas fermenter and fructose static

bottles, respectively. During fructose fermentation, mutants with insertions in *pta* were selected against with a 0.13-fold change, a comparable value to the fold change of *pduL1* insertion mutant under the same condition. mRNA levels and RPF (ribosome-protected fragments) levels of *pduL1* were 6 to 10 times those of *pta* during autotrophic growth (Shinet al., 2021). Additionally, while deletion of *ackA* was deleterious to *A. woodii* growth, deletion of *pta* (Awo_c19620) seemingly had no impact on growth (Dr. Jonathan Baker, personal communications). The enzyme encoded by *pduL1* is likely the enzyme responsible for acetyl-CoA to acetyl phosphate conversion in *A. woodii*.

In *Moorella thermoacetica*, no *pta* gene is annotated but two functional phosphotransacetylase genes, *pduL1* (Moth_0864) and *pduL2* (Moth_1181). Using transcript analysis, Breitkopf and co-workers showed the predominant presence of *pduL2* in cells grown on fructose and concluded that *pduL2* encodes the primary phosphotransacetylase for acetyl-CoA to acetyl phosphate (Breitkopf et al., 2016). These findings were further confirmed in 2017 by Iwasaki and co-workers (Iwasaki et al., 2017) who showed that deletion of both *pduL1* and *pduL2* abolished acetate production entirely. While deletion of *pduL1* had a marginal effect, the *pduL2* deletion significantly decreased acetate production. Although the Pta activity of PduL2 for acetyl-CoA was 20 times higher than that of PduL1 (Iwasaki et al., 2017), analysis of transcription of the encoding genes showed that deletion of one increased transcription of the other, effectively compensating for the deletion.

Study of compensation mechanisms between *pta* and *pduL1* in A. woodii by gene deletion and transcriptional analyses could inform on redundancy of the genes involved.

4.2.3.2.2 Energy conservation

i) Rnf complex

The Rnf complex comprises 6 genes (Awo_c22010 – Awo_c22060), it is membrane-bound and catalyses oxidation of reduced ferredoxin to the reduction of NAD^+ accompanied by sodium ion translocation. The sodium gradient is then used by an ATP synthase for ATP production. As such, the Rnf complex is believed to be responsible for energy conservation during autotrophic growth.

Surprisingly, none of the genes in the cluster appeared to be essential in any of the growth conditions tested (Figure 4.12). At the gene-level, the fold change of read abundance had values of between 0.84 and 1.98. Transposon insertions in the Rnf complex genes did not appear to confer any advantage or disadvantage for growth under the conditions tested (Figure 4.16).

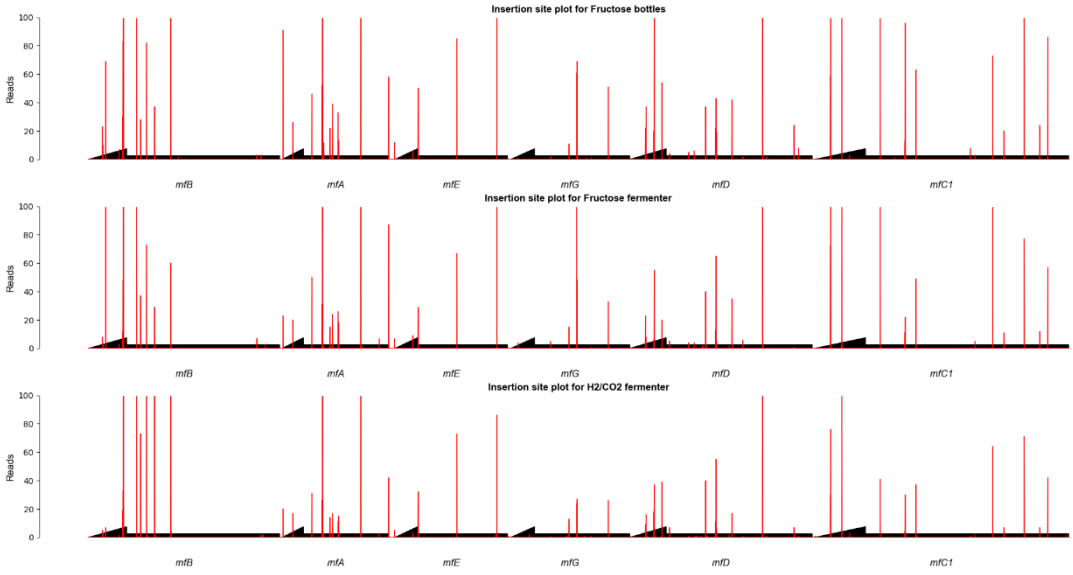


Figure 4-17: Insertion plots in genes of the Rnf cluster for the second fructose subculture and final time point of both fermenters. Plots were capped at 100 reads to improve resolution of insertion sites with lower read counts.

In *A. woodii*, the Rnf complex has been successfully deleted using fructose as a substrate (Westphal et al., 2018). Deletion was shown to result in complete loss of

ferredoxin:NAD^+ oxidoreductase activity and associated Na^+ translocation. When grown on fructose, the *rnf* deletion resulted in a lower acetate to fructose ratio and higher hydrogen production compared to WT (Wiechmann et al., 2020). While the absence of the Rnf complex resulted in a deficit in NADH for the oxidation of CO_2 and associated acetate production, this demonstrated activity of the HydABCDE hydrogenase.

While the Δ*rnf* strain did not grow on H_2/CO_2, acetogenesis was still possible in resting cells at a reduced rate (20%). ATP synthesis, however, was not observed (Westphal et al., 2018). The lack of ATP synthesis is likely caused by the absence of sodium ion gradient and subsequent inability of the ATP synthase to employ it for ATP synthesis.

In the present study, mutants carrying a transposon insertion in one of the *rnf* genes did not appear to have decreased fitness. Compared to the Rnf deletion strain described above, only one of the genes of the Rnf complex was disrupted in each mutant and the library population is not a pure culture. One or both of these factors might have had an impact on the essentiality status of the Rnf genes. Further investigation is required to elucidate the mechanisms that enabled survival of mutants with transposon insertion in the Rnf complex.

ii) F-type ATPase

The F-type ATP synthase is encoded by an operon of 11 genes (Awo_c02140 – Awo_c02240). It was shown to be dependent on the sodium ion gradient for ATP synthesis (Fritz & Müller, 2007; Müller et al., 2001) and its structure has been extensively studied (Brandt et al., 2013; Fritz & Müller, 2007; Müller et al., 2001),

Surprisingly, in static fructose growth, only *atpC* (Awo_c02240), *atpE3* (Awo_c02180), and *atpD* (Awo_c02230) were found to be ambiguous (Figure 4.12). In the fructose fermenter, only *atpC* was essential and *atpE3* was ambiguous. In the H_2/CO_2 fermenter,

both genes were essential, in addition, *atpH* (Awo_c02200), *atpA* (Awo_c02210), and *atpD* were ambiguous. However, abundance of cells carrying single transposon insertions in the genes of the F-type ATPase significantly decreased over the two subcultures in fructose and over the duration of both the fructose and the H_2/CO_2 fermenters. While the decrease was more apparent for the gas fermentation, this showed that the F-type ATPase is required in both heterotrophic and autotrophic conditions. Study of insertion plots confirmed the low prevalence mutants with transposon insertions in the F-type ATPase despite the non-essential status of several of its genes (Figure 4.17).

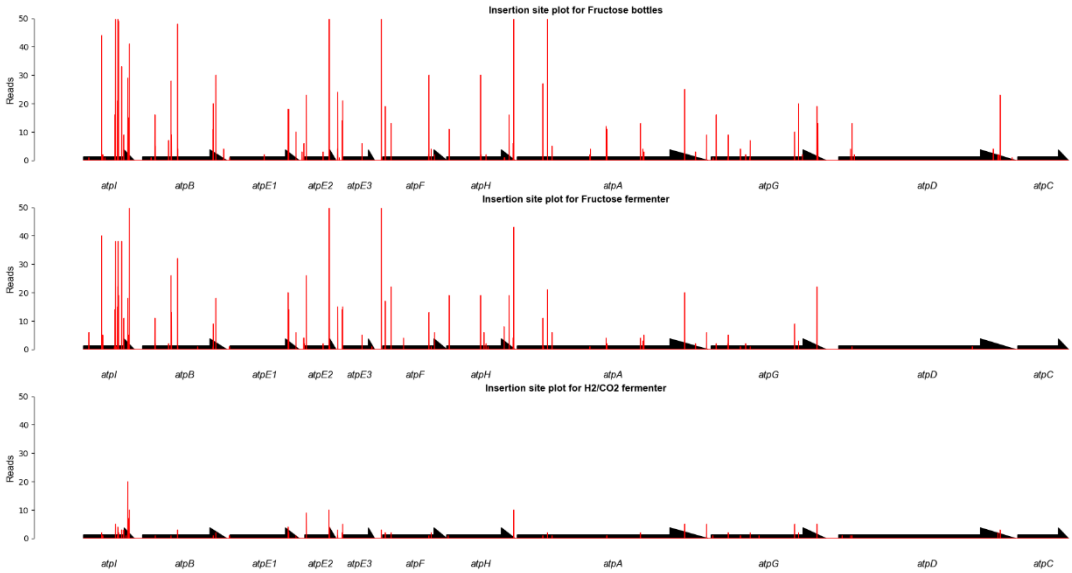


Figure 4-18: Insertion plots in genes of the F-type ATPase for the second fructose subculture and final time point of both fermenters. Plots were capped at 50 reads to improve resolution of insertion sites with lower read counts.

During H_2/CO_2 gas fermentation, the ATP synthase is responsible for the only 0.3 mol of ATP generated (Bertsch & Muller, 2015a). This can explain the dramatic decrease of the abundance of the single mutants with insertion in the genes not found essential. During growth on fructose, the ATP synthase is only responsible for generation of 0.3 mol of ATP while substrate-level phosphosphorylation generates 4 mol (Schuchmann &

Muller, 2016).This may explain the slower decline in abundance of mutants during fructose fermentation compared to gas fermentation.

iii) V-type ATPase: detrimental to H_2/CO_2 fermentation

Acetobacterium woodii contains an additional ATPase of the V-type, encoded by the gene cluster Awo_c023890 – Awo_c23970 (Figure 4.18). All its subunits were found to be non-essential for heterotrophic and autotrophic growth. During fructose fermentation, mutants with insertions in each of the different subunits are not positively or negatively selected. On the contrary, over the course of the H_2/CO_2 fermentation, single mutants with transposon insertion in the ATPase subunits display fold changes comprised between 22.23 and 11.94. *A. woodii* V-type ATPase appeared to be highly detrimental to H_2/CO_2 fermentation.

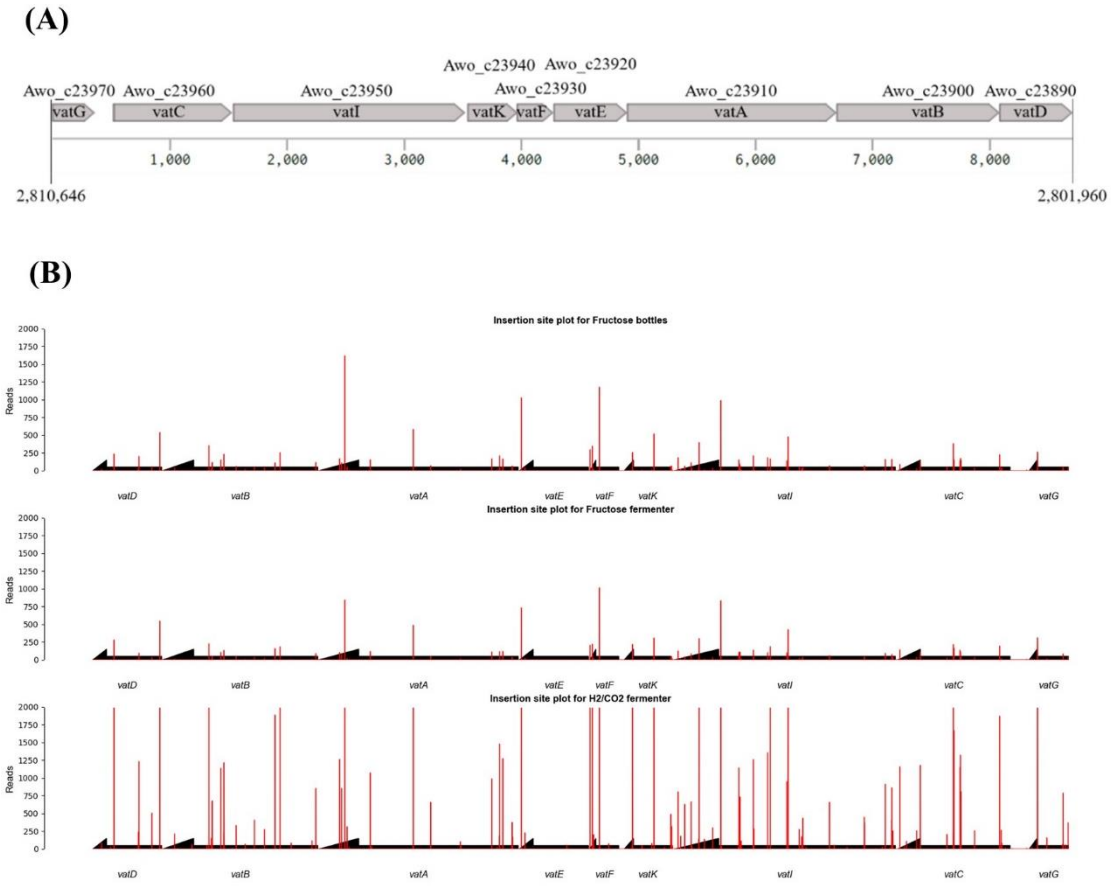


Figure 4-19: *A. woodii* V-ATPase under the different conditions. (A) V-ATPase gene organisation of *A. woodii*, genes are in reverse direction between the positions 2,801,960 and 2,810,646. **(B)** Insertion plots in genes of the V-ATPase cluster for the second fructose subculture and final time point of both fermenters. Plots were capped at 2,000 reads.

Unfortunately, the role of this ATPase in *A. woodii* remains elusive. Most bacteria use F-type ATPases for ATP generation, as described in *A. woodii*, and co-occurrence of F- and V-type ATPases is relatively rare. The distribution of these systems was studied in 700 genomes in 2011 and only 44 contained both (Peng et al., 2011), but co-occurrence of the two systems was found across the *Acetobacterium* genus (Ross et al., 2020). Both V- and F-type ATPases were found in the membrane of *C. thermocellum* but their roles remained unclear, the authors concluded that the presence of the two ATPases could be explained as different alternatives for ATP synthesis (Peng et al., 2011).

In archaeal A-Type ATPase ATP generation occurs in association with H^+ and/or Na^+ translocation (Grüber et al., 2014). Both F- and V-type ATPases were found and characterised in the archaeon *Methanosarcina mazei*; F- and V-type both allowed for ATP synthesis and they translocated Na^+ and H^+, respectively (Becher & Müller, 1994). Additionally, H^+-ATPases are efficient ion pumps responsible for pH regulation in cell compartments in eukaryotic cells by proton export linked to ATP hydrolysis at low pH and increase of cell survival (Cotter & Hill, 2003; Forgac, 1999; Guan & Liu, 2020). In contrast, in both *Clostridium fervidus* and *Enterococcus hirae*, the V-type ATPase was found to be responsible for Na^+ export of the cell with ATP hydrolysis (Höner zu Bentrup et al., 1997; Kakinuma et al., 1999).

Based on the literature and considering the well-studied Rnf and F-type ATPase of *A. woodii*, it can be hypothesized that the V-type ATPase is responsible for proton translocation. Two hypotheses could explain its detrimental effect during autotrophic

growth. At high acetate concentrations, proton translocation by the ATPase might be more costly than other alternatives present in the cells, such as ion antiporters. It is also possible that maintenance of the ATPase system was metabolically costly, in the ATP-limited gas fermenter, cells lacking it might have had a fitness advantage. The detrimental effect observed for the V-Type ATPase remains puzzling, and the study of its function could help elucidate its impact on gas fermentation.

iv) Electron bifurcation

Two systems co-exist in *A. woodii* for electron bifurcation, a flavin-based electron bifurcating hydrogenase HydABCDE (Awo_26970 to Awo_c27010) and a *Sporomusa* type Nfn (Stn) electron-bifurcating transhydrogenase (*nuoG* (Awo_c21410), *nuoF2* (Awo_c21420), and *nuoE* (Awo_c21430)). While the Hyd system has been historically described as responsible for redox balancing and is the most well studied (Müller et al., 2001), Stn complexes have been identified and described in acetogens more recently (Kremp et al., 2020; Marcellin et al., 2016).

The Hyd complex oxidizes H_2 into H^+ and transfers the electrons to NAD^+ and Fd. In contrast, the Stn complex balances NADP(H) with NAD(H) and ferredoxin by transferring the electrons from NADPH to NAD^+ and Fd (Kremp et al., 2020). The Hyd hydrogenase was found to be non-essential in all the conditions tested, except HydA1 and HydE that were found to be ambiguous in the fructose fermenter. However, Figure 4.12 showed a clear contrast in its role in fructose and H_2/CO_2 fermenters based on fold change of read abundance. Mutants with insertion in any of the genes of this hydrogenase appeared to have decreased fitness in the fructose fermenter but increased fitness in the H_2/CO_2 fermenter (Figure 4.19), indicating that the Hyd complex was important for growth in the fructose fermenter but detrimental to growth in the H_2/CO_2 fermenter. Mutants with transposon insertion in any of the genes of the hydrogenase

were almost equally negatively or positively selected depending on the condition, demonstrating the importance of each of the individual subunits.

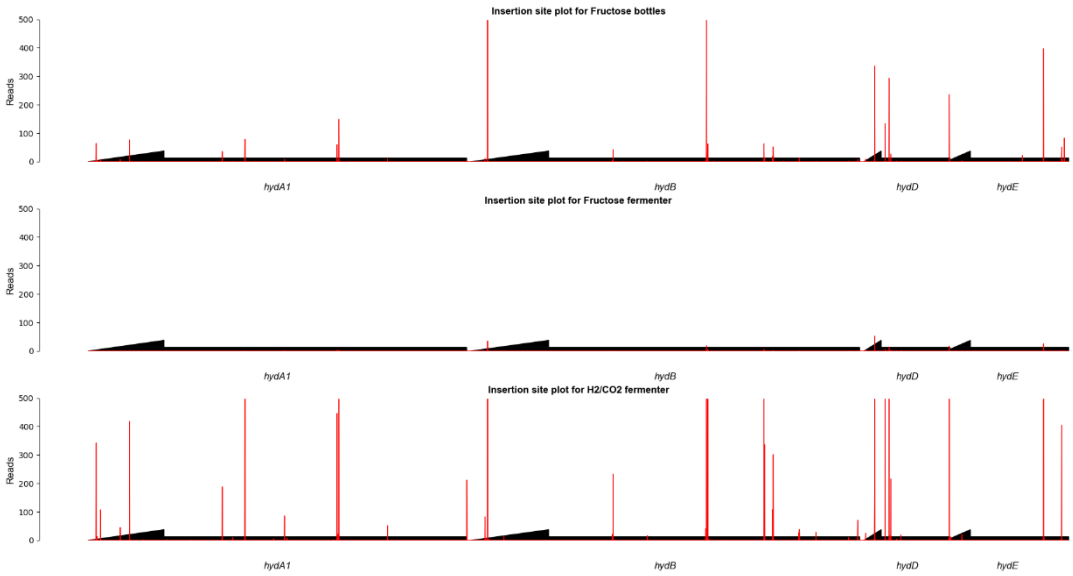


Figure 4-20: Insertion plots in genes of the electron-bifurcating hydrogenase for the second fructose subculture and final time point of both fermenters. Plots were capped at 500 reads.

In contrast, the Stn complex was found to be essential during heterotrophic growth and likely to be essential during autotrophic growth. Although the library was generated on fructose, creating a bias in the viable mutants, the remaining insertion mutants in the Stn genes were found to be eliminated from the library pool during gas fermentation (11, 5, and 2 UIS with 19, 12, and 5 reads in total remained at the start in genes *nuoG*, *nuoF2*, and *nuoE*, respectively, none were observed at the end of the fermentation).

Both Stn and Hyd are required for heterotrophic growth

During growth on fructose (in the absence of H_2), both Stn and Hyd complexes appeared essential. It has been predicted that NADH and Fd^{2-} is consumed by the Stn during fructose growth for NADPH production, necessary for anabolic reactions (Neuendorf et al., 2021), the present results confirmed its essentiality.

Fructose is converted to acetate though the EMP, generating ATP, CO_2, reduced ferredoxin and NADH. The CO_2 produced is further reduced to acetate through the WLP, the required hydrogen for formate production is generated from H^+, NADH, and reduced ferredoxin by the Hyd hydrogenase (Bertsch & Muller, 2015a; Müller et al., 2001; Neuendorf et al., 2021; Schuchmann & Muller, 2016). The HDCR can efficiently capture the hydrogen produced by the hydrogenase before it can diffuse though the membrane, it has been shown that when the hydrogenase is absent (Δ*hydAB*) *A. woodii* is unable to grow on fructose (Wiechmann et al., 2020). Heterotrophic growth could be restored by addition of molecular hydrogen or formate (Wiechmann et al., 2020). This does not conflict with the data presented here, indeed, the anaerobic cabinet used for library construction and static fructose growth was kept under anaerobic conditions by a CO_2:H_2:N_2 (10:10:80, v/v/v) gas mix while the fermenter used for continuous heterotrophic fermentation was only flushed with nitrogen. The difference in hydrogen presence can explain the successful generation of the mutants with transposon insertion in the Hyd, their survival during static fructose growth, and their elimination from the fructose fermenter. The hydrogen high volatility would have also prevented Hyd insertion mutants to utilize hydrogen produced by the rest of the transposon library population during continuous fermentation.

Hyd is detrimental during autotrophic growth

While Stn appeared essential, Hyd appeared detrimental to autotrophic growth. This is in direct contradiction with a part of the current literature, the hydrogenase has been shown to be essential for production of reducing equivalents though hydrogen oxidation during autotrophic growth. The Hyd hydrogenase was found to increase in abundance both transcriptionally and translationally between heterotrophic and autotrophic growth, demonstrating its apparent increased necessity during the latter (Shin et al., 2021).

Additionally, a *hydAB* deletion strain did not grow on H_2+CO_2+formate (Wiechmann et al., 2020) and Stn was not predicted to be responsible for any NAD(H) or ferredoxin consumption or production when grown on gas (Neuendorf et al., 2021).

As an example, *C. autoethanogenum* contains both a Hyt and an Nfn system, while the Nfn system has the same function as the predicted Stn system in *A. woodii*, the Hyt complex is similar to *A. woodii* Hyd system but uses NADPH and not NADH (Schuchmann & Müller, 2012; Wang et al., 2013). In *C. autoethanogenum*, deletion of the Nfn complex had no impact on growth on fructose and CO in the absence of H_2 but abolished growth in CO + H_2 (Marcellin et al., 2016). This indicated that redox balancing by Hyt cannot be performed in the presence of excess hydrogen.

Overall, the data presented here demonstrated that the deletion of the Hyd system in *A. woodii* likely resulted in all the electron bifurcation to be performed by the Stn system and improved fitness of the insertion mutants. Study of transcriptional and translational regulation of the Stn system in the presence and absence of Hyd would allow for quantification of the Stn compensation and for confirmation of this hypothesis. The deletion mutant could be studied for improved growth during autotrophic fermentations.

In 2016, Kantzow and co-workers studied the impact of increasing hydrogen partial pressure on acetate yields in *Acetobacterium woodii* (Kantzow & Weuster-Botz, 2016). An increase in partial pressure led to formate accumulation and they hypothesized that high hydrogen concentrations had an inhibitory effect on the "central carbon and energy metabolism" (Kantzow & Weuster-Botz, 2016). Overexpression of genes *pta* (phosphotransacetylase) and *ackA* (acetate kinase) from *C. ljungdahlii* improved acetate production and ATP production, and reduced formate accumulation (Kantzow & Weuster-Botz, 2016). Based on the present study, it can be hypothesized that growth and

acetate production could potentially be maintained at high H_2 pressure in a strain lacking Hyd.

4.2.3.3 Glycine cleavage system, acid stress and Na^+ transport

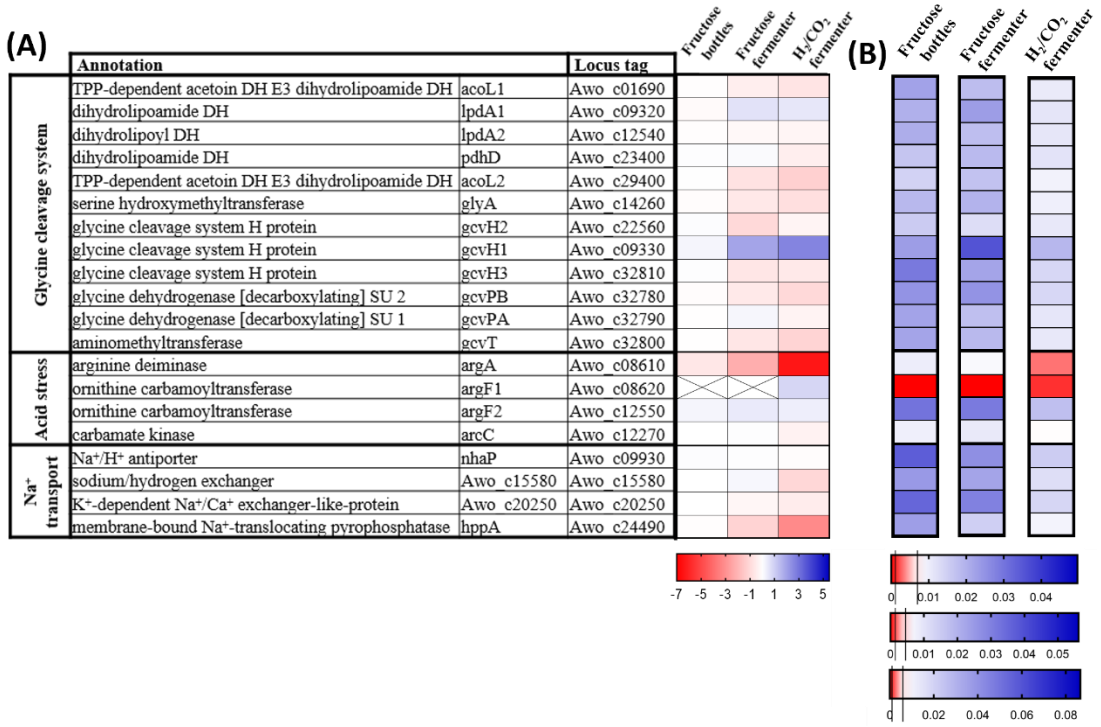


Figure 4-21: Gene-level abundance and essentiality studies of the genes involved in pyruvate metabolism, glycine cleavage system, acid stress and Na^+ transport. (A) Gene-level reads abundance, the colour gradient represents the $\log_2$ of the fold change of reads abundance at the gene-level crosses indicate instances where no fold change could be computed because of a null value. **(B)** Essentiality study with independent colour gradient for each condition, lowest value in red, central value is the ambiguous cut-off determined for each condition. Both the essential and ambiguous cut-offs are represented on the colour gradients.

4.2.3.3.1 Glycine cleavage system (GCS)

The glycine cleavage system (GCS) is part of the reductive glycine pathway and enables both glycine synthesis and degradation (Bar-Even et al., 2013). Unlike all other *Acetobacterium*, in *A. woodii*, genes *gcvH1* and *lpdA1* (dihydrolipoamide dehydrogenase) are directly downstream of cluster II of the WLP (Ross et al., 2020).

It has been recently hypothesized that this pathway could enable CO_2 fixation in acetogens, after synthesis, glycine can be reduced to acetyl-phosphate and further to acetate by *ackA* (acetate kinase) (Song et al., 2020). In *A. woodii*, the GCS-associated genes (genes *gcvPB*, *gcvPA*, *gcvT*, and *gcvH3* only) were shown to be upregulated in heterotrophic conditions compared to autotrophic conditions and, when expressed in *E. limosum*, increased growth rates, acetate production, and fructose consumption were observed (Shin et al., 2021). In contrast, genes *gcvH1* and *gcvH2*, as well as the genes forming the dihydrolipoyl dehydrogenase (including lpdA1), were upregulated during autotrophic growth (Shin et al., 2021).

During formate growth, the same set of genes (*gcvPB*, *gcvPA*, *gcvT*, and *gcvH3*) was reported to be upregulated during growth of *A. woodii* by comparison to H_2/CO_2 or fructose and the authors concluded that this upregulation enabled improved uptake of glycine provided in the media (Neuendorf et al., 2021). They found only a minor role of the GCS for carbon fixation during formate growth (Neuendorf et al., 2021).

In the present study, abundance of mutants with transposon insertion in all the genes from the glycine cleavage system except *lpdA1* and *gcvH1* appeared to be either staying constant or slightly decreasing during both fermentations (Figure 4.20). As such, genes *lpdA1* (Awo_c09320) and *gcvH1* (Awo_c09330) appeared to be slightly detrimental to both heterotrophic and autotrophic fermenter growth. Reads in the glycine cleavage system H gene *gcvH1* exhibited a 2.84- and 5.28-fold change of abundance in the heterotrophic and autotrophic fermenter, respectively. The number of single mutants was quite low, so the fold changes had a low confidence.

Taken with the previous transcriptomic and proteomic analyses (Moon et al., 2021; Neuendorf et al., 2021; Shin et al., 2021), this data indicated a separate function of the

gcvH1 and *lpdA1* from the rest of the glycine cleavage system. In contrast to the rest of the GCS, they appeared to be detrimental for both heterotrophic and autotrophic growth.

If the whole GCS was to work concomitantly with the WLP, it is surprising that only the two genes *gcvH1* and *lpdA1* are situated at the end of the WLP operon.

Poehlein and co-workers hypothesized that *gcvH1* and *lpdA1* encode a CO dehydrogenase-disulfide reductase responsible for acetyl-CoA formation based on research performed in *M. thermoacetica* (Pezacka & Wood, 1984; Poehlein et al., 2015). Another study appears to hypothesize that presence of a complete glycine cleavage/synthesis pathway results in down regulation of the WLP in *Clostridium drakei*, explaining the presence of *gcvH* and *lpdA1* only in acetogens (Shin et al., 2016).

4.2.3.3.2 Acid stress

The arginine deaminase pathway has been found to help mitigate acid stress by converting arginine to ornithine, ammonium, and CO_2 while generating ATP (Cotter & Hill, 2003; Guan & Liu, 2020). pH regulation results from production of NH_3 and neutralization of intracellular protons, effectively decreasing internal protons (Katsyv & Müller, 2020; Vrancken, Rimaux, Weckx, et al., 2009). First, arginine is converted to citrulline and ammonia (arginine deaminase), next citrulline is converted to ornithine and carbamoyl phosphate (carbamoyltransferase) and finally carbamoyl phosphate is further converted to carbon dioxide and ammonia, accompanied by ATP production (carbamate kinase). This mechanism has been extensively studied in *Propionibacterium acidipropionici* (Guan et al., 2013; Guan et al., 2014) and *Lactobacillus* (Vrancken, Rimaux, Weckx, et al., 2009; Vrancken, Rimaux, Wouters, et al., 2009). In *Lactobacillus fermentum*, fermentation pH modified the proportion of citrulline and ornithine produced. Sub-optimal pH led to mainly ornithine production whereas optimal pH

conditions led to equal amounts of citrulline and ornithine (Vrancken, Rimaux, Weckx, et al., 2009).

In *A. woodii*, the arginine pathway comprises the arginine deaminase *argA* (Awo_c08610), the carbamoyl transferase with genes *argF1* (Awo_c08620) and *argF2* (Awo_c12250) and the carbamate kinase *arcC* (Awo_c12270) (Figure 4.21).

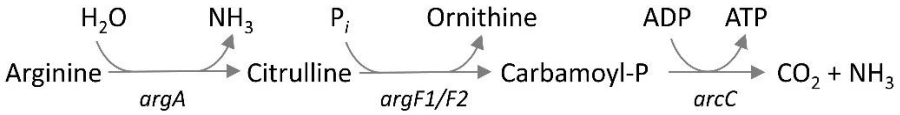


Figure 4-22: Arginine deaminase pathway in *A. woodii*.

In this study, gene *argF1* was found to be essential in all conditions tested and reads in *argA* showed a decrease in abundance and was found to be ambiguous during H_2/CO_2 fermentation. This corroborated the upregulation of *argA* observed during H_2/CO_2 growth compared to formate growth at the translational level (Neuendorf et al., 2021). In contrast, *arcC* had no impact on *A. woodii* fitness during fermentation. The essentiality study revealed that all the genes for UMP biosynthesis from carbamoyl phosphate were essential in both fermenters. It can be hypothesized that during fermentation, the arginine pathway plays a vital role in pH regulation and in pyrimidine biosynthesis.

To test if the arginine deaminase pathway could improve the bioenergetics of *A. woodii*, inducible and constitutive expression of the deaminase pathway genes of *C. autoethanogenum* was implemented (Beck et al., 2020). Strong constitutive expression resulted in poor or no growth in mixotrophic or autotrophic grown cells, potentially because of a metabolic flux imbalance (Beck et al., 2020). In contrast, moderate overexpression of the pathway when grown mixotrophically and with arginine as a nitrogen source resulted in higher optical densities but lower acetate production. As

less fructose was converted to acetate and more fumarate was observed, the authors hypothesized a downregulation of the WLP (Beck et al., 2020). This growth advantage was not observed during autotrophic growth (Beck et al., 2020).

4.2.3.3.3 Na^+ transport

None of the genes involved in sodium ion transport listed were found to be essential and the single mutants displayed minor variation of abundance. Gene *hppA* (Awo_c24490) had the highest fold change of the genes listed in this section. This observation correlated with the literature as the other genes could not be detected in proteomics studies (Neuendorf et al., 2021) and *hppA* had the highest transcript abundance when grown both heterotrophically and autotrophically (Moon et al., 2021).

4.2.3.4 Alternative substrates

Other characteristic features of the *A. woodii* genome include pathways to utilise a wide variety of substrates as carbon and energy sources (Drake et al., 2008), including formate (Moon et al., 2021), ethanol (Bertsch et al., 2016), methanol (Kremp et al., 2018), 2,3-butanediol (Hess et al., 2015), 1,2-propanediol (Schuchmann et al., 2015), lactate (Schoelmerich et al., 2018; Weghoff et al., 2015), alanine (Donig & Muller, 2018), glycine betaine (Lechtenfeld et al., 2018), and caffeate (Hess, González, et al., 2013).

Genes constituting the formate, ethanol, methanol, and caffeate pathways were not essential and mutant cells carrying transposon insertions in these genes exhibited little change in abundance during heterotrophic and autotrophic growth (Figure 4.22). Those genes pathways are thus unlikely to provide any increased or decreased fitness in the conditions tested and were not discussed. Genes from the 2,3-butanediol utilisation

pathway were non-essential but appeared to be slightly beneficial for heterotrophic and autotrophic growth.

Interestingly, the transcriptional regulator (*lctA*) of the lactate dehydrogenase genes was found to be ambiguous in the three conditions tested (Figure 4.22). The lactate gene cluster has been shown to only be expressed when lactate is present in the medium (Schoelmerich et al., 2018). Quantification of the gene cluster expression by qPCR revealed that its expression during fructose and H_2/CO_2 growth was approximately -6 and -8 $\log_2$ lower, respectively, than during lactate growth (Schoelmerich et al., 2018). The regulator *lctA* was found to bind to the DNA upstream of the gene cluster and repress expression. Binding and repression was relieved only in the presence of lactate. The ambiguous status of the regulator in the present study revealed that repression is necessary in all the conditions tested and the activity of the lactate dehydrogenase is detrimental to strain fitness when lactate is not present. This demonstrated the evolution of a strong repressor system to increase strain fitness in the absence of lactate.

Several genes of the 1,2-propanediol pathway were found essential during fermentation and are discussed below. Genes of the alanine and glycine betaine pathways were found to be detrimental to autotrophic and heterotrophic fermentation, respectively, they are discussed in the sections below.

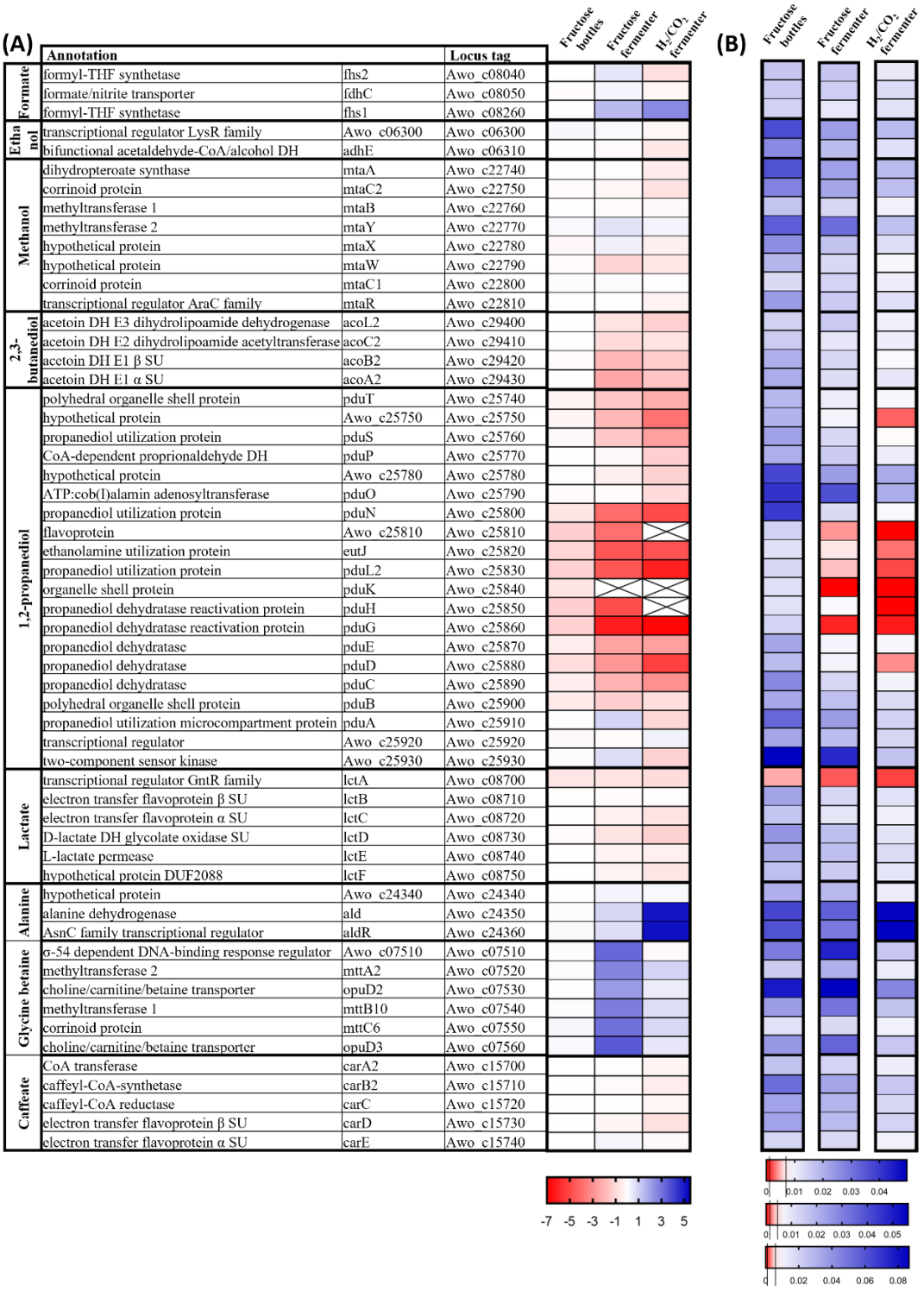


Figure 4-23: Gene-level abundance and essentiality studies of the genes involved in alternative substrates utilisation. (A) Gene-level reads abundance, the colour gradient represents the $\log_2$ of the fold change of reads abundance at the gene-level; crosses indicate instances where no fold change could be computed because of a null value. **(B)** Essentiality study with independent colour gradient for each condition, lowest value in red, central value is the ambiguous cut-off determined for each condition. Both the essential and ambiguous cut-offs are represented on the colour gradients.

4.2.3.4.1 1,2-propanediol pathway

In *A. woodii*, 1,2-propanediol degradation has been shown to lead to the production of propanol and propionate in equal proportions and to be independent from acetogenesis (Schuchmann et al., 2015). The 1,2-propanediol degradation pathway gene cluster is composed of 20 genes (Awo_c25740 to Awo_c25930), including a transcriptional regulator (Awo_c25920) and sensor histidine kinase (Awo_c25930). It was hypothesized to detect alcohol with chain lengths superior to 2 carbon or aldehydes intermediates (Chowdhury, Alberti, Linder, & Müller, 2020). 1,2-propanediol is first dehydrated to propionaldehyde by a propanediol dehydratase PduCDE and its associated proteins PduGH and PduS. Propionaldehyde is further converted to propanol by the alcohol dehydrogenase Adh4 (Chowdhury et al., 2021) and to propionyl-CoA by the aldehyde dehydrogenase PduP (Schuchmann et al., 2015). Propionyl-CoA is converted to propionate with propionyl-phosphate by PduL and an unidentified kinase (Schuchmann et al., 2015).

Genes *pduABKNT* and the chaperone encoding *eutJ* of the 1,2-propanediol cluster encode for organelle shell proteins that are part of bacterial microcompartments (BMC) (Chowdhury, Alberti, Linder, & Müller, 2020; Schuchmann et al., 2015). BMC have been shown to lower the toxicity of intermediate compounds by their encapsulation within the cells. In *Salmonella* it was shown to isolate aldehydes during metabolic processes to avoid toxicity and DNA mutations (Sampson & Bobik, 2008). The BMC encoded by the *pdu* gene cluster has been shown to have a prominent role in *A. woodii* and the genes shown to be expressed under a variety of conditions, including during growth on fructose and H_2/CO_2 growth (Chowdhury, Alberti, Linder, & Müller, 2020; Schuchmann et al., 2015).

Some genes of the pathway appeared to be important for both heterotrophic and autotrophic fermentations. Genes *pduK* and *pduG* were found to be essential in both fermentations, *eutJ* and *pduL2* were found to be ambiguous. Additionally, Awo_c25810 was ambiguous during growth on fructose and essential when growing on gas, *pduH* was essential in gas, and Awo_c25750, *pduS*, and *pduD* were ambiguous in gas only. Mutants in almost all the remaining genes of the cluster decreased in abundance during gas fermentation.

The alcohol dehydrogenase genes *adh4* and *adhE* were non-essential and mutant cells with transposon insertions displayed only minor variations of abundance under the conditions tested, with $\log_2$ fold change of -0.17 and -0.17, respectively, during growth on fructose and -0.86 and -0.64, respectively, during gas growth. This corroborated the observation that during growth on fructose and H_2/CO_2, while the alcohol dehydrogenase activity was 1.5% the activity recorded during growth on 1,2-propanediol, the activity of the aldehyde dehydrogenase (PduP) was 23% and 11%, respectively (Schuchmann et al., 2015). *A. woodii* has been shown to produce ethanol during fermentation (Buschhorn, Dürre, et al., 1989) and the analysis of the pathway revealed acetaldehyde as an intermediate of the conversion (Bertsch et al., 2016). Additionally, ethanol has been shown to activate expression of the *pduB* cluster in *A. woodii* (Chowdhury, Alberti, Linder, & Müller, 2020; Schuchmann et al., 2015).

As previously hypothesized, the 1,2-propanediol gene cluster appeared to be essential for mitigation of the toxicity of aldehyde intermediates by encapsulation and detoxification of aldehydes (Schuchmann et al., 2015).The phenomenon was observed during both fructose and H_2/CO_2 growth, likely as a result of ethanol production during fermentation.

4.2.3.4.2 Alanine pathway: detrimental to H_2/CO_2 fermentation

The alanine pathway is composed of genes encoding a transcriptional regulator *aldR* (Awo_c24360), a dehydrogenase *aldD* (Awo_c24350) and a transporter *aldT* (Awo_c24340)(Figure 4.23A). The alanine dehydrogenase requires NAD^+ for alanine oxidation to pyruvate, pyruvate can then be converted into acetyl-CoA by NifJ as discussed in section 4.2.3.1.

Genes *aldR*, *aldD*, and *aldT* are transcribed as one operon along with *nifJ* (Awo_c24330) (Donig & Muller, 2018). Reducing equivalents NADH and ferredoxin are re-oxidized by reduction of CO_2 to acetate through the WLP. Overall, alanine degradation generates 1.15 mol ATP / mol alanine and NH_4^+ and H^+ (Donig & Muller, 2018). Figure 4.22 shows that reads in both *aldR* and *aldD* drastically increased in abundance with a 33.31- and 35.84-fold change, respectively, while reads in *aldT* remained stable.

Analyses at the single mutant level confirmed that deletion of either gene *aldR* or *aldD* confered increased fitness on *A. woodii* during H_2/CO_2 fermentation, while deletion of *aldT* did not appear to have an impact (Figure 4.23B and E). Additionally, reads mapping against new loci in both *aldR* and *aldD* emerged over the course of the gas fermentation. In all genes analysed previously, some insertion sites absent from the first sample are sequenced in some of the later samples, but they never exceeded 10 reads and do not show any clear variation pattern, demonstrating transient sequencing of low frequency mutants. In the case of *aldR* and *aldD*, after their discovery some of the emerging reads proceeded to increase over the course of the gas fermentation (Figure 4.23C and D).

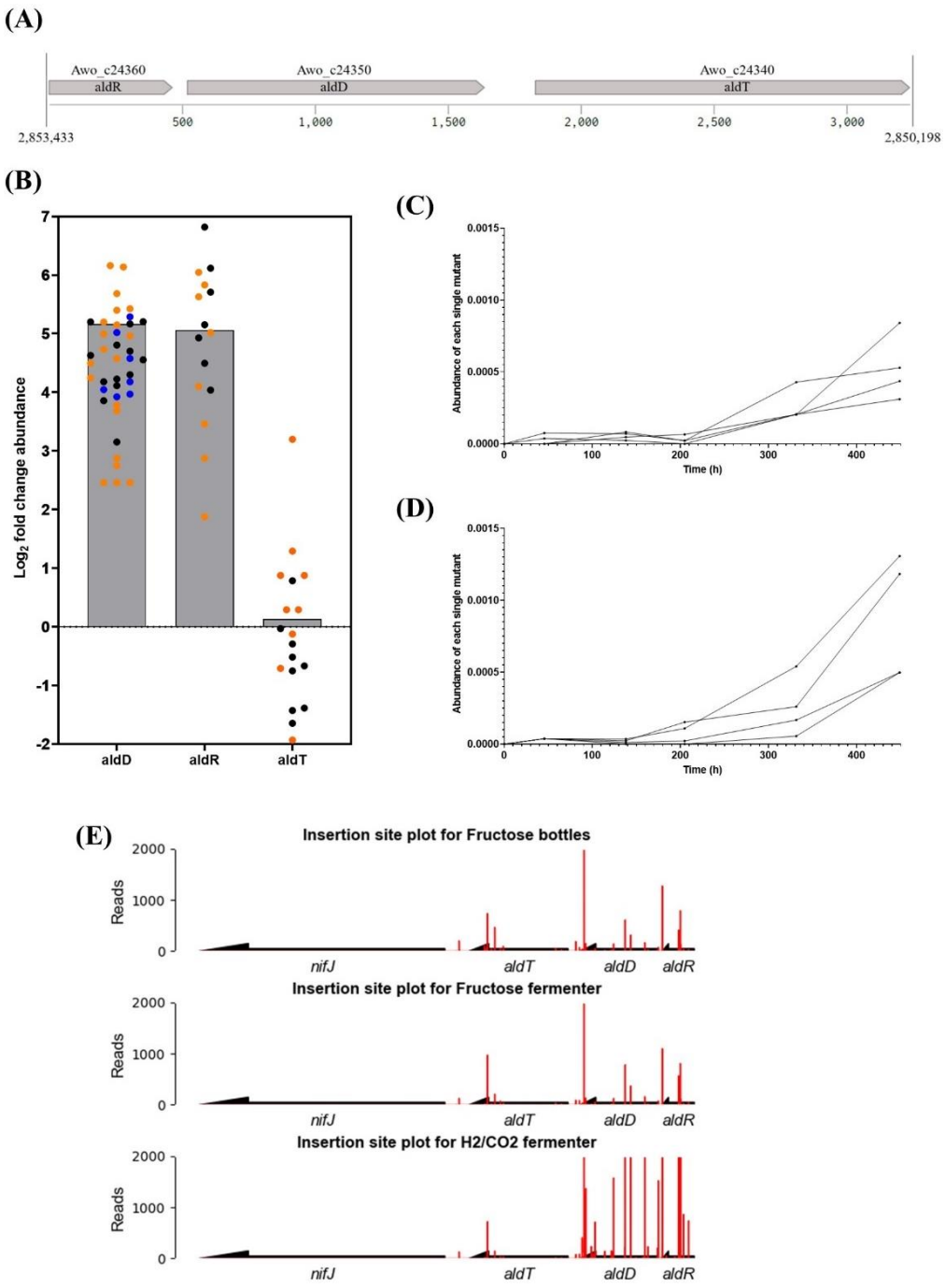


Figure 4-24: Analysis of the alanine pathway in *A. woodii* during H_2/CO_2 fermentation. **(A)** Gene organisation of the alanine pathway in *A. woodii*. Genes are situated between positions 2,850,198 and 2,853,433 in reverse orientation. **(B) G**ene-level log_2 fold change of reads abundance in each gene. Log_2 fold change of abundance of single mutants are represented as individual values. Mutants with a starting number of reads ≤10 and ≤20 are represented in orange and blue respectively, mutants starting with a number of reads >20 are represented in black. Abundance of single mutants not detected in the first sample for **(C)** *aldD* (Awo_c24350) and **(D)** *aldR* (Awo_c24360) over the course of the H_2/CO_2 fermentation, the mutants represented reach at least 10

reads during fermentation. **(E)** Insertion plots of genes of the alanine dehydrogenase cluster for the second fructose subculture and final time point of both fermenters. Plots were capped at 2,000 reads.

Because they started with an abundance of zero, emerging mutants cannot be represented in the single mutant level analyses. However, their emergence and increase in abundance provided further evidence that deletion of *aldR* and *aldD* is beneficial for autotrophic fermentation of *A. woodii.* This is surprising as genes *aldR*, *aldD*, and *aldT* are transcribed as one operon with *nifJ,* which was found to be essential in section 4.2.3.1. As such, it would be expected that insertions in the genes directly upstream would affect *nifJ* expression and be detrimental for growth.

Since the alanine transporter *aldT* did not display the same detrimental effect on autotrophic growth, it is likely that alanine production from pyruvate is detrimental rather than alanine uptake and degradation. Alanine was shown to be produced in both fermenters but the production profiles differed (Figure 4-25). Alanine was produced more rapidly in the autotrophic fermenter compared to the heterotrophic fermenter and peaked at 138 h. Alanine has previously been found as a fermentation product when phosphate was the limiting factor (Buschhorn, Durre, et al., 1989). It is likely that more recent studies did not measure alanine concentrations, more systemic measurements of amino acids concentrations would enable a better understanding of the conditions required for alanine production.

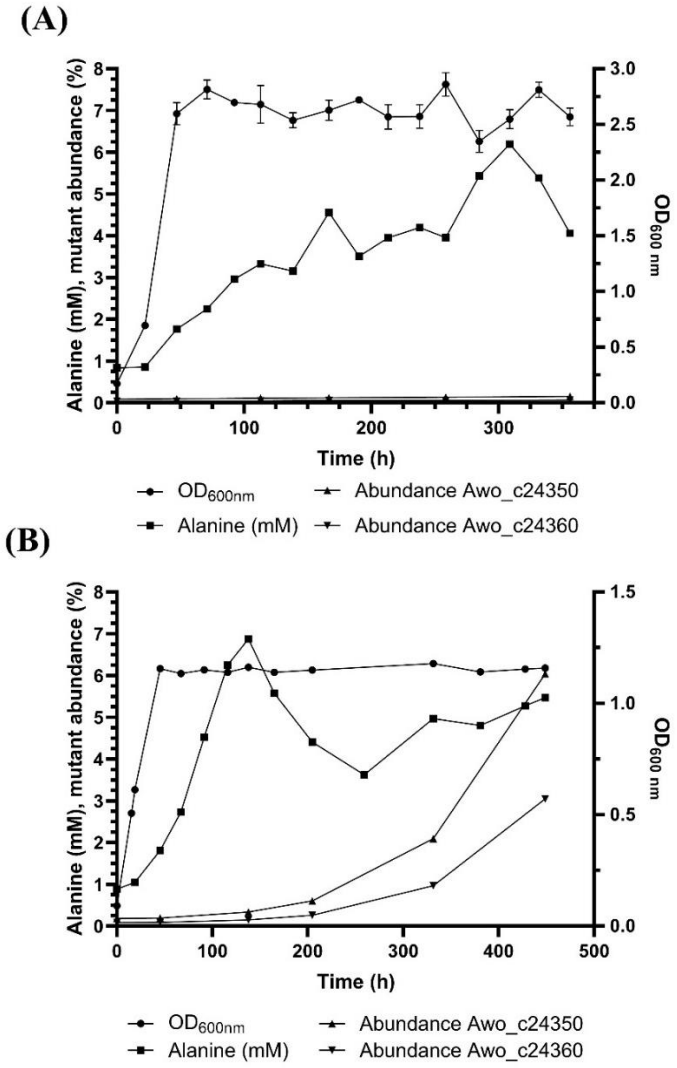


Figure 4-25: Alanine concentrations during both the (A) heterotrophic and (B) autotrophic (H_2/CO_2) continuous fermentations. $OD_{600\ nm}$ (●) and alanine concentrations (■) are represented over the course of the fermentations, as well as the cumulated abundance of mutants in Awo_c24350 (▲) and Awo_c24360 (▼).

A transaminase (Awo_c29660) is also annotated as responsible for alanine production. As such, disruption of the alanine dehydrogenase would likely not result in abolishment of alanine production. However, it appears that transamination is more favourable than the dehydrogenase system during gas fermentation. Further study of the function of both pathways would allow for a better understanding of the detrimental effect of the dehydrogenase.

Emergence of new insertion mutants in *aldD* and *aldR* was a new phenomenon not observed in previous genes and can strengthen the analysis of detrimental genes.

4.2.3.4.3 Glycine betaine pathway: detrimental to fructose fermentation

In 2018 Lechtenfeld and co-workers (Lechtenfeld et al., 2018) showed that *A. woodii* can utilise glycine betaine (present in yeast extract) as an energy and carbon source, as previously discovered in other acetogens (Andreesen, 1994; Eichler & Schink, 1984; Moller et al., 1984; Muller et al., 1981). Genes Awo_c07520 to Awo_c07560 were found to form an operon (Figure 4.24); gene *opuD3* was identified as encoding a glycine betaine transporter from the betaine-choline-carnitine family; genes *mttB10*, *mttA2*, and *mttC6* were identified as coding for the glycine betaine-specific methyltransferase I, methyltransferase II and corrinoid protein, respectively; the function of gene *opuD2* could not be established (Lechtenfeld et al., 2018). The σ^{54}-regulator upstream and in the opposite direction of the operon was hypothesized to play a role in the regulation of the pathway (Lechtenfeld et al., 2018).

Glycine utilisation was hypothesized to take place as follows: glycine betaine is demethylated into dimethylglycine, the methyl group is oxidized to CO_2 by reversal of the WLP, CO_2 is reduced to CO, and acetate is produced (Lechtenfeld et al., 2018).



Figure 4-26: Gene organisation of the glycine betaine uptake and utilisation in *A. woodii*, comprised between position 860,864 and 868,984.

Mutants with transposon insertion in each of the genes of the glycine betaine pathway, exhibited an increased abundance over time in the fructose fermenter (Figure 4.22), demonstrating that the pathway is detrimental during heterotrophic fermentation.

This data corroborated the observation that *A. woodii* cannot co-utilise glycine betaine and fructose (Kremp et al., 2018), further, the expression of the pathway is detrimental to growth on fructose. Unsurprisingly, fructose utilisation yields more ATP (4.3 ATP / mol fructose) than glycine betaine utilisation (0.625 ATP / mol glycine betaine without taking into account export of dimethylglycine (Lechtenfeld et al., 2018)).

As the presence of the regulator was detrimental, it is likely that it is an activator rather than a repressor. Glycine betaine in the medium is likely to have resulted in the activation of the pathway. Fructose utilisation is more favourable than glycine betaine, expression of the pathway might represent an unnecessary metabolic burden or compete with fructose utilisation.

4.2.4 Additional detrimental genes

While studying the gene-level fold change of the read abundance of genes in the central metabolism of *A. woodii*, other genes appeared detrimental to growth in the different conditions tested, they were studied at the single-mutant level. The results will be discussed in the following sections. The list of genes with the highest fold change of abundance during heterotrophic and autotrophic fermentations can be found in Supplementary Table 14 and Supplementary Table 15.

4.2.4.1 Detrimental genes for heterotrophic and autotrophic fermenter growth

4.2.4.1.1 Awo_c35490

Gene Awo_c35490 had a fold change of 11.99 and 10.65 in the fructose and H_2/CO_2 fermenter, respectively. Gene Awo_c35490 overlaps with its neighbouring genes (Figure 4.25A). Gene *rplI* is essential in both fermenters. Awo_c35490 is annotated as a

hypothetical membrane protein with a DUF2232 domain and Awo_c35480 as a signal transduction phosphoesterase, YybT-like protein.

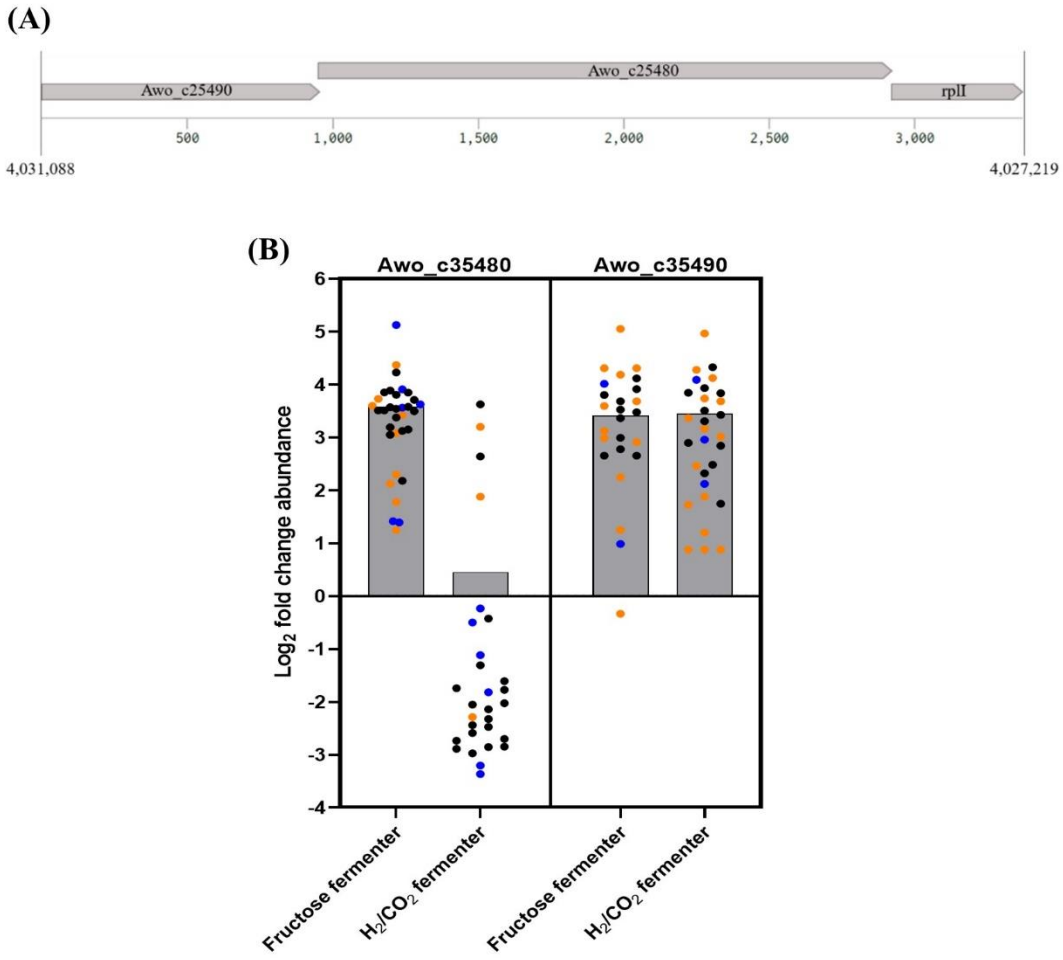


Figure 4-27: Gene Awo_c35480 and Awo_c35490 mutant abundance during fructose and H_2/CO_2 fermentations. (A) Relative gene positions of Awo_c35480 and Awo_c35490, in reverse position between positions 4,028,168 and 4,031,088, Awo_c35480 overlaps with *rplI* (Awo_c35470). **(B)** Bars represent the gene-level log_2 fold change of reads abundance. Log_2 fold change of abundance of single mutants are represented as individual values. Mutants with a starting number of reads ≤10 and ≤20 are represented in orange and blue respectively, mutants starting with a number of reads >20 are represented in black.

Single mutant analyses confirmed Awo_c35490 as detrimental to growth in both fermenters (Figure 4.25B). It is interesting to note that reads in this gene represented a high percentage of the population even in the first sequenced samples, with

approximately 0.22% of all mutants for both fermenters. This high abundance demonstrated that gene Awo_c35490 is detrimental during the library construction and revival.

Gene-level study of Awo_c35490 revealed that while KO of Awo_c35480 was very favourable for fructose fermentation, it appeared marginally favourable for H_2/CO_2 fermentation. Single-mutant level analyses confirmed the results observed for fructose fermentation but contradicted them for gas fermentation. Closer examination of the data revealed that one mutant had a $\log_2$ fold change of 3.62 and a high starting number of reads, creating a bias in the gene-level fold change of abundance. A majority of mutants with transposon insertion in Awo_c35480 displayed a $\log_2$ fold change of abundance between 0 and -3.37, demonstrating the importance of the gene during gas fermentation.

4.2.4.1.2 Awo_c26530 / Awo_c26540, *bcaT1* and Awo_26510

Five other genes were found to be detrimental in both fermenters (Figure 4.26).

The detrimental effect of gene Awo_c13070 could not be confirmed because of the low number of single mutants observed. Neighbouring genes Awo_c26530 and Awo_c26540 were also detrimental to both fermentations, this effect was confirmed by single mutant analyses. They encode a putative HIT family protein and a putative oxidoreductase, respectively.

As confirmed by single mutant analysis, Awo_c26510 and *bcaT1* were detrimental to both fermentations but more specifically to gas fermentation. Awo_c26510 has been annotated as a potential cation transporting P-type ATPase by motif analogy. It is likely that it had a detrimental effect on the homeostasis of the cells during fermentation. Interestingly, gene *bcaT1* is annotated as a branched-chain AA aminotransferase. In Gram-positive bacteria, branched chain amino acids are precursors for the biosynthesis

of branched chain fatty acid (Kaiser et al., 2016). In *A. woodii*, Gene *bcaT1* is identified as responsible for catalysis of the first step, the conversion of branched-chain amino acids to their corresponding branched-chain α-keto acids, genes for the following steps are not identified in *A. woodii*.

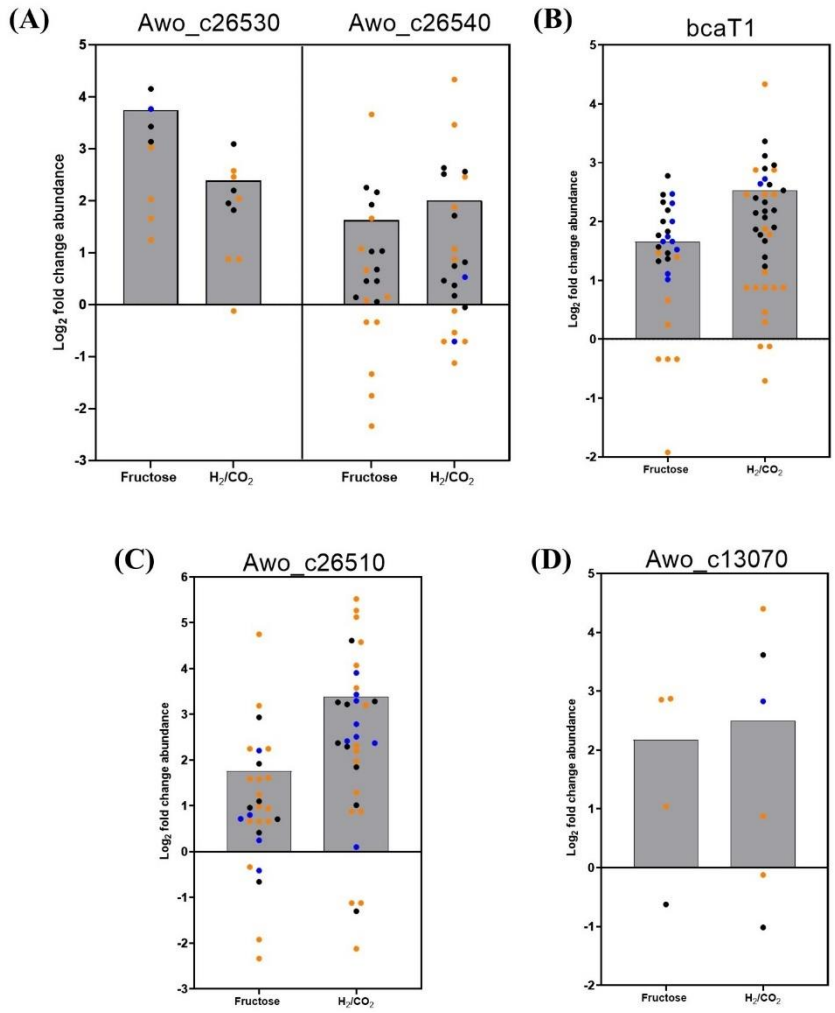


Figure 4-28: Log_2 fold change of genes detrimental to both fermenters. Genes **(A)** Awo_c26530 / Awo_c26540, **(B)** *bcaT1*, **(C)** Awo_c26510, and **(D)** Awo_c13070 mutant abundance during fructose and H_2/CO_2 fermentations. Bars represent the gene-level log_2 fold change of reads abundance. Log_2 fold change of abundance of single mutants are represented as individual values. Mutants with a starting number of reads ≤10 and ≤20 are represented in orange and blue respectively, mutants starting with a number of reads >20 are represented in black.

The fatty acid composition of the cell membrane has been shown to have an impact on membrane fluidity, including increasing the ratio of unsaturated fatty acids, increasing

fatty acids chain length, increasing cycloprane fatty acid, and altering the branching of fatty acids (Guan & Liu, 2020; Kaiser et al., 2016; Wu et al., 2012). Membrane fluidity was, in turn, found to participate in acid stress resistance in *Lactobacillus*, *Listeria*, and *Staphylococcus* (Guan & Liu, 2020; Kaiser et al., 2016; Sen et al., 2015; Sun et al., 2012; Wu et al., 2012). It can be hypothesized that the deletion of the *bcaT1* gene has an impact on membrane fatty acid composition, affecting the functionality of the cell membrane.

4.2.4.2 Detrimental genes for heterotrophic fermenter growth only

4.2.4.2.1 Methyltransferase MttC7 is extremely detrimental

A. woodii genes Awo_c08080 and Awo_c08070 (Figure 4.27A) have the highest and second highest fold change increase of read abundance at the gene-level during fructose fermentation with values of 45.79 and 61.70, respectively. During H_2/CO_2 fermentation, they appear to have no impact on strain fitness with fold change of 0.66 and 0.42, respectively. Study of single mutants during fructose growth confirmed the extremely detrimental status of both genes (Figure 4.27B). Single mutants in both genes consistently increased in abundance over the course of the fructose fermentation. One single mutant not detected in the first sample is detected in the following samples and proceeds to increase in abundance, as previously observed in the alanine pathway (section 4.2.3.4.2).

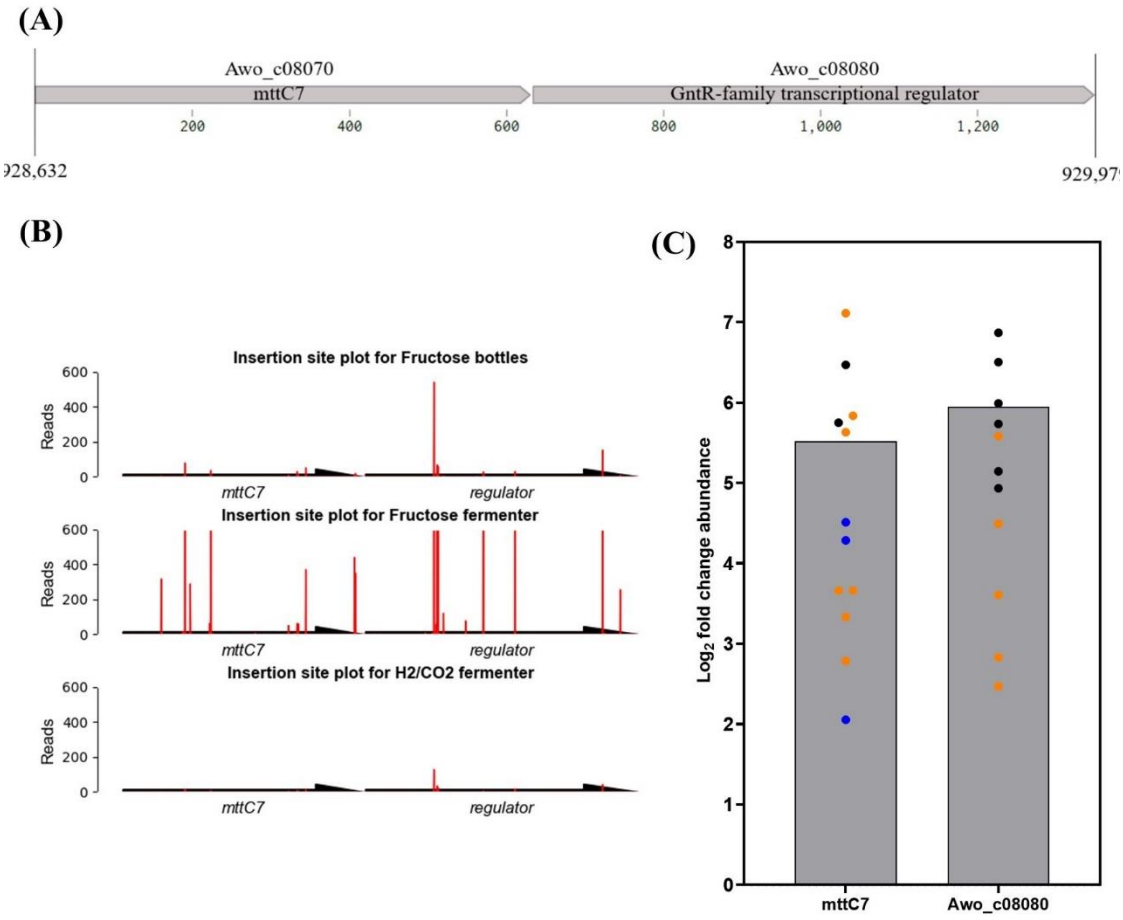


Figure 4-29: Methyltransferase mttC7 and GntR-family regulator mutant abundance in fructose fermenter. **(A)** Relative positions of genes mttC7 and GntR-family transcriptional regulator in *A. woodii*. **(B)** Insertion plots in genes of the *mttC7* and Awo_c08080 genes for the second fructose subculture and final time point of both fermenters. Plots were capped at 600 reads to improve resolution of insertion sites with lower read counts. **(C)** Bars represent the gene-level log_2 fold change of reads abundance. Log_2 fold change of abundance of single mutants are represented as individual values. Mutants with a starting number of reads ≤10 and ≤20 are represented in orange and blue respectively, mutants starting with a number of reads >20 are represented in black.

Gene Awo_c08080 is annotated as a GntR-family transcriptional regulator, taking into consideration the similar fold change of abundance in reads in both Awo_c08070 and Awo_c08080, it can be hypothesized that it regulates the activity of *mttC7* (Awo_c08070). Gene *mttC7* is annotated as a corrinoid protein. Analysis on InterPro revealed that it is likely a cobalamin-dependent methionine synthase (Supplementary Figure 3).

Methionine synthases catalyse methyl transfer from methyl tetrahydrofolate to homocysteine to form methionine. However, genes *mttC19* and *metH* are already identified in the genome of *A. woodii* as responsible for L-homocysteine to L-methionine conversion. In *E. coli,* high acetate concentrations were shown to inhibit methionine biosynthesis leading to accumulation of homocysteine, a toxic intermediate. Addition of methionine to the growth medium relieved that toxicity (Pinhal et al., 2019; Roe et al., 2002), as well as cytoplasmic accumulation of methionine because of metabolic deficiencies (Roe et al., 2002).

The function of *mttC7* remains unclear, but its impact on fructose fermentation could be characterised with its targeted deletion.

4.2.4.2.2 Miscellaneous genes

Several other genes exhibited a high fold change in read abundance during fructose fermenter, but contrary to the genes presented in the above sections, they are isolated genes. Miscellaneous genes with gene-level reads abundance fold change higher than 4 were analysed at the single mutant level during fructose fermentation and at the gene-level during H_2/CO_2 fermentation (Figure 4.28).

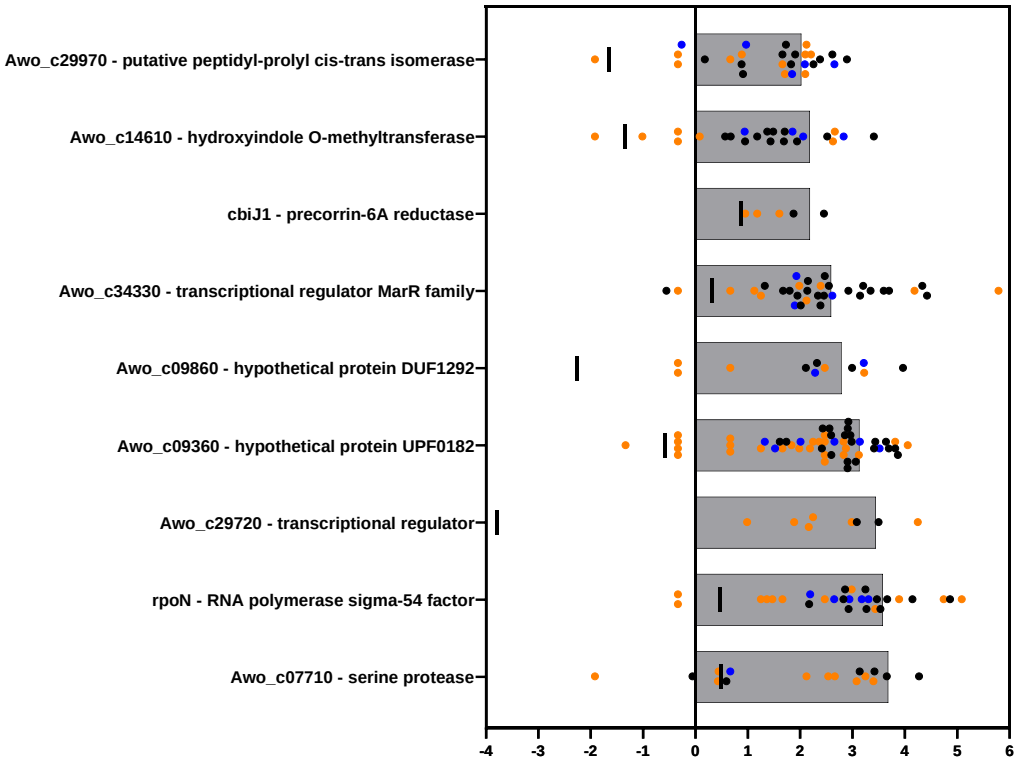


Figure 4-30: Miscellaneous genes mutant abundance in fructose fermenter. Bars represent the gene-level $\log_2$ fold change of reads abundance during fructose fermentation, lines represent the log2 fold change during H2/CO2 fermentation. Individual values are the $\log_2$ fold change of abundance of single mutants; mutants with a starting number of reads ≤10 and ≤20 are represented in orange and blue respectively, mutants starting with a number of reads >20 are represented in black.

The combined study of gene-level reads abundance in both fermenters and single mutant level can enable identification of isolated gene targets for genetic engineering.

Genes Awo_c29720 and Awo_c09860 have a low number of single mutants, however their gene-level $\log_2$ fold change of reads abundance in each fermenter confirmed that they are detrimental to fructose fermentation but important for H_2/CO_2 fermentation. Genes *rpoN*, Awo_c09360, and Awo_c34330 have a smaller difference in $\log_2$ fold change between the fermenters, the study of single mutants revealed a high number of single mutants with a consistently high $\log_2$ fold change. Those genes are likely to play

a detrimental role when growing on fructose even if this role is not important to growth on H_2/CO_2.

The emergence of single mutants was observed as follows: two in *rpoN* and Awo_c29970, one in Awo_c09360 and Awo_c34330, attesting to their highly detrimental role during fructose fermentation. In contrast, gene *cbiJ1* and Awo_c07710 are unlikely to play a definite role specific to the fructose fermenter.

4.2.4.3 Detrimental genes for autotrophic fermenter growth only

A set of genes appeared to be extremely detrimental to H_2/CO_2 fermenter growth with gene-level read abundance fold changes higher than 30. Genes *aldR* and *aldD* have high fold change values of 35.85 and 33.31, respectively, as discussed in a section 4.2.3.4.2. Genes *recF* and *gyrB* also displayed high fold changes at the gene-level, however, they were found to be essential in both fermenters. Closer examination of the single mutants revealed that one and three single mutants, respectively, were creating a bias in the general fold change. These mutants have insertions in the last 2% of the genes, and while the reason for their important increase in abundance is unclear, gene *recF* and *gyrB* can still be considered as essential in both fermenters.

4.2.4.3.1 Awo_c03590 is extremely detrimental

Among all of *A. woodii* genes and in both heterotrophic and autotrophic fermentations, the highest fold change calculated was 91.40. It was observed for gene Awo_c03590 during the H_2/CO_2 fermentation. In contrast, during the fructose fermentation, its fold change was 0.02, a value comparable to the fold change of reads in genes from the F-type ATPase and the HydABCDE hydrogenase during the same fermentation.

The study of single mutants confirmed that Awo_c03590 has opposite functions in the heterotrophic and autotrophic fermentations (Figure 4.29A and B).

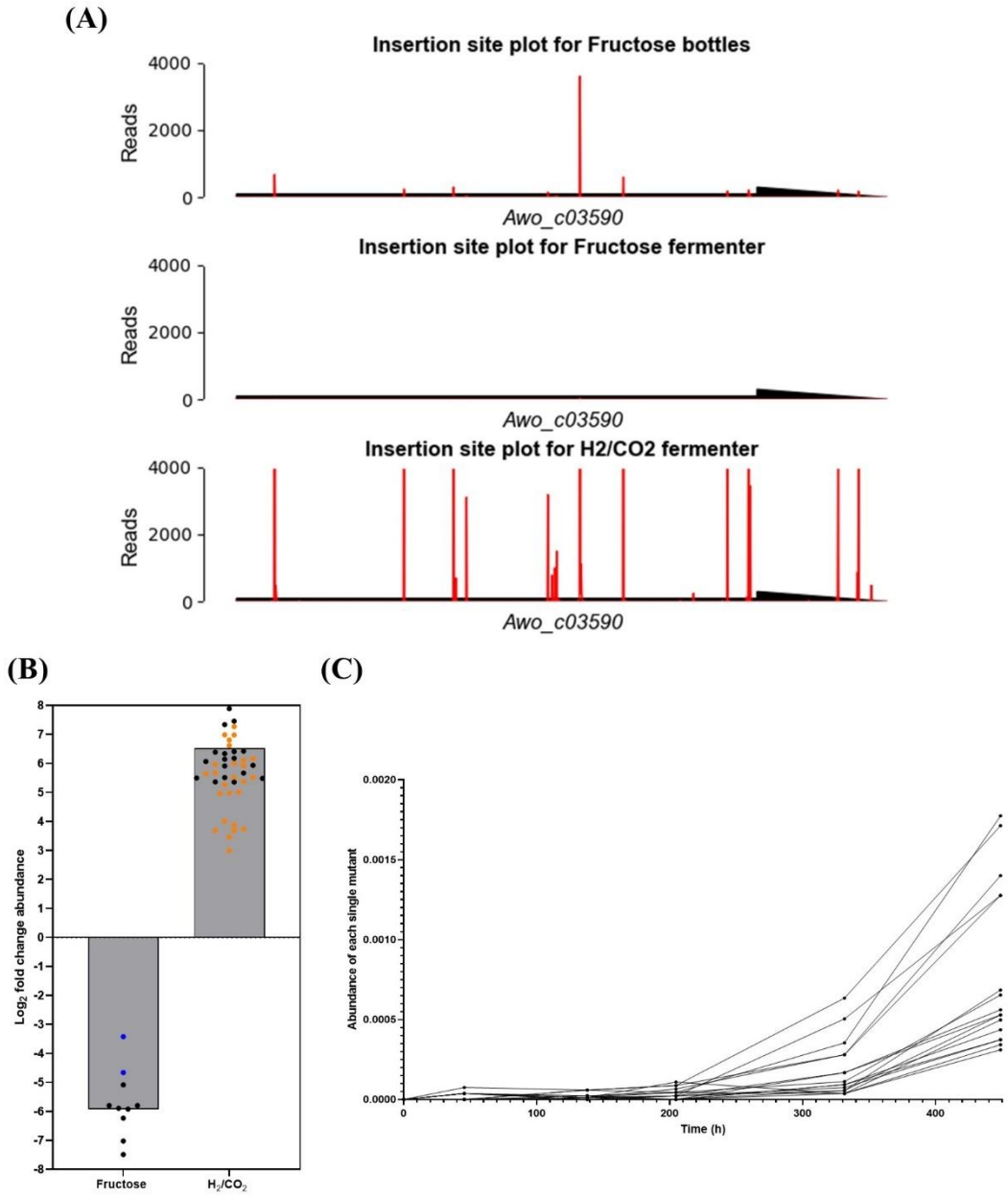


Figure 4-31: Analysis of Awo_c03590 insertion mutants during fermentation. (A) Insertion plots in genes of the Awo_c03590 gene for the second fructose subculture and final time point of both fermenters. Plots were capped at 4,000 reads. **(B)** Abundance of reads in Awo_c03590 during both fructose and H_2/CO_2 fermentations. Bars represent the gene-level log_2 fold change of reads abundance. Log_2 fold change of abundance of single mutants as individual values. Mutants with a starting number of reads ≤10 and ≤20 are represented in orange and blue respectively, mutants starting with a number of reads >20 are represented in black. **(C)** Abundance of single mutants over the course of the H_2/CO_2 fermentation, those mutants were not detected in the first sample and reach at least 10 reads during fermentation.

It appeared to be extremely detrimental to gas fermentation while being essential for fructose fermentation. The phenomenon of emergence of new insertion mutants was important, with more new mutants being observed than in the previous sections (4.2.3.4.2 and 4.2.4.2) and their increase in the second half of the gas fermentation (Figure 4.29C).

Awo_c03590 is annotated as a PilT-like protein or a twitching mobility protein in *A. woodii* genome. Analysis on InterPro unsurprisingly revealed a transmembrane structure with membrane embedded and cytoplasmic domains as well as small extracellular domains. Overall, it is a membrane-associated protein with RNA binding (TRAM domain) and nuclease activity (PIN domain) (Supplementary Figure 4).

The nomenclature of PIN domains originates from N-terminal domain of type IV pili twitching motility (PilT) proteins, hence the annotation as PilT-like protein in the *A. woodii* genome. There is, however, no functional link between PIN domains and type IV pili (Matelska et al., 2017). PIN-like domains are a widespread superfamily of nucleases with diverse functions in central cellular processes, including quality control and maturation of RNA (Matelska et al., 2017; Winther & Gerdes, 2011). PIN domains have been found in conjunction of the TRAM RNA-binding domain in the protein VapC (Matelska et al., 2017). In prokaryotes, the VapBC complex constitutes a toxin-antitoxin complex which controls translation. The VapB inhibitor is co-expressed to offset the toxicity of the VapC ribonuclease (Arcus et al., 2011; Winther & Gerdes, 2011). When located on plasmids, toxin-antitoxin systems enable plasmid maintenance, the role of chromosomally encoded TA system is more uncertain. They can play a role in altruistic cell death and growth regulation in stress conditions (amino acid starvation, antibiotic treatment, oxidative shock for example) by altering the proteome of the cells (Arcus et al., 2011; Van Melderen & Saavedra De Bast, 2009; Winther & Gerdes, 2011).

VapB and VapC are usually adjacent in the genome, as such, it could be hypothesized that, in *A. woodii*, the transcriptional regulator (Awo_c03580) upstream of Awo_c03590 is the VapB component in the system. If Awo_c03590 (potential VapC) is detrimental to growth on gas, the VapB inhibitor (potentially Awo_c03580) is expected to be essential, which was not the case in *A. woodii* in all the conditions tested, refuting this hypothesis.

The exact function of Awo_c03590 remains unclear, however, it does appear to play a role in RNA metabolism, directly upstream is a transcriptional regulator of the CarD family and a transcriptional accessory protein Tex, while downstream is a ribosomal RNA methyltransferase F. However, they did not display the same pattern of reads abundance evolution over the course of the fermenters and were not found to be essential.

Generation of a deletion strain could confirm that mutants harbouring the insertion did not benefit from metabolites produced by the rest of the library. Study of its potentially complex phenotype could help elucidate its function in *A. woodii*.

4.2.4.3.2 Awo_c14190 and Awo_c14200

Genes Awo_c14190 and Awo_c14200 also had high fold change values of 41.23 and 33.05, respectively. While there was a clear pattern between heterotrophic and autotrophic fermentations (Figure 4.30) the really low number of single mutants affected the certainty of any conclusions drawn. They are annotated as glutaredoxin / thioredoxin and hypothetical protein, for Awo_c14190 and Awo_c14200, respectively. None of the surrounding genes have a similar reads abundance pattern. Gene Awo_c14200 contains a PC4 domain, also referred to as a transcriptional coactivator p15. Considering insertions in both genes lead to similar increase in reads abundance, it can be hypothesized that Awo_c14200 is a transcriptional regulator of gene Awo_c14190.

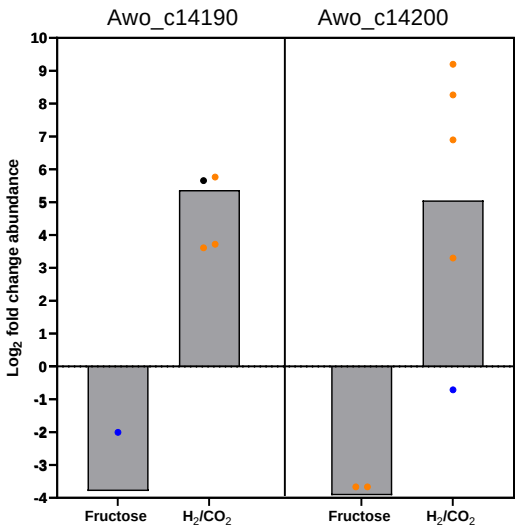


Figure 4-32: Analysis of Awo_c14190 and Awo_c14200 during H_2/CO_2 fermentation. Log_2 fold change of abundance of single mutants, bars represent the gene-level log_2 fold change of reads abundance. Mutants with a starting number of reads ≤10 and ≤20 are represented in orange and blue respectively, mutants starting with a number of reads >20 are represented in black.

4.2.4.3.3 NeuBCA

The abundance analysis at gene level revealed that reads in genes *neuB* (awo_c24840) and *neuC* (Awo_c24830) become more abundant during the course of the H_2/CO_2 fermentation. They are flanked by *neuA* and Awo_c24850, while reads in Awo_c24850 also become more abundant, reads in *neuA* diminished. Genes *neuB* and *neuC* form an operon with gene *neuA* (awo_c24820), with an overlap between each of the three genes (Figure 4.31A). Interestingly, the abundance of reads in those genes seem to decrease during fructose fermentation. The gene level data seems to point towards the operon being detrimental to gas fermentation but important for fructose fermentation.

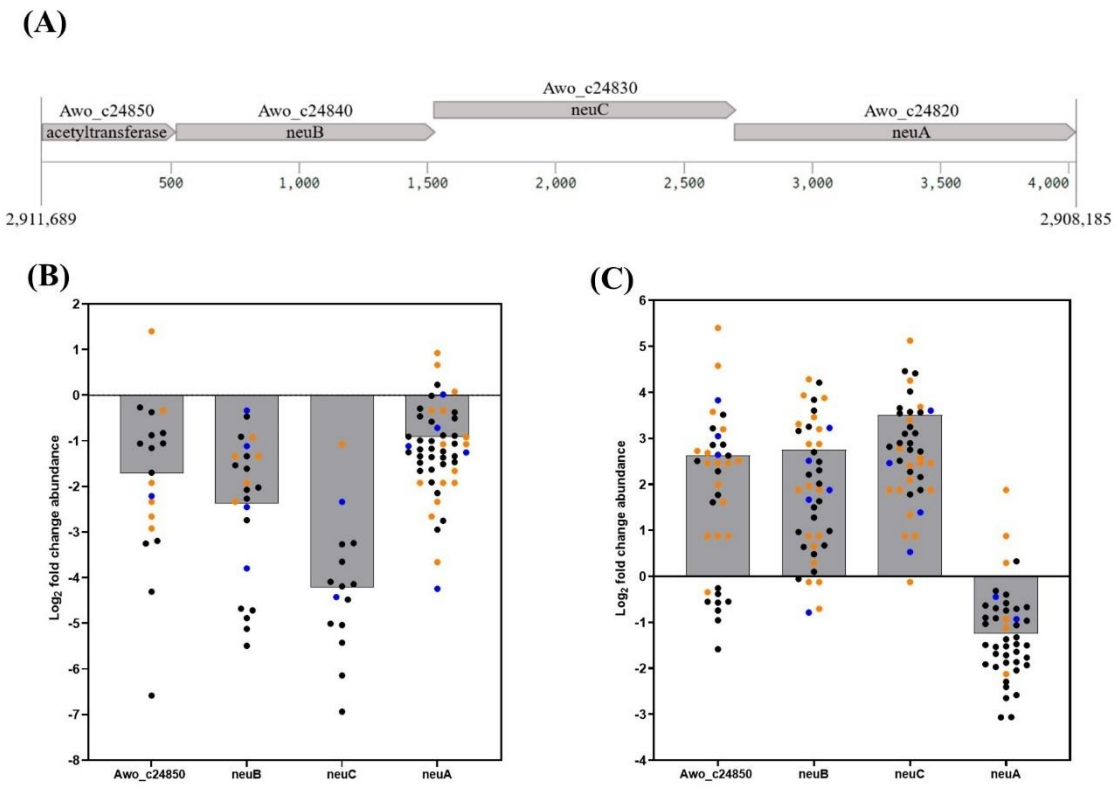


Figure 4-33: Genes Awo_c24820 to Awo_c24850 mutant abundance in fructose and H_2/CO_2 fermenters. (A) Gene organisation of *A. woodii* genes Awo_c24820 to Awo_c24850, genes are in reverse direction between the positions 2,908,185 and 2,912,210. Log2 fold change of abundance during **(B)** fructose fermentation and **(C)** H_2/CO_2 fermentation. Bars represent the gene-level log_2 fold change of reads abundance. Mutants with a starting number of reads ≤10 and ≤20 are represented in orange and blue respectively, mutants starting with a number of reads >20 are represented in black. **(D)** Log2 fold change of abundance of single mutants in gene Awo_c24850 plotted as a function of their position in the gene. Gene Awo_c24850 is situated between positions 2,911,692 and 2,912,210 in reverse orientation. Mutants with a starting number of reads ≤10 and ≤20 are represented in red and orange respectively, mutants starting with a number of reads >20 are represented in black

The single mutant data for genes Awo_c024820 to Awo_c24850 confirmed the gene level fold change of abundance (Figure 4.31B and C). It is interesting to note that gene Awo_c24850 has two clear populations of single mutants, with either a positive or a negative log_2 fold change of abundance. The position of these single mutants in the gene could not be explained by their position in the gene.

Awo_c24850 is annotated as an acetyltransferase of the GNAT family. According to InterPro records, *neuCBA* are involved in the biosynthesis of CMP-N-acetylneuraminate (sialic acid) from UDP-N-acetylglucosamine. However, the KEGG pathway identifies CMP-N,N'-diacetyllegioaminate as a final product. It can be hypothesized that Awo_c24850 corresponds to enzyme EC 2.3.1.203 from the Uridine diphosphate sugar metabolism, converting UDP-N-acetylbacillosamine to UDP-N,N'-diacetylbacillosamine. Genes *neuC, neuB* and *neuA* form the remainder of the pathway with CMP-N,N'-diacetyllegioaminate as a final product.

4.2.4.3.4 Miscellaneous genes

Miscellaneous genes with gene-level reads abundance fold change higher than 4 were analysed at the single mutants level during H_2/CO_2 fermentation and at the gene-level during fructose fermentation (Figure 4.32). Genes Awo_c06020, Awo_c03380, and Awo_c2000 presented $\log_2$ fold changes of 2.23, 2.20, and 3.05, respectively, but really low number of mutants, they were excluded from Figure 4.32.

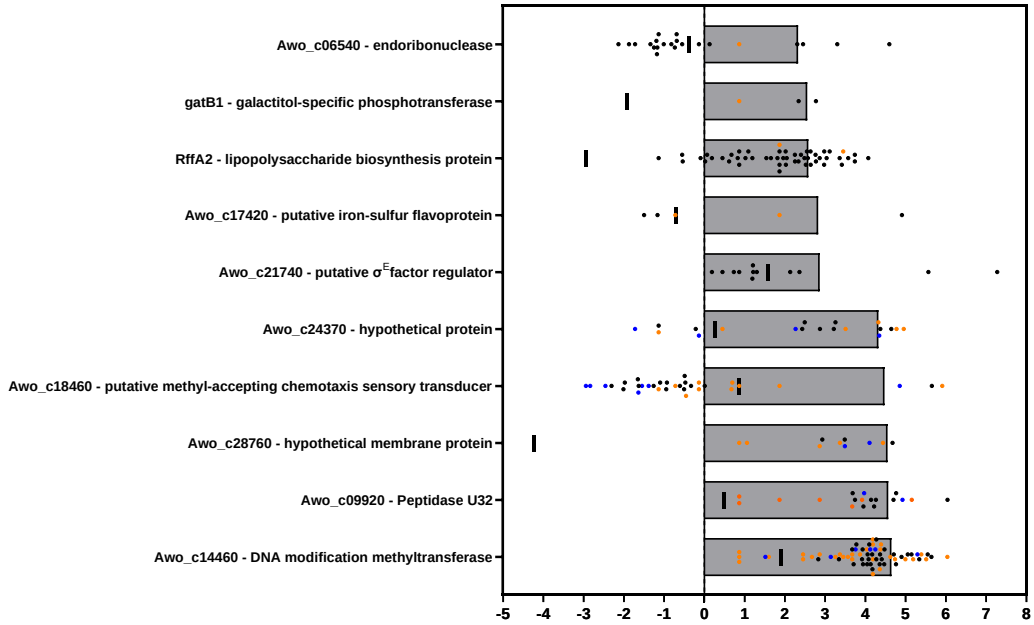


Figure 4-34: Miscellaneous genes mutant abundance in H_2/CO_2 fermenter. Bars represent the gene-level log_2 fold change of reads abundance during H_2/CO_2 fermentation, lines represent the log_2 fold change during fructose fermentation. Individual values are the log_2 fold change of abundance of single mutants; mutants with a starting number of reads ≤10 and ≤20 are represented in orange and blue respectively, mutants starting with a number of reads >20 are represented in black.

Gene Awo_c28760 has a low number of single mutants, however its gene-level log_2 fold change of reads abundance in H_2/CO_2 is very low. Gene *rffA2* has a high number of single mutants and significantly different log_2 fold changes in the fructose fermenter compared to the H_2/CO_2 fermenter. It can be hypothesized that the roles of these genes are detrimental to H_2/CO_2 fermentation but important for fructose fermentation.

Genes Awo_c14460 and Awo_c09920 had a smaller difference in log_2 fold change in the fermenters, the study of single mutants revealed a high number of single mutants with a consistently high log_2 fold change. They are likely to play a role in the fructose fermenter even if this role is not detrimental to H_2/CO_2 fermentation.

New single mutants emerged in the population as follows: 4 in Awo_c14460, 3 in Awo_c18460, and 1 in Awo_c21740 and Awo_c24870. In contrast, genes Awo_c21740, Awo_c17420, and Awo_c06540 had gene-level log_2 fold change values different in the two fermenters. However, the single mutant level analyses revealed only a few mutants responsible for the high log_2 fold change. Those genes are unlikely to play a definite role specific to the H_2/CO_2 fermenter.

Finally, the study of single mutants in Awo_c18460 showed that a bias was created by a few mutants, but it also revealed emergence of new mutants during fermentation, its role remains unclear.

4.3 General discussion

4.3.1 Residual plasmid presence and mutant emergence

Reads mapping to the pMTL-GL15 plasmids were observed in all but one of the samples sequenced (section 4.2.2.3). Since the plasmid did not contain a Gram-positive replicon, it was hypothesized that the whole plasmid was integrated in the *A. woodii* genome. Two explanations were put forward for the apparent decrease of the plasmid: it lowered cell fitness or led to transient transposition events during fermentation.

At the same time, new mutants with insertion in genes found to be detrimental emerged from the population. Once again two hypotheses can explain this phenomenon: they were undetected in the earlier samples because of their low frequency or new transposition events had occurred.

It is possible that genome-based transposase expression occurred and led to the concomitant decrease of plasmid reads and increase of new mutants in the library. However, the other hypotheses cannot be ruled out.

Regardless, in the context of our experiment, this phenomenon proved to be advantageous as it enabled identification of highly detrimental genes. Genes displaying emerging single mutants represent ideal candidates for study using directed mutational tools. Some of the most attractive targets are *mttC7* for fructose fermentation and both the alanine pathway and Awo_c03590 for H_2/CO_2 fermentation.

4.3.2 Combination of essentiality and abundance study

Essentiality studies enabled confirmation of the genes crucial for heterotrophic and autotrophic fermentation by *A. woodii*. Genes involved in the WLP, energy conservation, and carbon metabolism of *A. woodii* were studied in detail and, for most

genes, the results corroborated the transcriptomic and proteomic studies performed (Moon et al., 2021; Neuendorf et al., 2021; Shin et al., 2021).

In the HDCR, it was interesting to note that the formate dehydrogenase catalytic subunit appeared duplicated with compensation mechanisms likely in place when *fdhF2* or *hycB2* are disrupted. The non-essentiality of the Rnf complex was puzzling as it conflicts with the current knowledge of energy conservation in *A. woodii*.

As hypothesized, the abundance study was extremely valuable as it enabled confirmation of the importance of genes found to be non-essential and led to the identification of detrimental genes under the conditions studied. Notably, several subunits of the F_1F_o-ATPase were found to be non-essential but the dramatic decrease in abundance of mutants with transposon insertion in their coding sequence confirmed their crucial role under autotrophic conditions.

Other genes were highly detrimental to either the fructose or the gas fermenter, demonstrating metabolic burden in the respective condition. A study in *S. pneumoniae* demonstrated that 6% of the genotype-phenotype interactions identified were detrimental for growth (van Opijnen & Camilli, 2012). *A. woodii* cannot co-utilize fructose with glycine betaine (Lechtenfeld et al., 2018) and expression of the glycine betaine pathway appeared detrimental to heterotrophic growth in the present study. Similarly, alanine production using the dehydrogenase was found to be extremely detrimental to autotrophic growth. It is an interesting observation that no regulatory processes are in place for these genes to increase strain fitness when their expression is not required. This can also be said about the *mttC7* gene and Awo_c03590, in the case of *mttC7*, its hypothesized transcriptional regular was also found to be detrimental.

In the present study, the pivotal role of transcriptional regulators was demonstrated by the lactate dehydrogenase repressor LctA. Essentiality studies confirmed that expression of the lactate dehydrogenase was detrimental for the cells in the absence of lactate. The role of the repressor was thus crucial to strain fitness.

The laboratory conditions in which the transposon library was grown were quite different from the natural growing conditions of *A. woodii*, the observations made and conclusions drawn should take this fact into consideration. It is likely that the microbial communities and substrates they encounter would have never resulted in such a decrease in strain fitness.

For most of the detrimental genes identified, their function or impact on fermentation remains elusive. The study of those detrimental genes provides an exciting new avenue for a better understanding of gene networks in *A. woodii* and strain improvements for biotechnology applications.

Interestingly, the combination of essentiality and abundance studies provided new insights into the electron bifurcation mechanism in *A. woodii*. Electron bifurcation is accomplished by both the Hyd and Stn complexes. Hyd has been shown to be the main contributor for electron bifurcation and works in opposite direction in heterotrophic and autotrophic conditions (Bertsch & Muller, 2015a; Schuchmann & Muller, 2016). During heterotrophic growth, the Hyd utilizes NADH and reduced ferredoxin for the hydrogen production required to re-capture the CO_2 generated by substrate level phosphorylation (Schuchmann & Muller, 2016). In contrast, during autotrophic growth, Hyd oxidizes hydrogen to produce NADH and reduced ferredoxin (Bertsch & Muller, 2015a). The data presented here indicates that while the Stn complex was found to be essential in both conditions, Hyd was found to be important in heterotrophic conditions but

detrimental in autotrophic conditions. These data showed that Hyd cannot compensate for the loss of function of Stn and led to the hypothesis that without the requirement for H_2 production of heterotrophic conditions, deletion of Hyd is compensated by Stn and is beneficial for growth in autotrophic conditions. In *A. woodii*, the study of the Stn complex offers a promising avenue for further understanding of redox balancing and compensation mechanisms between the electron bifurcating complexes. This observation also calls into question the lack of regulation for Stn and Hyd, one would expect that such crucial functions of the cellular processes would be activated in the most adequate conditions.

Maintenance of intracellular pH, despite the proton imbalance caused by high acetate concentrations, was found to be a prominent factor of mutant fitness, especially during autotrophic growth. Degradation of arginine into citrulline and the linked production of NH_3 was found to be essential during both heterotrophic and autotrophic growth. As *de novo* pyrimidine synthesis was found to be essential during fermentation, the essentiality of the arginine deaminase pathway could have also been linked to the production of carbamoyl-phosphate and subsequent UMP synthesis. During autotrophic growth, a V-Type ATPase was found to be detrimental and hypothesized to translocate protons. This ATPase has never been studied in *A. woodii* but could provide further insights into ATP generation and maintenance of pH homeostasis. If the detrimental effect of the Hyd complex during autotrophic growth is due to proton generation, it can be hypothesized that deletion of the Hyd complex would mitigate the detrimental effect of the V-type ATPase.

4.3.3 Limitations

Overall, the data presented here allowed for a more in-depth understanding of *A. woodii* metabolic networks during heterotrophic and autotrophic growth and enabled the identification of targets for future gene editing. It is, however, important to consider some of the limitations of the method, they are discussed below.

First, construction of the library was undertaken on fructose, all the mutants with insertion in genes essential for fructose growth were eliminated at an early stage, introducing a bias in the insertion sites. As a result, genes essential to growth on H_2/CO_2 cannot be isolated from genes essential to fructose growth. Similarly, genes that would be detrimental to gas growth but essential to fructose growth cannot be discovered. Abundance studies mitigate this to some extent but to dissociate gene essentiality on different substrates it would be necessary to build a different library on each of the selected substrate. For gas growth, recoveries could be plated without fructose and placed in an anaerobic jar with adequate gas mix. Achieving high mutant numbers would likely constitute a logistical challenge. Construction of separate libraries for each substrate would also restrict direct comparison. This bias is, however, important to take into consideration when analysing results from transposon libraries.

Second, for transposon experiments of static growth in static bottles, triplicates are performed to ensure the reproducibility of the results obtained. This was not feasible for the fermentations at the time of the study for logistical reasons. The absence of replicates might call into question the results as a small variation in media composition might result in significant changes in the bottlenecks applied to the system.

Finally, social cheating has been described in microbial communities (García-Contreras & Loarca, 2020; Smith & Schuster, 2019), the survival of some of the insertion mutants

might have only been possible because of the rest of the population. For genes found to be detrimental but with an unknown function, social cheating can be an important potential limitation, as their deletion might result in growth defects when grown in pure cultures.

4.4 Key outcomes

This study enabled the identification of the core essential gene set for growth in rich medium and essential genes specific to heterotrophic and autotrophic growth. The study of abundance of mutants over time was complementary to the essentiality study, it provided two major advantages: (i) confirmation of gene essentiality and (ii) discovery of detrimental genes for the specified conditions.

The detrimental effect of the Hyd hydrogenase during gas fermentation revealed potential beneficial compensation by the Stn system for redox balancing. The detrimental effects of alanine production, V-Type ATPase, and a potential nuclease (Awo_c03590) were also observed and provide exciting new avenues for improvement of gas fermentation of *A. woodii*.

During fructose fermentation, degradation of glycine betaine and another unknown molecule, potentially degraded by MttC7, were found to be detrimental.

Overall, while transcriptomic and proteomic studies enabled an understanding of gene regulation in different conditions, the present study revealed detrimental pathways and compensation systems active during fermentation and identified several potential targets for directed genome editing that would improve *A. woodii* as an industrial chassis.

CHAPTER 5: Towards improvements of solvent yields in *A. woodii*

5.1 Introduction

5.1.1 Acetone production in *A. woodii*

Acetone production in *Acetobacterium woodii* was established in 2016 with the introduction of an acetone operon formed of the *adc*, *ctfA/B* and *thlA* genes (Figure 5.1A) of *C. acetobutylicum* under the control of the promoter of the *C. acetobutylicum* thiolase gene (Hoffmeister et al., 2016). A maximum yield of 15.2 mM of acetone was produced in serum bottles under autotrophic conditions when the operon was localised to a plasmid with the pIP404 Gram-positive replicon. The impact of Gram-positive replicons was studied, while pBP1 and pCD6 yielded a comparable amount of acetone than pIP404, pCB102 results in approximately half the acetone yield. Promoters also had an important impact, the *pta-ack* promoter from *C. ljungdahlii* resulted in similar acetone yields as the *thlA* promoter from *C. acetobutylicum* whereas the *ackA* promoter from *A. woodii* resulted in approximately half the yield. Continuous gas fermentation (H_2 + CO_2) of the recombinant strain resulted in a productivity of 26.4 mg/L/h with media being fed at 0.3 to 0.6 mL/min.

Within our group, acetone production was achieved in *A. woodii* with the same acetone operon from *C. acetobutylicum* (Baker, 2020; Hoffmeister et al., 2016). In the expression plasmid, the operon was placed downstream of the *C. difficile tcdB* promoter (P_{tcdB}) that is recognised by the TcdR sigma factor. The plasmid also contained the ColE1 *E. coli* Gram-negative replicon, the Gram-positive replicon pCD6, and a thiamphenicol selection marker (pMTL24-1, Figure 5.1B).

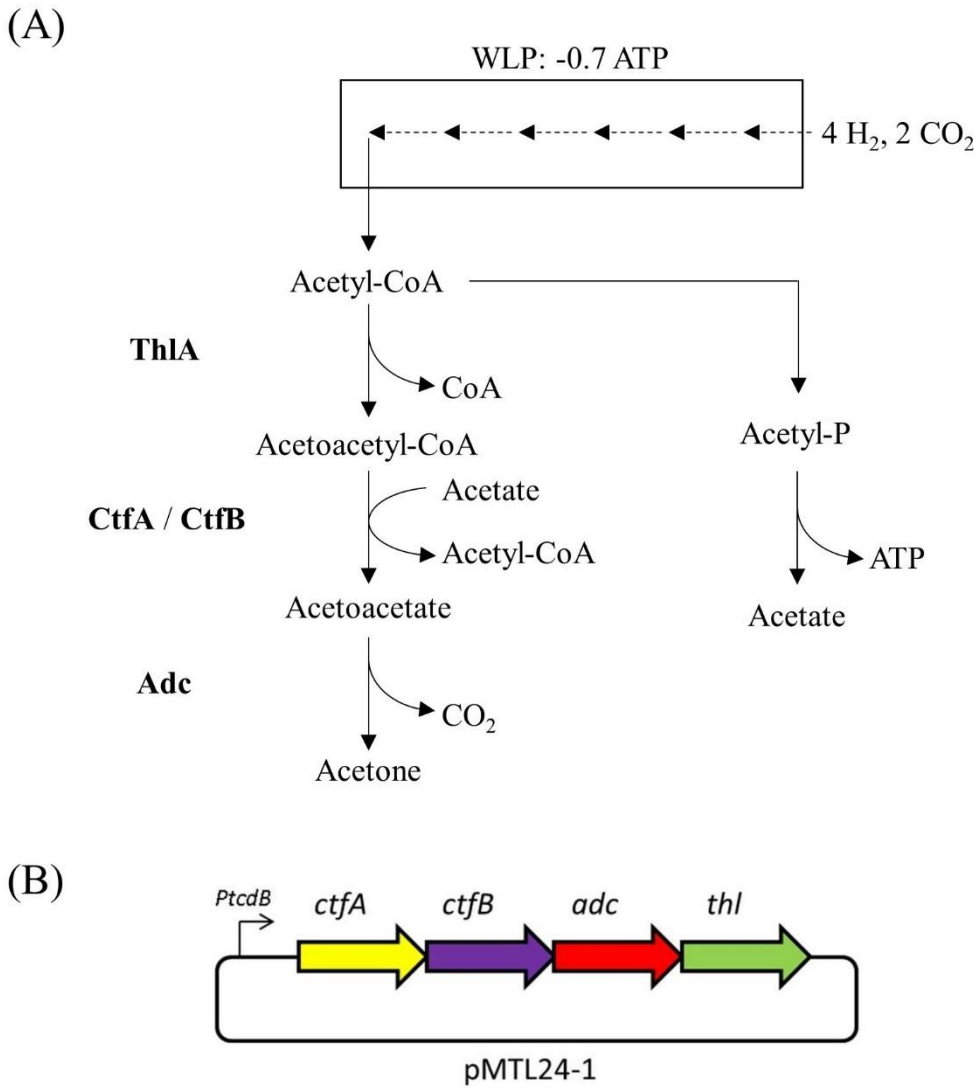


Figure 5-1: Acetone production from acetyl-CoA in the acetogen *A. woodii*. (A) Acetyl-CoA is synthesized from the Wood-Ljungdahl Pathway (WLP). A native operon from *C. acetobutylicum* comprised of a thiolase (Thl), an acetoacetyl-CoA:acetate/butyrate CoA transferase (CtfA/CtfB), and an acetoacetate decarboxylase (Adc) allows for synthesis of acetone from acetyl-CoA (Hoffmeister et al., 2016). **(B)** pMTL24-1 vector used to produce acetone in *A. woodii* in the SBRC laboratory (Baker, 2020).

An *A. woodii* strain with the *tcdR* gene inserted in the genome at the *pyrE* locus was transformed with the pMTL24-1 plasmid. The resulting strain yielded 18 mM of acetone when grown in an autotrophic fed batch reactor (Baker, 2020), comparable to that achieved by Hoffmeister et al. (2016) under similar conditions.

The thiolase gene is thought to be the bottleneck for acetone production. Accordingly, gene organisation of the operon was modified to have the thiolase gene in first position for potential higher expression. Concomitantly, alternative thiolase genes from other bacterial species were tested for improved acetone yields. Both strategies resulted in lowered acetone yields when compared to the original pMTL24-1 plasmid (Baker, 2020), with some thiolase genes resulted in the complete abolishment of acetone production.

A subsequent strategy involved codon harmonisation based on *A. woodii* codon usage. It has been shown to improve production through improving protein translation and folding (Mignon et al., 2018). In *A. woodii*, this technique was enabled by a Java script developed by Dr. Gareth Little. The genes from the acetone operon were codon harmonized and replaced in the pMTL24-1 plasmid separately in order to assess their isolated impact on acetone production. Attempts at cloning the codon harmonized version of *adc* failed and neither the *ctfAB* nor *thl* codon harmonized genes resulted in measurable acetone production (Baker, 2021).

5.1.1.1 Limitations of plasmid-based expression and genomic integration

Plasmid maintenance requires antibiotic in the growth medium, which represents strain stability concerns and additional costs for industrial purposes. As a result, plasmid-based production cannot be applied to commercial strain and for scale-up purposes. Moreover, the expression of a fluorescent reporter gene on an autonomous plasmid was shown to lead to an heterogenous population in *C. acetobutylicum* (Streett et al., 2019). This heterogeneity was dependent on the growth stage and the first hypotheses for its cause were plasmid loss and difference in cell morphogenesis in sporulating bacterial populations (Streett et al., 2019). The former was invalidated by re-culturing of the high

and low fluorescence populations in selective media, both led to indistinguishable populations (Streett et al., 2019).

To circumvent plasmid-based expression of gene of interest, genomic integration is required. In *A. woodii*, integration of the acetone operon from the pMTL24-1 plasmid at the *pheA* locus resulted in a significant decrease of acetone yields compared to expression from the plasmid (Baker, 2020). This result was expected and can be explained by the lower copy number of genes when integrated in the genome.

In the literature, along with lack of genome engineering tools, loss of yield is the main hurdle for genomic insertion of pathways of interest associated with the generation of efficient industrial strains.

5.1.1.2 Conversion of acetone into isopropanol

During characterisation of the acetone producing strain in autotrophic conditions, isopropanol was detected by HPLC and constituted potential evidence of native conversion of acetone to isopropanol in *A. woodii* (Baker, 2021). This is not unprecedented, as *C. beijerinckii* was shown to possess a secondary alcohol dehydrogenase (SADH) capable of the conversion of acetone to isopropanol (Hanai et al., 2007). *A. woodii* encodes for 11 alcohol dehydrogenases and only a few have been functionally characterised (Bertsch et al., 2016). It is possible that one of those alcohol dehydrogenases (Adh) naturally converts acetone to isopropanol. The ENGICOIN report showed that, after the addition of 25 mM of acetone in the growth medium, growth of *A. woodii* WT did not have a significant impact on acetone concentrations, as quantified by HPLC. Additionally, differences in acetate and acetone yields in the WT compared to the acetone producing strain led to the conclusion that no additional carbon

products were being produced and that the isopropanol detected was an artefact of the HPLC method (Baker, 2021).

However, acetone and isopropanol co-elute with the HPLC technique that was employed for acetone quantification (Baker, 2021). As such, acetone to isopropanol conversion could still constitute a significant loss of yield for acetone production and may still benefit from knock-out of the SADH responsible for the conversion.

Expression of the known SADH from *C. beijerinckii* in the acetone producing *A. woodii* strain could help better understand acetone to isopropanol conversion and enable validation of an HPLC technique for isopropanol measurement. As isopropanol production requires only one additional gene compared to acetone production, its expression could also be used to study the impact of gene integration locus in the genome on production yields.

5.1.2 Random knock-in for identification genome loci favourable for expression of integrated genes

In *A. woodii*, only a two insertion sites have been used for gene expression, namely *pheA* and *pyrE*. Accordingly, it would be beneficial to identify other genomic loci which may result in more favourable gene expression and therefore higher yields of products.

It is generally accepted that the level of expression observed when genes are integrated in the genome can be affected by the distance from the origin of replication. Traditionally, a handful of insertion sites are tested for genomic gene expression to account for local effects. However, recent analysis of randomly inserted reporters in *E. coli* has shown tremendous variations of transcription across the genome (Bryant et al., 2014; Scholz et al., 2019). When randomly inserted, a reporter cassette comprised of a *gfp* gene under the control of a *lac* promoter cassette, showed gene expression

variations of up to 300-fold depending on genomic position (Bryant et al., 2014). This study hypothesized that a major part of the regulation variations is at the transcriptional level. A subsequent study looked at transcription propensity after random genome insertion. They showed that there is significant variation in transcription depending on genome position, with prophage and MGEs regions displaying the lowest transcription propensity (Scholz et al., 2019).

In 2018, a similar study was performed in *Bacillus subtilis* and certain insertion sites were found to increase up to 2.9-fold *gfp* fluorescence (Jeong et al., 2018). The authors found that expression was affected by gene knock-out, adjacent promoters, and metabolic pressures. This suggests that favourable insertion sites depend on the gene integrated, as favourable knock-out and metabolic balance depends on the protein produced.

For this reason, in 2021, Saleski and co-workers studied the random insertion of part or a whole pathway in *E. coli* for isobutanol production (Saleski et al., 2021). A biosensor strain with isobutanol-dependent fluorescence expression allowed for analysis of the mutant strains generated. The production strains were encapsulated with the biosensor strain in an emulsion and subsequently sorted through flow cytometry. This study successfully identified multiple patterns of insertions resulting in high isobutanol production. Interestingly, some mutants resulted in production levels similar to the ones obtained with plasmid-based expression of the pathway. They also showed that there is an apparent lack of correlation between level of transcription and isobutanol production (Saleski et al., 2021). This highlights the need for phenotype analyses of strains with insertions in different places in the genome instead of transcriptional analyses.

For studies in *A. woodii*, no strains are available for either acetone or isopropanol detection. As an alternative, a fluorescent reporter gene can be utilized as a first test (Bryant et al., 2014; Jeong et al., 2018).

5.1.2.1 Flow cytometry in anaerobic conditions

Random knock-in and utilization of a reporter gene for characterization of the cells entails the use of flow cytometry. Subsequent sequencing of the best-performing mutants requires survival and culturing of the cells to obtain sufficient DNA material. Dedicated anaerobic gloveboxes have been designed in the past specifically for anaerobic cell sorting (Bellais et al., 2022; Thompson et al., 2015), they require extensive logistics for their installation. Alternatively, re-culturing of *C. acetobutylicum* was successful after cell sorting in aerobic conditions (Streett et al., 2019).

5.1.2.2 Fluorescence reporter gene

Fluorescent reporters are essential for the development of synthetic biology as they allow for a range of techniques to be applied from promoter assay to protein tagging and flow cytometry as described in the previous section. Among fluorescent reporters, the green fluorescent protein (*gfp*) is the most used. However, the GFP gene is big, pH sensitive, and needs oxygen for the chromophore to mature. In conditions lacking oxygen other solutions have been put in place to be able to still use the GFP-based genes, including fixation of the cells before exposure to oxygen for fluorescence maturation or simple exposure of cells to oxygen in the case of oxygen-tolerant organisms (Gyulev et al., 2018). For the development of new tools in acetogens, fluorescent reporter genes functional in anaerobic conditions are required.

5.1.2.2.1 LOV-domain containing genes

Recently, new oxygen-independent blue-light photoreceptors have been investigated in a variety of organisms (Chapman et al., 2008; Drepper et al., 2007; Mukherjee et al., 2015). These photoreceptors contain a photoactive LOV (Light Oxygen Voltage) domain with a conserved region containing a cysteine (Krauss et al., 2009). Chromophore formation requires Flavin mononucleotide (FMN), upon irradiation with blue-light the protein undergoes a conformational change that results in the formation of an FMN-cysteine adduct with weak auto-fluorescence (Mukherjee et al., 2015).

The most used examples are the LOV domains from *Bacillus subtilis* and *Pseudomonas putida* that were the starting point of Drepper and co-workers (Drepper et al., 2007). Mutation of the cysteine into an alanine increased 10-fold the fluorescence signal even in the absence of oxygen after codon optimization and production in either *E. coli* or *Rhodobacter capsulats* (Drepper et al., 2007). In a further study Drepper and co-workers showed that the FbFP (Flavin-binding Fluorescent Protein) is quick to fold and is not dependent on oxygen availability, enabling in improved real-time and quantitative protein expression measurement compared to Yellow Fluorescent Protein (a GFP variant), especially in hypoxic conditions (Drepper et al., 2007). Multiple studies have shown the successful expression of genes Bs2 (from *B. subtilis*) and Pp1 (from *P. putida*) and good fluorescent signal in anaerobic conditions in a variety of Gram-positive bacteria including *C. acetobutylicum* (Bruder et al., 2016; Cho & Lee, 2017), *Clostridium cellulolyticum* (Li et al., 2014), *C. ljungdahlii* (Molitor, 2016), *Bifidobacteria* spp (Montenegro-Rodriguez et al., 2015), *Clostridium perfringens* (Wakabayashi et al., 2019), Listeria spp (Landete et al., 2017), *Lactococcus* spp, *Lactobacillus* spp and *Enterococcus* spp (Landete et al., 2015). In most of these articles the cells are washed and resuspended in PBS before imaging or FACS in order to lower

the background fluorescence, in *C. acetobutylicum*, however, FACS was performed directly in the growth medium (Cho & Lee, 2017).

In 2012, the phenylalanine in position 37 of Pp1 was shown to be crucial for protein brightness, with its replacement by a serine or threonine resulting in two-fold greater fluorescence. The authors hypothesized that the aforementioned changes result in a "stronger binding between the FMN fluorophore and FbFP in the mutants" (Mukherjee et al., 2012).

Another example is the LOV domain from *Arabidopsis thaliana*, that has been successfully engineered by DNA shuffling to render a more stable variant (iLOV) that also has a better recovery after photobleaching in *E. coli* (Chapman et al., 2008). The iLOV protein was further engineered to increase its photostability by rendering the protein less flexible and more compact, locking the chromophore (phiLOV variant), however, the higher photostability resulted in decreased brightness (Christie et al., 2012). Lastly the creiLOV protein identified by Mukherjee and co-workers is more stable at a wide-range of pH and thermostable up to 60ºC (Mukherjee et al., 2015).

Overall, the LOV proteins are found to be smaller than GFP, oxygen-independent and have a more stable fluorescence across a wide range of pH.

5.1.2.2.2 Fluorescence-Activating and Absorption-Shifting Tag Protein (FAST)

The fluorescence-activating and absorption shifting tag (FAST) was developed in 2018 for mammalian cells. It is a protein tag which is fluorescent only when associated with the fluorogenic ligand 4-hydroxy-3-methylbenzylidene-rhodanine (HMBR). Since the ligand is cell-permeable and binding is reversible, fluorescence is also completely reversible (Plamont et al., 2016). Further engineering resulted in variants with 3.8-fold brighter fluorescence (Tebo et al., 2018).

In *C. acetobutylicum*, while it was shown that codon-optimized phiLOV2.1 and creiLOV did not fluoresce above background, FAST resulted in high fluorescence (Mordaka & Heap, 2018; Streett et al., 2019). The fluorescence signal was adequate for microscopy, promoter assay, protein tagging and flow cytometry applications (Streett et al., 2019).

FAST is a very promising fluorescent reporter gene for the random knock-in application discussed. One potential limitation, however, is the requirement for ligand addition before measurement, potentially extending the oxygen exposure of the cells. This requirement for cell washing and addition of ligand also prevents real-time fluorescence measurements.

5.1.3 Aims

The overall aim of this study was to improve acetone production in *A. woodii*.

First, the conversion of acetone into isopropanol was studied to explore potential carbon redirection. Employing a known SADH could illustrate the impact of isopropanol production in *A. woodii*. The identification of candidates SADH through *in silico* and *in vivo* approaches was undertaken, the latter through overexpression and KO.

Secondly, position of genomic insertion was investigated as an avenue for better yields and a tool to allow for characterization of new genomic loci was developed combining a fluorescent reporter-containing transposon library and flow cytometry. For this purpose, the most important goal was the formulation of a method for fluorescence-based sorting of cells in *A. woodii*. This involved the characterization of an appropriate reporter gene and validation of the flow cytometry protocol, including cell survival and contamination. Based on the method developed, a transposon library was designed for identification of genome loci resulting in the highest fluorescence levels.

Concomitantly, the plasmid-based expression was compared to genome expression both for the acetone operon and for a fluorescent reporter gene.

Acetone and isopropanol productions were used to exemplify the impact of the choice of genome insertion for exogenous gene expression. Acetone production is well established in *A. woodii* and could lead to robust data while isopropanol production from acetone requires only one additional enzyme/gene, circumventing any potential problems related to size limitation of cargo.

5.2 Results

5.2.1 Isopropanol production with *C. beijerinckii* SADH

The SADH from *C. beijerinckii* was used to show the impact of the overexpression of the gene on isopropanol titers in the *A. woodii* strains already producing acetone. Plasmid-based expression of the SADH was achieved either at the end of the acetone operon (pMTL24-1+Adh or pAcetone+Adh) or on its own under the control of the *tcdB* promoter (pMTL84151_PtcdB_CbeSADH, also referred to as pAdh). For the former, the SADH with its native RBS was added as the last gene of the acetone operon under the control of the *C. difficile tcdB* promoter. *A. woodii tcdR* was transformed with p24-1+Adh and *A. woodii tcdR*/PtcdB_acetone was transformed with pAdh. Both parental strains served as negative controls for acetone and isopropanol production, respectively. Strain growth, acetate, acetone, and isopropanol were measured over a period of 140.5 h in serum bottles in heterotrophic conditions (Figure 5.3). The product profile at the final time point was studied after conversion of the concentrations into c-molar, to reflect the carbon content of each product (Figure 5.2).

As expected, *A. woodii* with only *tcdR* in the genome produced solely acetate, the strain with the additional acetone operon integrated in its genome resulted in a lower acetate concentration at the end of the fermentation. This decrease of acetate was accompanied by production of acetone and isopropanol. As the HPLC method employed was designed to avoid co-elution of isopropanol and acetone, this result confirmed the presence, in *A. woodii,* of an alcohol dehydrogenase natively converting acetone to isopropanol.

Addition of the SADH gene from *C. beijerinckii* in the acetone producing *A. woodii* resulted in an increase of isopropanol concentrations. However, the relative position of the gene had an impact on isopropanol production levels. The addition of the SADH

gene on an autonomous plasmid in a strain containing *tcdR* and P_{tcdB}_acetone integrated in the genome resulted in a higher isopropanol concentration than when added at the end of the acetone operon in the pMTL24-1 plasmid (7.7 mM average compared to 1.8 mM average). When added at the end of the acetone operon, the final acetone concentration remained high and the isopropanol concentration only slightly above the concentration obtained without the SADH (1.8mM average compared to 0.7mM). This demonstrated that in this strain the bottleneck for isopropanol production is likely the low expression of the SADH gene. On the contrary, when the SADH gene was expressed alone on a plasmid and acetone operon was integrated in the genome, the bottleneck for isopropanol production is the expression of the acetone operon.

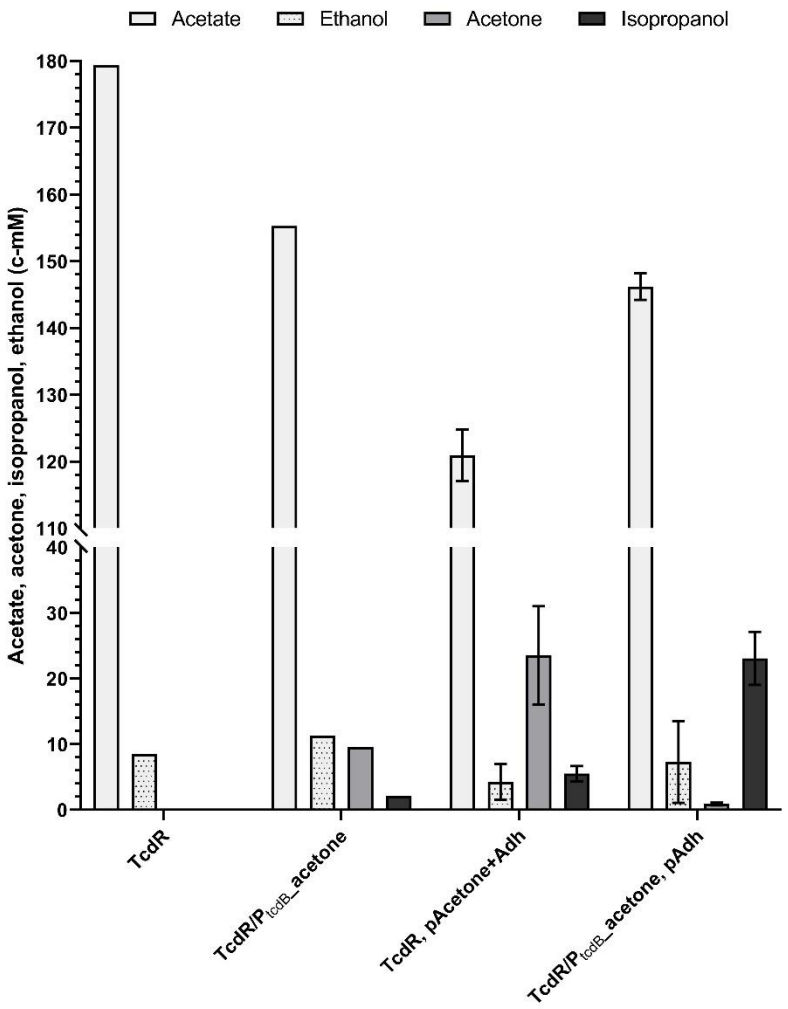


Figure 5-2: Product profiles of *A. woodii* strains with the acetone operon and/or the SADH gene from *C. beijerinckii*. Acetate (white), ethanol (white dotted), acetone (light grey), and isopropanol (dark grey) concentrations obtained with *A. woodii* strains

after 140.5 h of heterotrophic batch fermentations, expressed relative to the carbon content of each product (c-mM). Ethanol concentrations were calculated by subtracting the starting concentrations for each strain. The data represent the *tcdR* background strain with and without plasmid pAcetone+Adh and *tcdR*/P_{tcdB}_Acetone background strain with and without the pAdh plasmid. Strains containing the SADH gene were studied in duplicate.

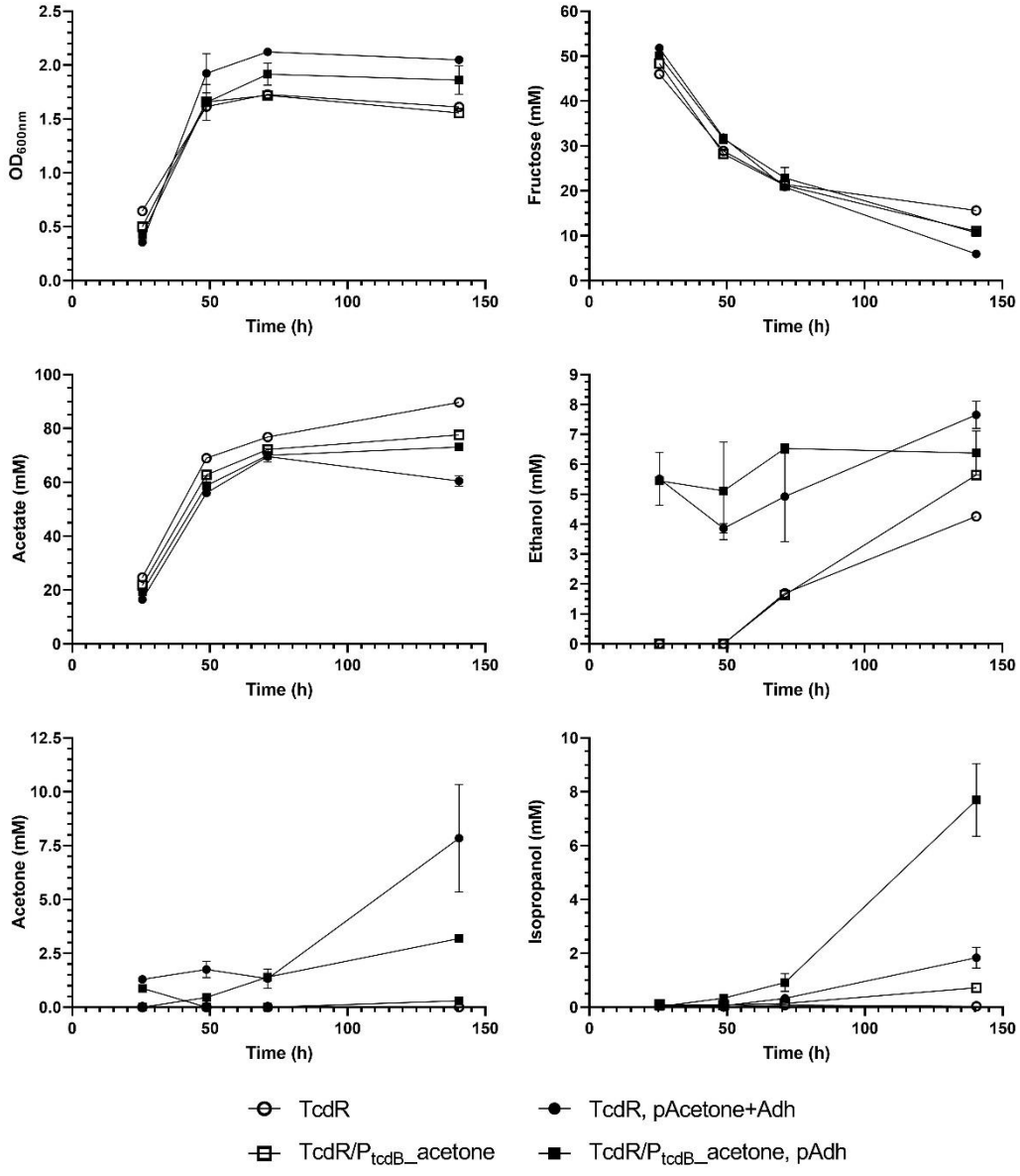


Figure 5-3: Growth and product profiles of *A. woodii* strains with the acetone operon and/or the SADH gene from *C. beijerinckii*. OD_{600nm}, acetate, ethanol, acetone, and isopropanol concentrations of *A. woodii* strains during heterotrophic batch fermentations in serum bottles over 140.5 h. Strains with *tcdR* (circle: ● and ○) or *tcdR* and P_{tcdB}_Acetone (square: ■ and □) integrated in the genome were analysed in the

absence (empty: ○ and □) or presence (full: ● and ■) of the SADH gene from *C. beijerinckii.* Strains containing the SADH gene were studied in duplicate.

As observed above, addition of the SADH gene at the end of the acetone operon increased isopropanol concentration minimally. As a result, comparison of the *tcdR*/P$_{tcdB}$_acetone strain with the *tcdR* strain containing pMTL24-1+Adh can inform on the impact of genome integration on yields. As expected, plasmid-based expression of the acetone operon, even in the presence of the SADH gene at the end of the operon, resulted in higher acetone concentrations than when integrated in the genome. A strain with pMTL24-1 alone would allow for precise quantification of the difference in production. However, expression of the acetone operon on a plasmid resulted in the highest fructose consumption, highest acetone concentration and lowest acetate concentration.

Comparisons of the different products over time showed that acetate was produced during exponential phase, peaking after 70 h. Acetone and isopropanol production started at 70 h and did not reach a plateau at the end of the experiment. Considering there is fructose remaining in the cultures, it can be hypothesized that acetate reaches its limit of toxicity, production continues at a lower rate and part of it gets converted to acetone and isopropanol. This is demonstrated with the *tcdR* strain containing the pAcetone+Adh plasmid, acetate concentration decreased as acetone concentration increased.

Ethanol was also detected in all the samples. Thiamphenicol in 50% ethanol was added, when required, for plasmid maintenance. This can explain the starting concentrations of ethanol obtained for strains containing a plasmid (between 4.9 and 6.1 mM). For those strains the ethanol concentrations were quite variable between the two replicates and no clear increase was observed over time. An increase in ethanol concentration was

observed for both the *tcdR* and *tcdR*/P_{tcdB}_acetone strains after 50 h of culture and reached 4.3 mM and 5.6 mM, respectively.

Overall, plasmid-based expression resulted in higher yields than genome integration, but the latter is needed for the development of industrial strains. Accordingly, high-throughput characterization of new genomic loci is necessary for the potential discovery of loci with improved productivity, ideally similar to the productivity obtained with plasmid-based expression.

5.2.2 Redirecting the carbon flux for improvement of acetone yields

5.2.2.1 *In silico* identification of enzymes responsible for acetone to isopropanol conversion in *A. woodii*

As demonstrated in the previous section, redirection of the carbon flux away from isopropanol is a promising avenue for increased acetone production. *A. woodii* has 11 identified alcohol dehydrogenases, most of them have no assigned function. To identify a candidate SADH for acetone to isopropanol production, homologies to known enzymes were explored.

C. autoethanogenum has been shown to possess an alcohol dehydrogenase (CAETHG_0553) capable of converting acetone to isopropanol (Kopke et al., 2014). *C. beijerinckii* was also shown to convert acetone to isopropanol and its Adh was purified and characterized (Ismaiel et al., 1993). The function of *C. beijerinckii* SADH was further confirmed by the successful expression of its gene and the production of isopropanol in *E. coli* (Hanai et al., 2007). The SADH from *C. autoethanogenum* and *C. beijerinckii* are extremely similar, their BLAST alignment results in an e-value of zero. When searched against *A. woodii* proteins, they matched solely with a galactitol-

1-phosphate 5-dehydrogenase (gatD, Awo_c01070) with e-values of 3e-26 and 5e-26, respectively (Figure 5.4).

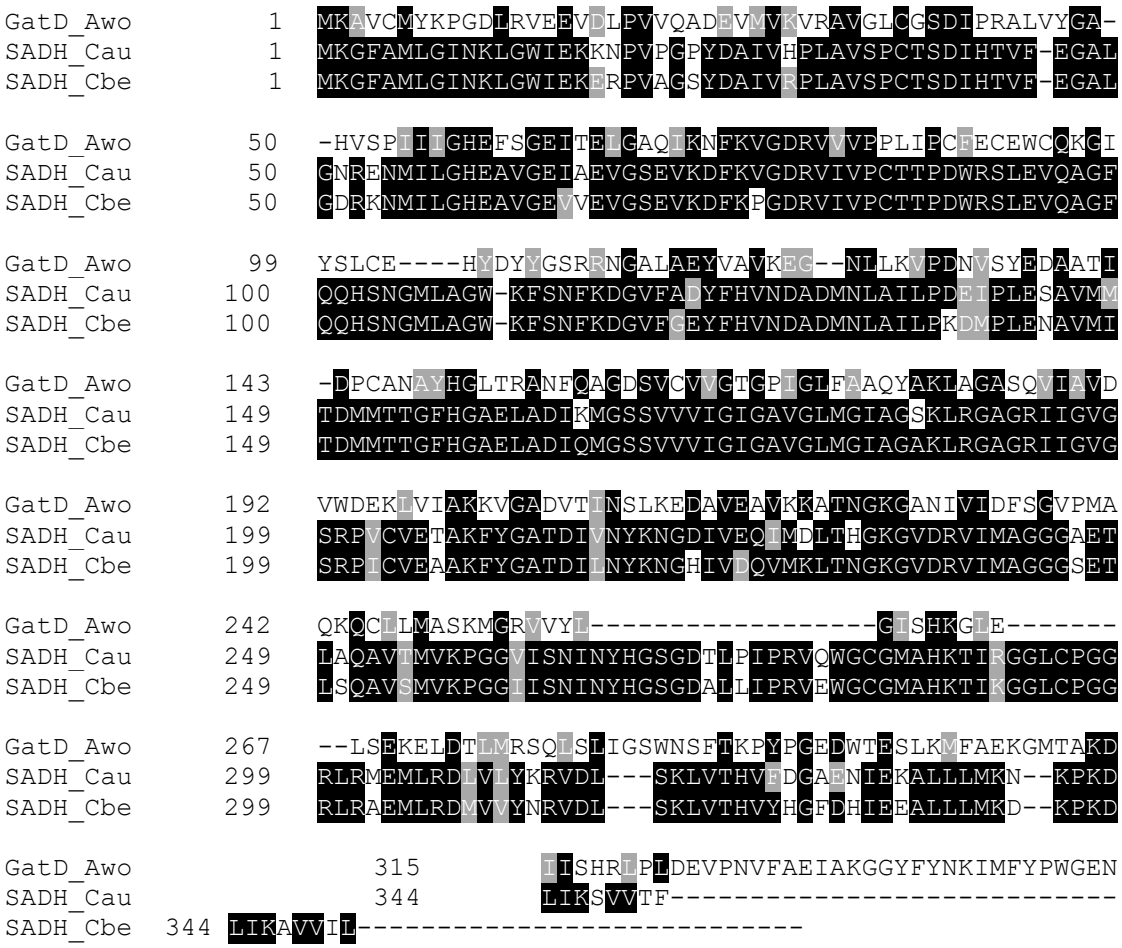


Figure 5-4: Protein similarity between *A. woodii* GatD and the SADH from *C. autoethanogenum* (Cau) and *C. beijerinckii* (Cbe). Protein alignment performed by T-Coffee (Notredame et al., 2000) and shading performed by Boxshade.

In *E. coli* K-12 galactitol-1-phosphate 5-dehydrogenase is involved in galactitol metabolism; the 2-alcohol group of L-galactitol 1-phosphate is oxidized into a keto group to form D-tagatose 6-phosphate by GatD in a NAD^+ dependent manner (Mayer & Boos, 2005). Additionally, the galactitol dehydrogenase from *Rhodobacter sphaeroides* D was found to be able to use secondary alcohols and their corresponding ketones as substrate (Carius et al., 2010).

Since protein alignment of GatD against the two SADH revealed some conservation of residues, but no extended identical domains (Figure 5.4), functional domains of the three proteins were studied. In *A. woodii*, *gatD* is identified as responsible for D-sorbitol to D-fructose interconversion. Comparison of the domains of *gatD* from *A. woodii* and the SADH from *C. beijerinckii* and *C. autoethanogenum* in the InterPro database (IDs H6LF11, P25984 and U5RQB1 respectively) revealed high similarities (Figure 5.5, Supplementary Figure 5 and Supplementary Figure 6. Although GatD is one amino acid shorter than the two SADH, they all contain oxidoreductase functions with a zinc-binding domain, a NAD(P)-binding domain and alcohol dehydrogenase-like domains. The *gatD* gene appeared as an interesting candidate for overexpression and knock-out in *A. woodii.*

5.2.2.2 Study of function of the candidate

5.2.2.2.1 Gene overexpression

The easiest way to assess the impact of the *gatD* gene is to overexpress it in *A. woodii,* in a strain already producing acetone and compare it to the results obtained in the previous section with the secondary alcohol dehydrogenase from *C. beijerinckii*. *A. woodii tcdR*/P_{tcdB}_acetone was transformed with pMTL84151_P_{tcdB}_AwoGatD (pGatD), the resulting strain was compared to the strain containing pAdh studied in section 5.2.1. The parental strains *A. woodii tcdR* and *A. woodii tcdR*/P_{tcdB}_acetone were used as controls for acetone and isopropanol production, respectively. Since acetone and isopropanol concentrations were still increasing after 140.5 h in the previous experiment, fermentation time was extended to over 230 h. Growth and production overtime (Figure 5.6) and final concentrations overtime were represented (Figure 5.7).

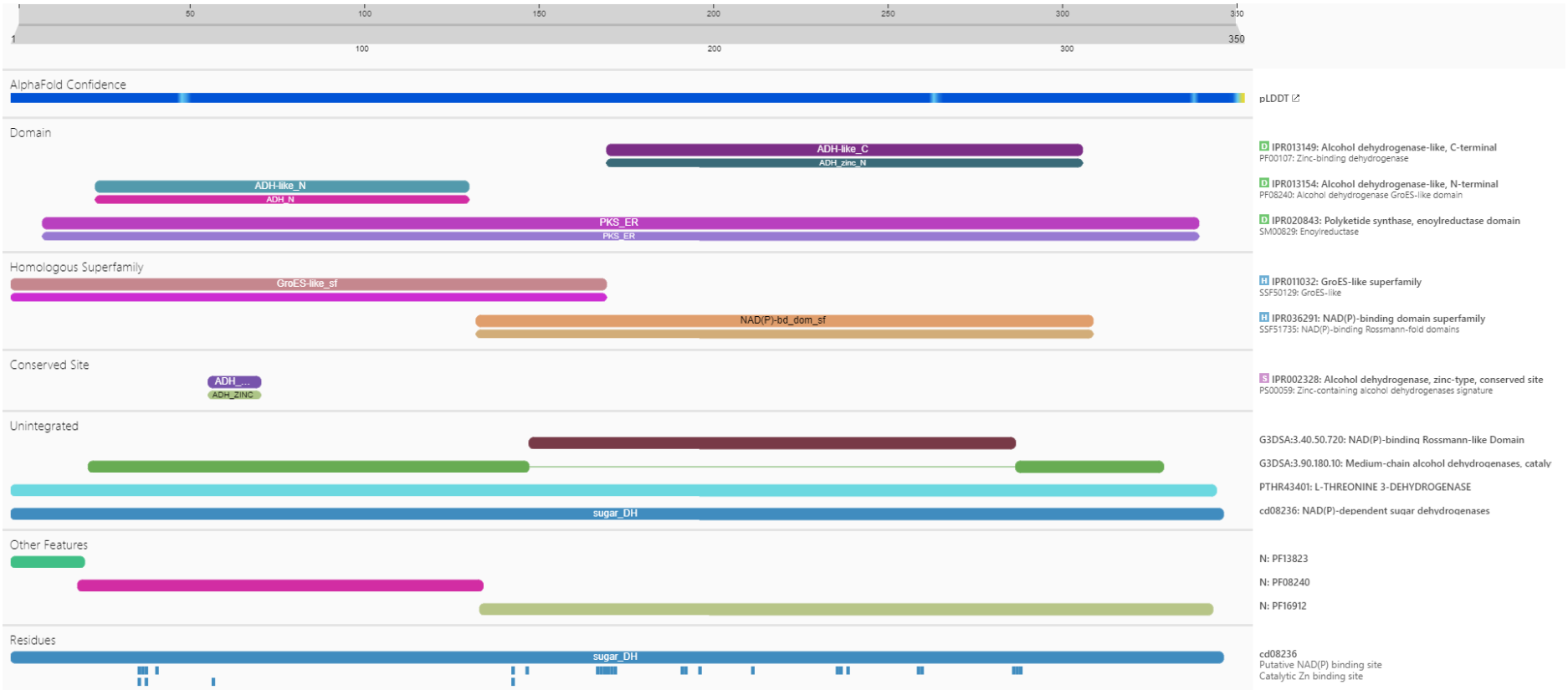


Figure 5-5: InterPro analysis of *A. woodii* GatD protein. The protein contains alcohol dehydrogenase domains, including a zinc-binding domain and an NAD(P)-binding domain. The InterPro analyses of the SADH from *C. autoethanogenum* and *C. beijerinckii* are presented in Supplementary Figure 5 and Supplementary Figure 6.

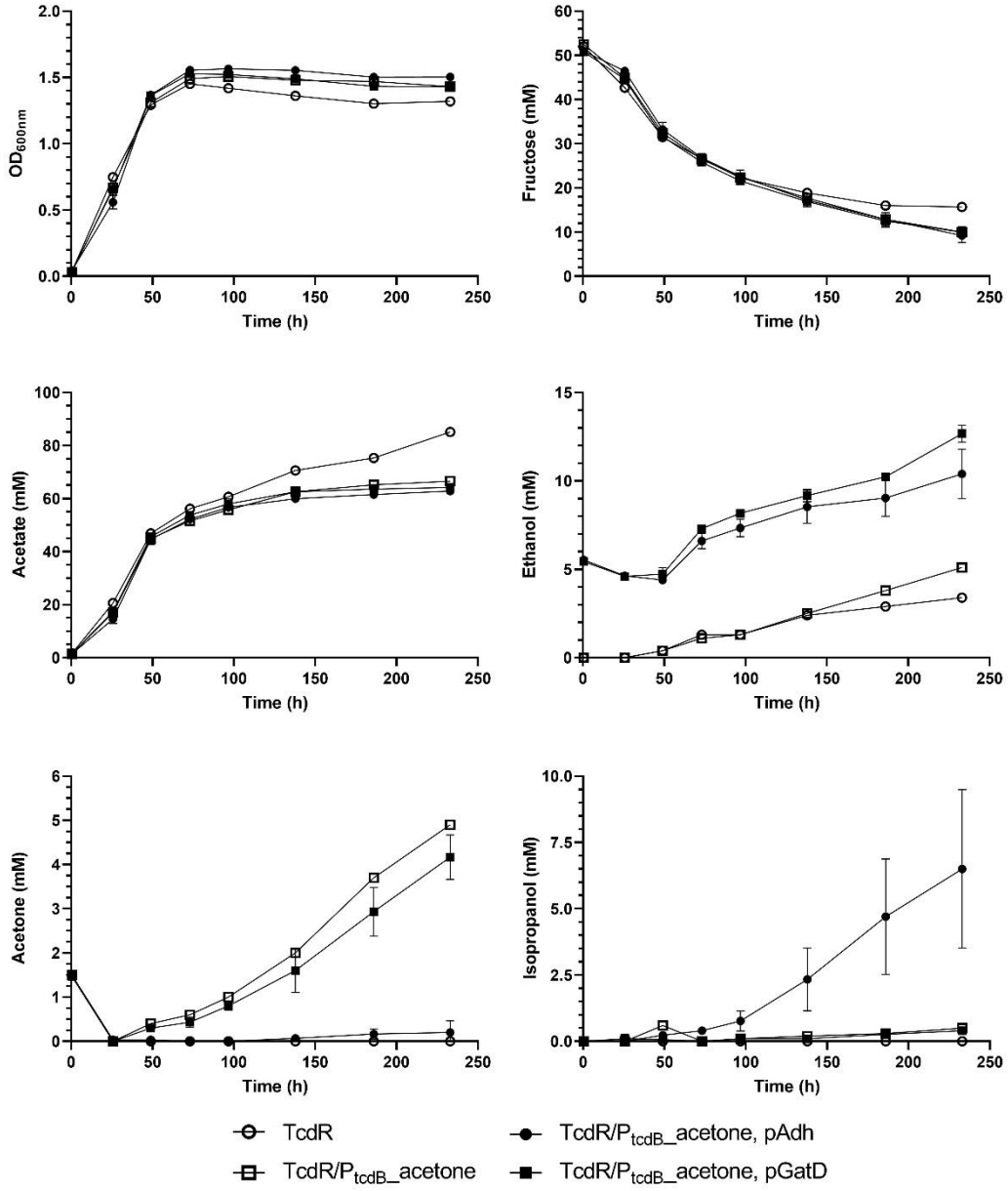


Figure 5-6: Growth and product profiles of *A. woodii* strains with the acetone operon and overexpressed SADH from *C. beijerinckii* or *gatD* from *A. woodii*. OD_{600nm}, acetate, ethanol, acetone, and isopropanol concentrations of *A. woodii* strains during heterotrophic batch fermentations in serum bottles over 233 h. *A. woodii tcdR* (○) and *tcdR*/P_{tcdB}_acetone (□) were studied as controls. Strains with *tcdR*/P_{tcdB}_acetone and either plasmid pAdh (●) or plasmid pGatD (■) were studied in triplicate.

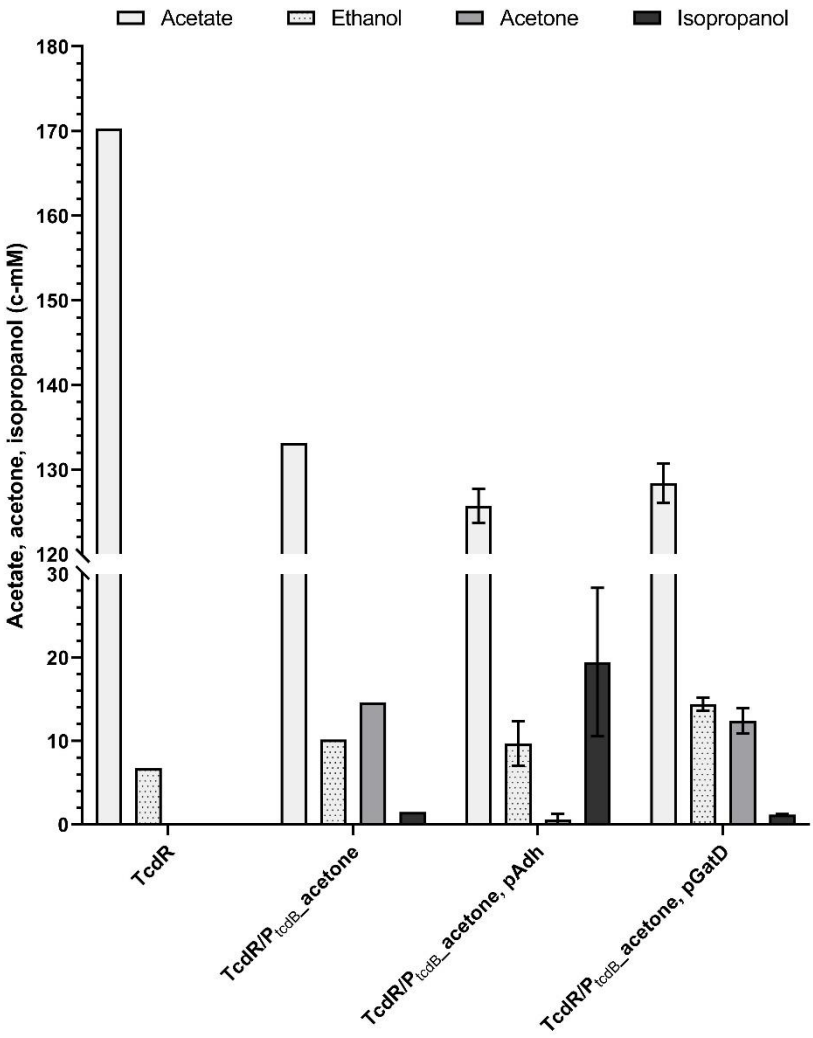


Figure 5-7: Product profiles of *A. woodii* with the acetone operon and overexpressed SADH from *C. beijerinckii* or *gatD* from *A. woodii*. Acetate (white), ethanol (white dotted), acetone (light grey), and isopropanol (dark grey) concentrations obtained with *A. woodii* strains after 233 h of heterotrophic batch fermentations, expressed relative to the carbon content of each product (c-mM). Ethanol concentrations were calculated by subtracting the starting concentrations for each strain. *A. woodii tcdR* and *tcdR*/P_{tcdB}_acetone were studied as controls. Strains with *tcdR*/P_{tcdB}_acetone and either plasmid pAdh or plasmid pGatD were studied in triplicate.

As previously observed, *A. woodii tcdR* resulted in the lowest OD_{600nm}, lowest fructose consumption and highest acetate concentration. Demonstrating once more that in the absence of the acetone operon, acetate reached its maximal toxic level and no further fructose was consumed. As opposed to strain *tcdR*/P_{tcdB}_acetone in which acetate is produced and converted into acetone, resulting in higher fructose consumption and lower acetate concentrations.

The production profiles of acetate, acetone, and isopropanol were similar to the ones observed in section 5.2.1. Acetate was produced exponentially from the start of growth, while acetone and isopropanol increased linearly after a 25 h or 50 h lag phase, respectively.

Plasmid pAdh resulted in a higher isopropanol concentration and lower acetate concentration compared to the *tcdR*/P$_{tcdB}$_acetone background strain alone. The concentrations obtained were comparable to the ones obtained in section 5.2.1 but variations between replicates of the isopropanol concentration highlighted the need for replicates for accurate conclusions.

Ethanol concentrations confirmed ethanol production for *tcdR* (3.4 mM) and *tcdR*/P$_{tcdB}$_acetone (5.1 mM) strains, production levels were comparable to section 5.2.1. Variations between replicates was lower than previously observed and all four strains displayed the same production profile. Similar to isopropanol production, the concentration increase was linear and started after a 50 h lag phase. Plasmids pGatD and pAdh resulted in average ethanol production of 7.23 mM and 5.0 mM, respectively.

The addition of plasmid pGatD in the *tcdR*/P$_{tcdB}$_acetone background strain did not result in higher acetone and isopropanol levels. This result invalidated the hypothesis that *gatD* is responsible for acetone to isopropanol conversion in acetone-producing *A. woodii* strains.

5.2.2.2.2 Gene knock-out

To be able to determine the role of the *gatD* gene, a knock-out strain was created using the tool developed in Chapter 3. A 1,041 bp in-frame deletion was targeted, to leave the six first and last nucleotides of the gene, as previously done for the *hsdR1* deletion mutant in section 3.2.3.6. The HAs designed were 630 bp and 1,000 bp, and the spacer

was 36 bp in length and 156 bp from the RHA. *A. woodii tcdR*/P_{tcdB}_acetone was transformed with the editing plasmid pMTL8415_GatD_KO (Figure 5.8A) and colonies displaying the gene KO were successfully generated (Figure 5.8B).

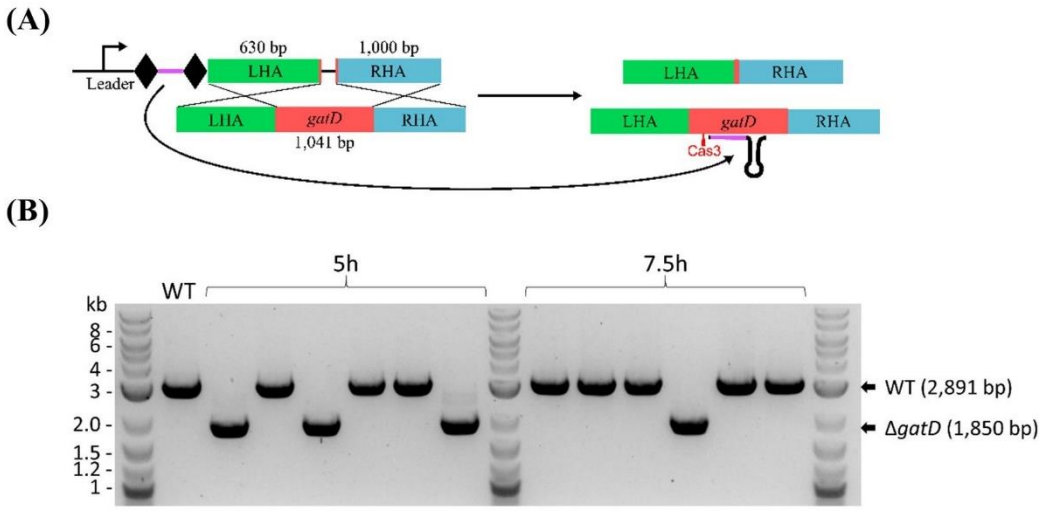


Figure 5-8: Endogenous CRISPR based deletion of the *A. woodii gatD* gene. (A) Representation of the application-specific module of the editing plasmid for *gatD* (red) KO in *A. woodii*, pMTL8415_GatD_KO. The 36 bp spacer (purple) is flanked by the native DR (diamond) and expressed under the control of the native leader sequence. The editing template with a 630 bp LHA (green) and a 1,000 bp RHA (blue) allows for homologous recombination and deletion of a 1,041 bp region. **(B)** Colony PCR results with primers FW_gatD_seq and RV_gatD_seq of the *gatD* locus of colonies transformed with the editing vector for *gatD* KO. WT amplicon is expected at 2,891 bp and KO amplicon is expected at 1,850 bp.

Due to lack of time, and considering the results obtained in the previous section, the impact on conversion of acetone to isopropanol was not evaluated in the knock-out mutant. Measurement of the acetone to isopropanol ratio in the *A. woodii tcdR*/P_{tcdB}_acetone and *A. woodii tcdR*/P_{tcdB}_acetone Δ*gatD* strains would allow verification of the absence of implication of *gatD* in the acetone to isopropanol conversion as hypothesized in the previous section. Measurements of ethanol could also provide insights in the implication of *gatD* in ethanol production in *A. woodii*.

5.2.3 Transposon library for discovery of novel genome loci for gene expression

5.2.3.1 Fluorescent reporter genes

To enable the random knock-in and identification of loci with improved fluorescence in *A. woodii*, a functional fluorescent reporter gene needs to be identified. The gene must be easy to knock-in and allow for identification of fluorescent and non-fluorescent populations by flow cytometry. Additionally, fluorescence-activated cell sorting and re-culturing of the cells must be feasible.

5.2.3.1.1 Reporter genes with LOV domains

Genes Pp1, Bs2, and creiLOV were selected to be tested in *A. woodii*. The first attempts at assembling vectors containing the genes downstream of the P_{thl} promoter were unsuccessful, resulting in deletions and mutations of the genes. To resolve potential toxicity issues in *E. coli*, the high copy number ColE1 replicon was changed for the low copy number replicon p15a. Cloning resulted once more in multiple mutations in the three genes, abolishing fluorescence (Supplementary Figure 2).

LOV-domain genes require FMN for the chromophore to fold and fluoresce, it was hypothesized that riboflavin was the bottleneck causing metabolic burden to the cells. After the discovery of the R-M system recognition site (characterized in section 3.2.1.1) in the p15a sequence, cloning of the genes was performed anew with the high copy number Gram-negative replicon ColE1 and with riboflavin supplementation in the LB medium. This experiment allowed for the isolation of colonies without mutations in *E. coli*, confirming the riboflavin requirement. Accordingly, *A. woodii* transformation with the plasmids was performed in the presence of additional vitamins (including riboflavin) and colonies containing the plasmids were successfully obtained.

Emission and excitation scans were performed on strains containing Pp1 to confirm the fluorescence characteristics of the protein (Supplementary Figure 7A and B). The presence of Pp1 shifted the emission maximum and resulted in the generation of an additional excitation peak, compared to the medium only and empty vector strain controls. These results showed that the medium contained a high fluorescence background. Subtraction of the empty vector values from the samples confirmed the maximum emission and excitation values reported in the literature (Supplementary Figure 7C and D). It was observed that the control with the medium had emission and excitation values higher than the control with the empty vector. Riboflavin has been found to have a maximum excitation of 442 nm and maximum emission of 533 nm (Ahmad et al., 2017; Zhang et al., 2012), and was therefore hypothesized to be the cause of the high fluorescence background. A standard curve confirmed that the *A. woodii* vitamin solution fluorescence overlaps with the fluorescence of LOV-domain genes (Supplementary Figure 8).

Taken together, these results demonstrate the metabolic burden of the expression of the LOV genes on the cells and the risk of mutations in the genes in the absence of available riboflavin. For the proposed random knock-in experiment, the risk of variation of riboflavin availability between each cell would result in mutation of the gene. This would compromise the experiment as it would result in variation of fluorescence due to mutation of the gene and not genomic locus insertion.

5.2.3.1.2 FAST

The FAST reporter gene was already successfully cloned between the thiolase promoter (P_{thl}) and the *pyrE-hydA* terminator to function as a positive control for transcriptional control for the leader sequence in section 3.2.3.3. This demonstrated a lower burden on the cells in normal growing conditions compared to the LOV-domain proteins. FAST

was therefore assessed in more details to ensure its suitability for the fluorescent reporter gene knock-in library.

i) Knock-in of the Pthl_FAST cargo at pheA locus

The KI of the P_{thl}_FAST cassette was performed with the endogenous CRISPR/Cas system developed in Chapter 3. As no 5′ CCA trinucleotide could be found in the target intergenic region downstream of *pheA*, an alternative 5′ CCG PAM was chosen for spacer design. Its selection was based on the interference assay (section 3.2.3.2) suggesting that the double cytosine is essential for efficient targeting. A 36 bp spacer was chosen in the intergenic region downstream *pheA* and homology arms of 1,496 bp and 1,465 bp were designed. The homology arms were designed for insertion of a cargo in place of the PAM. A small cargo (614 bp) containing the FAST gene flanked by the *thl* (thiolase) promoter and *pyrE-hydA* (orotate phosphoribosyltransferase and hydrogenase I) terminator from *C. acetobutylicum* was inserted between the homology arms. WT *A. woodii* was transformed with the editing vector, pMTL8415_leader_pheA CCG_KI (Figure 5.9A) following the protocol named T3 (section 3.2.3.6).

Colony PCRs confirmed that successful knock-ins were obtained with plating at both 5 h and 7.5 h with an overall efficiency of around 57% (Figure 5.9B). Following plasmid loss, phenotypic analysis of the knock-in strains was undertaken. Plating of two of the clones onto minimal media with and without phenylalanine supplementation confirmed that the insertion had not disrupted *pheA* (Figure 5.9C).

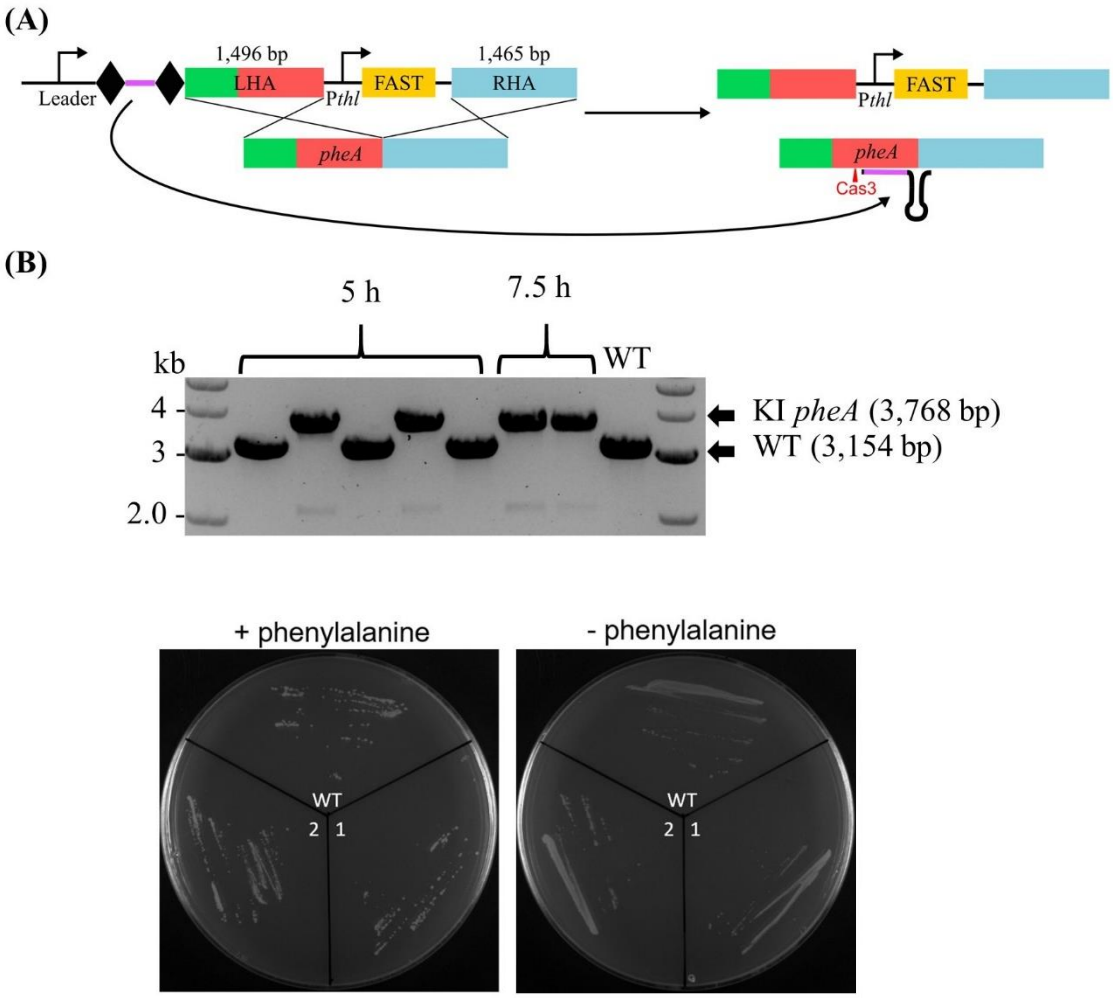


Figure 5-9: Endogenous CRISPR based insertion of P_{thl}_FAST at the *pheA* locus in the *A. woodii* genome. (A) Representation of the application-specific module of the editing plasmid for P_{thl}_FAST insertion at the *pheA* (red) locus in *A. woodii*, pMTL8415_leader_pheA_CCG_KI. The 36 bp spacer (purple) is flanked by the native DR (diamond) and expressed under the control of the native leader sequence. The editing template with a 1,496 bp LHA (containing *pheA*) and a 1,465 bp RHA (blue) allows for homologous recombination and insertion of the cargo. **(B)** Colony PCR results with primers FW_pheA_KI_seq and RV_pheA_KI_seq of the *pheA* locus of colonies transformed with the editing vector. WT amplicon is expected at 3,154 bp and KI amplicon is expected at 3,768 bp. **(C)** Growth on minimal media with and without phenylalanine supplementation (20 µg/mL) of two *A. woodii pheA* KI colonies (marked 1 and 2) in comparison to WT.

ii) Flow cytometry analyses

Assessment of the suitability of FAST for flow cytometry measurement was performed by analysing WT *A. woodii*, WT *A. woodii* with P_{thl}_FAST in a pMTL84151 plasmid, and *A. woodii* with FAST at the *pheA* locus (Figure 5.10).

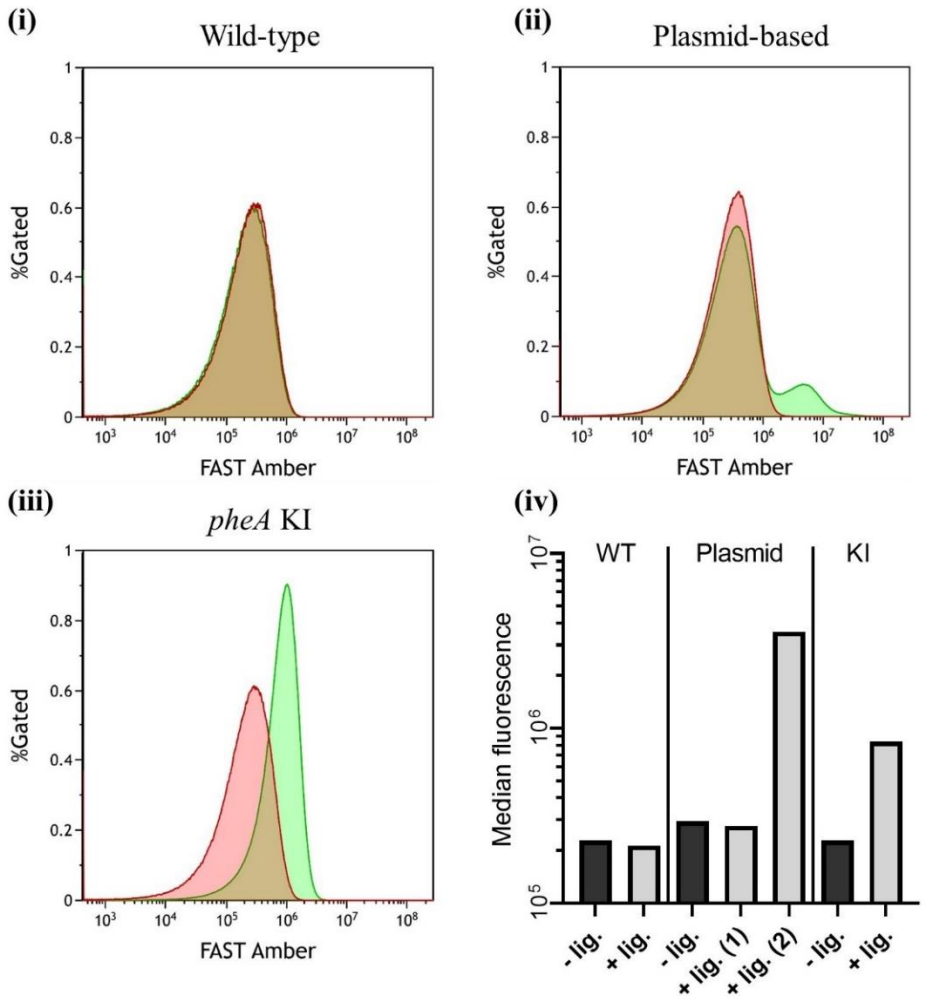


Figure 5-10: Flow cytometry analyses of *A. woodii* using the FAST reporter gene (P_{thl}_FAST) expressed on a plasmid or in the genome. (A) Flow cytometry results for **(i)** *A. woodii* WT, (**ii**) *A. woodii* WT with pMTL8415_P_{thl}_FAST, and **(iii)** *A. woodii* P_{thl}_FAST *pheA* KI. Cell populations with PBS only (**red**) or with the ^{TF}Amber ligand (**green**), the percentage of cells gated is represented as a function of the FAST Amber fluorescence intensity. The number of events represented is > 440,000 for WT and > 1,200,000 for all other strains; exact numbers can be found in Materials and Methods (section 2.9.3). **(iv)** Median fluorescence of the populations with PBS only (**dark grey**) or with the ^{TF}Amber ligand (**light grey**). For the plasmid-based expression strain, the median fluorescence was calculated separately for the two peaks observed and annotated (1) and (2).

As expected, WT *A. woodii* displayed the same level of fluorescence with and without ligand and the plasmid-based FAST resulted in detectable fluorescence only in the presence of ligand. The plasmid-based population fluorescence is heterogeneous with two clear cell populations: one population with a similar level of fluorescence as without

a ligand and a smaller population displaying a higher level of fluorescence. The KI strain, on the contrary, displayed a homogeneous population with a level of fluorescence higher than without the ligand. Comparison of the median fluorescence for each condition shows that the fluorescence of the KI population (KI + ligand on Figure 5.10B(iv)) is lower than the fluorescence of the plasmid-based fluorescent population (Plasmid + ligand (2) on Figure 5.10B(iv)). It is also interesting to note that the auto-fluorescence of the plasmid-based population is also slightly higher than both the WT and KI strains.

These results confirm the heterogeneity of plasmid-based expression of exogenous genes. They also confirm the suitability of FAST for flow cytometry and the functionality of the inserted FAST gene in the *pheA* locus of *A. woodii*, demonstrating the knock-in capabilities of *A. woodii* Type I-B CRISPR/Cas system.

iii) Cell survival and contamination

Cell survival was assessed by sorting and re-culturing different number of single cells from the flow cytometry experiment described in the previous section. Concomitantly, re-culturing was used to confirm that the heterogeneity of the plasmid-based expression was not caused by plasmid loss (Streett et al., 2019). 100 to 500,000 cells of *A. woodii* WT (with increments of 1,000 and 10,000 cells) and 1,000 cells to 500,000 cells of the plasmid and KI strains (with increments of 10,000 and 100,000 cells) were sorted. For the plasmid strain, both fluorescent and non-fluorescent cells were sorted, and re-culture was performed in the presence of antibiotic selection.

Unfortunately, the volumes of culture after cell sorting were too low to allow for OD_{600nm} measurements. Nonetheless, all the cultures grew and as expected the lag phase decreased as the number of cells increased, showing cell survival and plasmid retention

for both fluorescent and non-fluorescent population in the case of the plasmid-based strain.

The cell sorting experiment contains a few steps performed in non-sterile environments, including addition of the ligand and re-injection of the sorted cells into sterile anaerobic tubes. Microscope observation, *pyrE* locus amplification, and sequencing of the 16S locus of all the liquid cultures confirmed the identity of the cultures as *A. woodii* and apparent absence of contamination in the sorted cells.

FAST is the ideal candidate for use as a reporter for construction of a library with random knock-in and cell sorting.

5.2.3.2 Library construction

5.2.3.2.1 Terminator characterization

The transposon library aims to evaluate the impact of insertion of the FAST reporter gene in different loci of *A. woodii* genome. To ensure that expression of the FAST gene after insertion in the genome does not cause read-through into the downstream genes, a strong terminator needs to be chosen. To quantify terminator strength, an RT-qPCR experiment was designed to evaluate the ratio between mRNA fragments upstream and downstream of the terminator.

For measurement, *catP* was placed between the P_{thl} promoter and each terminator studied and FAST was placed downstream of each terminator. For each terminator, a negative control without FAST was constructed (Figure 5.11A). The terminator of the *C. acetobutylicum pyrE-hydA* locus (orotate phosphoribosyltransferase and hydrogenase I) and a terminator from *C. tetani* were evaluated. The former was employed when inserting FAST in the genome at the *pheA* locus (section 5.2.3.1.2) and the latter was shown to be a strong promoter in previous experiments (Dr. Jonathan

Baker, personal communication). All the primers employed are listed in Supplementary Table 9, primers to amplify *gyrA*, *orfB*, and FAST have already been tested in section 3.2.3.3. The terminator activity was determined by the ratio of mRNA levels of FAST in relation to *catP* (Figure 5.11B).

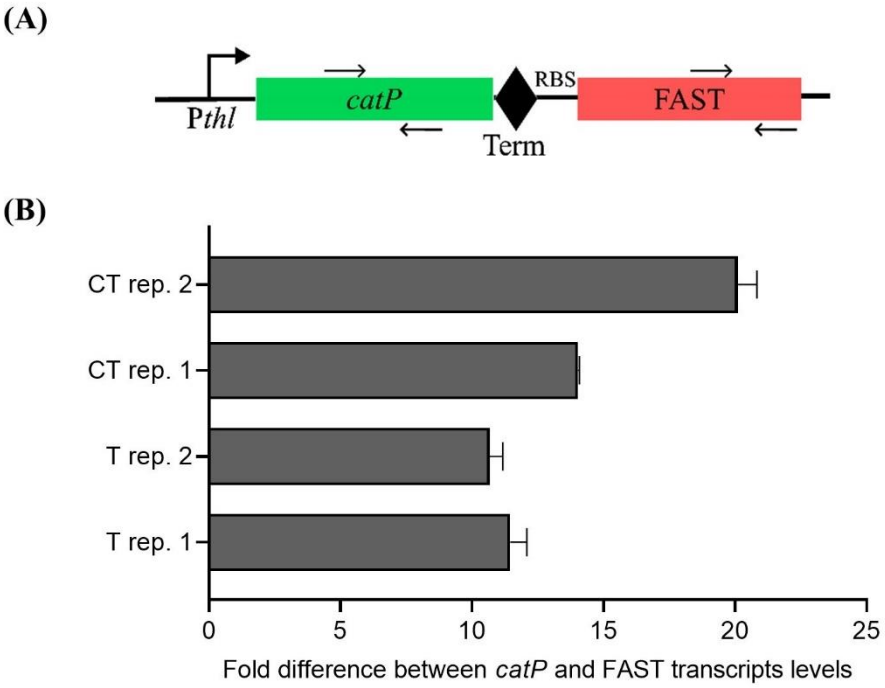


Figure 5-11: Terminator characterization for use in the FAST transposon library. **(A)** Schematic representation of the modules built for characterization of the terminators, with Term being either the *pyrE-hydA* terminator from *C. acetobutylicum* or the terminator from *C. tetani*. **(B)** Fold difference between the transcript levels of *catP* and FAST after normalization to the transcript levels of *orfB* to take into account plasmid copy number. Fold difference was calculated by dividing the transcript level of *catP* by the transcript level of FAST. T = *pyrE-hydA* terminator from *C. acetobutylicum*; CT = terminator from *C. tetani*. Terminator were assessed in technical triplicates and biological duplicates.

As expected, no FAST transcript was detected in the absence of the FAST gene (data not shown) demonstrating the specificity of the primer employed. The two biological replicates containing the *C. acetobutylicum* terminator were similar whereas the two replicates containing the *C. tetani* terminator led to variable repression of transcription. On average, the terminator from *C. tetani* led to stronger repression of transcription.

In the pooled library, all the mutants will be assessed at the same time for increased fluorescence. Thus, it is crucial that all the mutants contain the exact same module inserted in the genome and that the module behaves in a similar manner in all the mutants. Accordingly, the FAST transposon library was built with the *pyrE-hydA* terminator from *C. acetobutylicum*.

5.2.3.2.2 Plasmid design and assembly

Based on the library analysed in Chapter 4 and in accordance with all the preliminary experiments performed, the plasmid for library generation was built to contain a cassette for FAST expression alongside the thiamphenicol resistance cassette between the ITR (Figure 5.12). The FAST cassette was the same as the one used on the plasmid pMTL8415_P_{thl}_FAST and for KI at the *pheA* locus, with the P_{thl} promoter and the *pyrE-hydA* terminator. Successful assembly was confirmed by Sanger sequencing.

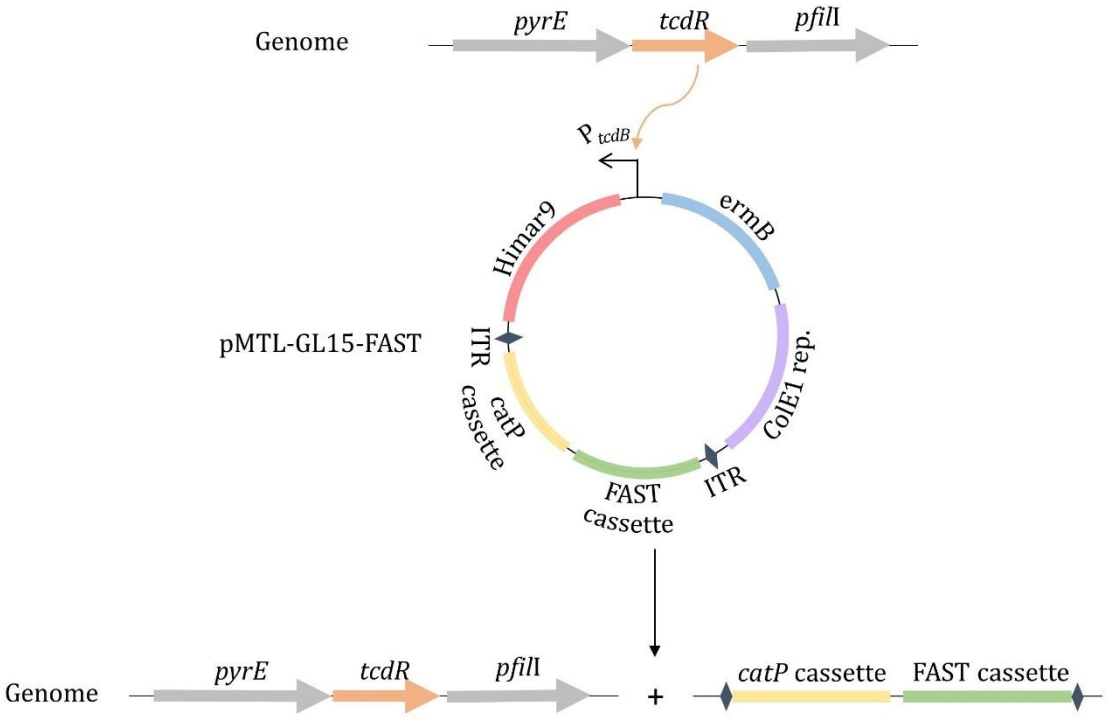


Figure 5-12: Transposon library construction for random insertion of both an antibiotic resistance gene and the FAST fluorescent reporter using plasmid pMTL-GL15-FAST. pMTL-GL15-FAST was built from plasmid pMTL-GL15 employed in

Chapter 4, it contains the Himar1 transposase under the control of the P_{tcdB} promoter, the Gram-negative replicon ColE1, the resistance marker gene *ermB*. The FAST gene is present between the ITRs alongside the *catP* cassette. Upon transformation of an *A. woodii* strain containing the *tcdR* gene in its genome, the transposase is expressed and the *catP*-FAST cassette randomly inserted in the genome. pMTL-GL15-FAST does not contain a Gram-positive replicon, it is not maintained in the cells after transformation.

5.2.3.2.3 Plasmid retention

The potential retention of the replicon-deficient plasmid employed for construction of the library in the cells was assessed based on the antibiotic resistance profile of the cells. Plasmid retention results in thiamphenicol and clarithromycin resistance whereas successful transposition events result in thiamphenicol resistance only. For this purpose, both *A. woodii* WT and *tcdR* were transformed with the pMTL-GL15-FAST plasmid. For both the WT and *tcdR* strain, competent cells were prepared according to the high-density protocol. The recoveries were plated on both clarithromycin and thiamphenicol selection plates and colony numbers recorded (Figure 5.13).

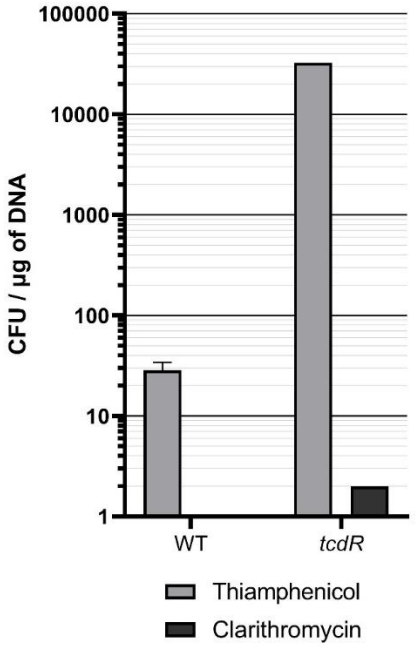


Figure 5-13: Resistance profile of colonies obtained after transformation of *A. woodii* WT and *tcdR* with plasmid pMTL-GL15-FAST for FAST transposon library creation. The number of colonies was normalised to the amount of DNA used for transformation and expressed as CFU / µg of DNA. Number of colonies obtained on thiamphenicol (light grey) and clarithromycin (dark grey) are shown.

A comparison of CFU on thiamphenicol or clarithromycin indicated plasmid retention of 0.0062%. Surprisingly, *A woodii* WT transformation with the pMTL-GL15-FAST plasmid generated thiamphenicol resistant colonies. An equal number of thiamphenicol and clarithromycin resistant colonies would have indicated plasmid retention in the cells. Generation of thiamphenicol only resistant cells indicated transposition events occurring in the WT. These transposition events occur in the absence of *tcdR* and can be hypothesized to be caused by leakage of the P_{tcdB} promoter expressing the transposase. Transformation of *A. woodii tcdR* resulted in a more than 1,000-fold increase in thiamphenicol resistant colonies compared to *A. woodii* WT, this value is higher than the one observed with the library employed in Chapter 4 (265-fold). This result indicated a relatively low leakage level of the promoter. Considering the promoter is activated upon transformation of the cells with the plasmid by the presence of *tcdR* in the genome, P_{tcdB} leakage is not a concern for construction of the library.

This experiment also allowed for estimation of the number of transformation necessary for generation of an adequate single mutant library pool. $1.3x10^5$ thiamphenicol resistant colonies were generated with the protocol, to obtain a mutant library of around 1 million single mutants, nine identical transformations are required. Accordingly, the library was built using the same batch of *A. woodii tcdR* competent cells as tested here in nine replicates.

5.2.3.3 Library analysis

5.2.3.3.1 Fluorescence

The most interesting analysis performed on the pooled transposon library is the analysis of the fluorescence of the population at the cell level. For this purpose, WT *A. woodii* and *A. woodii* with P_{thl}_FAST on a plasmid (pMTL8415_P_{thl}_FAST) or integrated in the

genome (both generated in Chapter 3) were employed as controls alongside the pooled library during flow cytometry measurements (Figure 5.14).

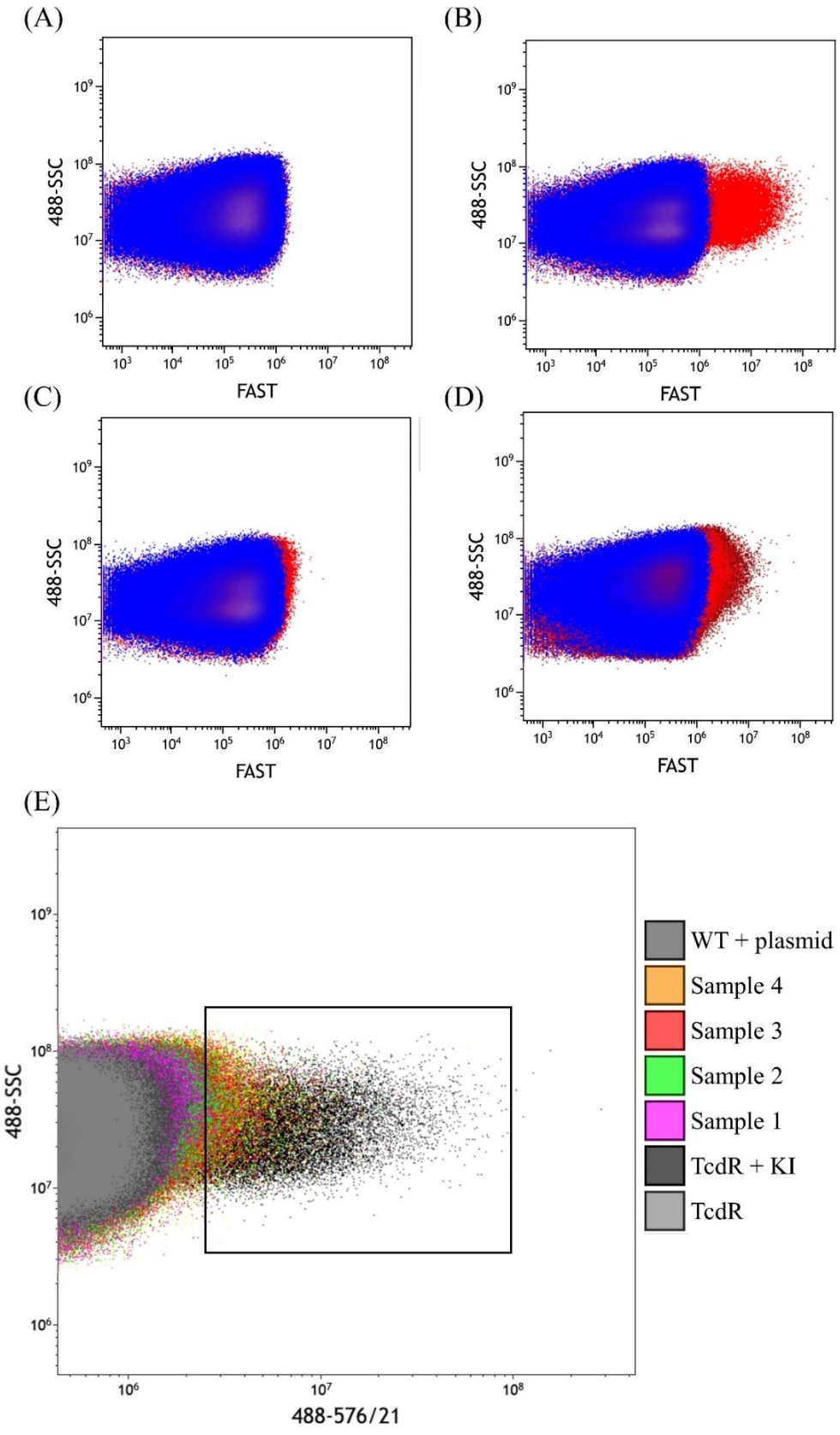


Figure 5-14: Flow cytometry analysis of the FAST transposon library in *A. woodii*. Histogram of the cells obtained with the **(A)** *tcdR* background strain, **(B)** the plasmid strain (pMTL8415_P$_{thl}$_FAST), **(C)** the FAST KI strain, and **(D)** the pooled transposon library. In blue are represented the cells without ligand and in red with the ^{TF}Amber ligand. For the pooled library, four samples were analysed with the ligand, they are represented in varying shades of red. **(E)** Superimposition of the plots of the fluorescence analyses in presence of the ligand for all the strains, with the axes adapted for a better view of the fluorescent populations. In grey the *tcdR* background strain and the plasmid strain, in the foreground and background, respectively. In black the FAST KI strain and in orange, red, green, and pink the four samples of the pooled library. The rectangle delineates the outliers discovered in the samples of the pooled library with a high fluorescence intensity and separate from the outlier cells obtained with the FAST KI strain. Number of single cells analysed for each sample can be found in Materials and Methods (section 2.9.3).

A first analysis was performed to assess the fluorescence of the pooled library and put in place the gates for sorting of the cells. Surprisingly, the plasmid-based expression resulted in a smaller percentage of fluorescent cells in comparison to the experiment presented in section 5.2.3.1.2 (Figure 5.14B) but with the same fluorescence intensity. Interestingly, the library population samples show a shift in fluorescence similar to, or higher than, the shift observed with the FAST KI at the *pheA* locus. They also contain some cells with high fluorescence (Figure 5.14E), including as high or higher than the average fluorescence obtained with the plasmid strain.

It was decided to set sorting gates in both the fluorescent and non-fluorescent populations, four sorting gates for cells in the main non-fluorescent population and three sorting gates for the cells with higher fluorescence. Each set of gates was designed to sort cell population of increasing fluorescence and to not overlap with its two neighboring gates. A hundred cells were sorted for each gate and the gates for high fluorescence were sorted in duplicate. The sorted cells were re-introduced in the anaerobic cabinet and growth media was added to be able to perform further phenotypic assays as described thereafter.

5.2.3.3.2 Verification of single copy of the FAST gene in the sorted cells

In the library, assessment of FAST expression and improved fluorescence relies on the presence of only one copy of the FAST gene in each mutant. Two phenomena can result in multiple copies of the FAST gene, plasmid retention and multiple transposon insertion into the genome.

i) Plasmid retention

The high copy number of plasmids could result in high copy number of the FAST gene in the sorted cells. Consequently, it was essential to check for plasmid presence in the sorted cells.

After cell sorting, the cells were re-cultured in medium containing thiamphenicol for selection of the transposon presence, after growth of the cultures, they were sub-cultured maintaining the selection. Both the cultures and the sub-cultures were plated onto medium containing thiamphenicol to obtain single colonies. Those single colonies were streaked onto thiamphenicol and clarithromycin. Growth on both antibiotics would indicate plasmid presence, while growth on thiamphenicol only would indicate transposon insertion in the genome. For each gate described in the previous section and for both pre-cultures and sub-cultures, 30 colonies were analysed.

Surprisingly, none of the colonies tested were found to grow on both antibiotics, demonstrating the absence of the plasmid in the cells. The colonies from those tests were grown in liquid cultures and were used for single mutant phenotypic work.

ii) Single insertions

Another potential risk is the insertion of the transposon into multiple loci of the genome resulting in multiple copies of the FAST gene. Inverse PCR allows to assess the number of insertions of the transposon for each mutant. The size of the amplicon also informs

on the insertion sites and can help verify that the insertion sites are different between a range of mutants. Sequencing of the amplicon allows for identification of the transposon insertion locus.

Briefly, fragmentation of the DNA by restriction digestion followed by ligation of the fragments generates circular DNA. For each mutant, at least one of the circular fragments generated contains the transposon insertion, design of primers facing outwards allows for amplification for the circular DNA generated. Single insertions result in only one amplification whereas multiple insertions result in multiple amplifications.

The enzyme used for digestion should not digest between the primers employed inside the transposon cargo. The enzyme HindIII was used to demonstrate single insertion in the library analyzed in Chapter 4 (Dr. Jonathan Baker, personal communications).

The first attempt at inverse PCR using the protocol described in Materials and Methods (section 2.10) was unsuccessful albeit successful amplification of the plasmid template demonstrating that the primers are functional.

Troubleshooting of the protocol was undertaken with different combination of changes as follows. For both digestion and ligation, different DNA amounts were tested. The digestion was tested with incubation between 1 h to overnight and the ligation was tested with incubation between 1 h or 2 h at room temperature and overnight at 16ºC. PCR was performed with different conditions, different template amounts and different primer pairs catP_INV_F1_bis / catP_INV_R1_bis with annealing at 65ºC or catP_INV_F1 / catP_INV_R1 at 72ºC.

All the parameters tested did not affect the outcome of the inverse PCR for the library samples. More fundamental tests were subsequently performed to check for the integrity of the samples and the enzymes involved.

Sample integrity was tested on a small number of samples by amplification of the *A. woodii* genomic *pheA* locus and verification of the transposon presence. The former was performed with primer pair FW_pheA_cargo_seq / FW_pheAHA1.5_out and the latter with FW_catp_cassette_corrected / catP_INV_R1 (Figure 5.15).

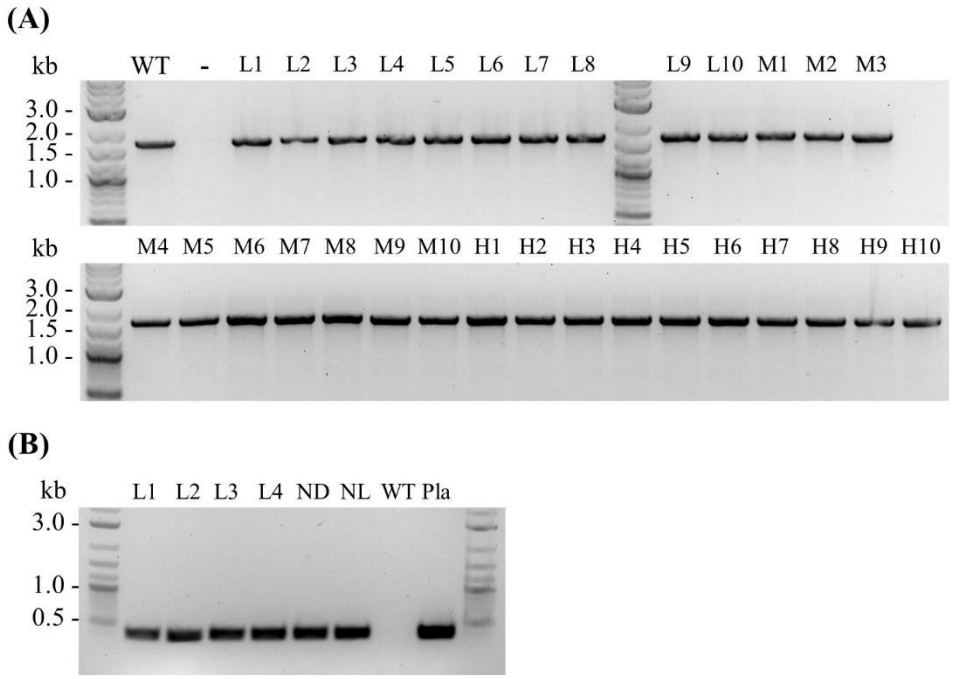


Figure 5-15: Analysis of genomic DNA samples from single colonies obtained by FACS. (A) *A. woodii pheA* locus, primers FW_pheA_cargo_seq / FW_pheAHA1.5_out, WT amplicon is expected at 1,671 bp. **(B)** Verification of transposon sequence presence, primers FW_catp_cassette_corrected / catP_INV_R1 were employed and the amplicon is expected at 400 bp. For both **(A)** and **(B)**, L, M, H represent the different gated population Low, Medium, and High respectively and the number represents the different single colonies analysed; WT: wild-type; -: no template; ND: non-digested; NL: non-ligated, Pla: pMTL-GL15-FAST

The PCR reactions demonstrated that the genomic DNA extracted is *A. woodii* (Figure 5.15A) with the transposon sequence inserted (Figure 5.15B). This did not show the position of the transposon but confirmed its presence in the samples.

To test the enzymes involved, the inverse PCR protocol was performed on the plasmid used for library construction (pMTL-GL15-FAST) as well as on a sample containing a mix of mutants from the library analyzed in Chapter 4 (G9). To test the effective DNA cutting by HindIII, AscI was used as a control, unlike HindIII, restriction with AscI results in two fragments when cutting pMTL-GL15-FAST. To test the impact on restriction enzyme inactivation and its presence with associated buffers on the subsequent ligation, a step of DNA purification was added between the digestion and the ligation.

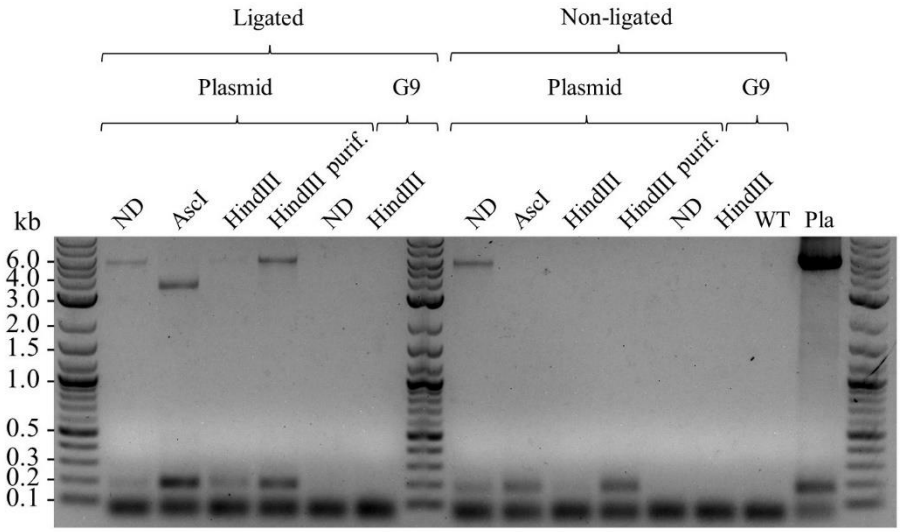


Figure 5-16: Inverse PCR protocol performed on plasmid pMTL-GL15-FAST and sample G9 of the library of Chapter 4 (mix of mutants). Primer pair catP_INV_F1_bis / catP_INV_R1_bis, for HindIII and AscI digestions the fragments are expected at 5,622 bp and 3,486 bp, respectively. Half of the HindIII digestion was purified before the ligation (HindIII purif.). As a control, PCRs were also performed on samples that were not ligated. ND: non-digested.

The inverse PCR protocol was not successful when applied to the G9 sample. It contains a high number of mutants and a DNA smear was expected as a result of the inverse PCR protocol.

Digestion of the pMTL-GL15-FAST plasmid with both AscI and HindIII were successful (Figure 5.16). Subsequent ligation and PCR resulted in successful amplification in all the conditions tested. Digestion with AscI resulted in higher PCR efficiency compared to HindIII, purification of the HindIII digestion before ligation increased PCR efficiency to a similar level as obtained with AscI digestion. (Figure 5.16). Both AscI and HindIII were inactivated at 80ºC as per provider instructions. It can be hypothesized that inactivation of HindIII is not complete, and some enzyme remained active in the ligation reaction.

Accordingly, the same procedure was applied to the samples from the library. However, all the PCRs were, once more, unsuccessful. Polymerases Taq, Q5, and KOD were then tested on the resulting ligation, but amplification was unsuccessful.

While inverse PCRs have been unsuccessful so far, successive experiments have shown that the samples are *A. woodii* genomic DNA and the transposon sequence is present; the enzymes used are functional and the protocol is successful when applied to plasmid pMTL-GL15-FAST. This confirmed the resistance profile observed in section 5.2.3.3.2 when assessing plasmid-retention.

5.2.3.4 Control strains

After identification of interesting genomic loci with the FAST reporter gene expression, the loci will be further characterized for their impact on production yields with other exogenous genes. The impact of the genomic position could be tested on both acetone and isopropanol production.

Impact on acetone yields would be evaluated by knocking-in the whole acetone operon at the different genomic loci, in a strain that already contains *tcdR* at the *pyrE* locus. The knock-in of the acetone operon in new genomic loci might prove challenging due to its length (3,677 bp).

Study of isopropanol yields obtained with the *tcdR*/P_{tcdB}_acetone background strain and the SADH encoding gene inserted at different genomic loci offers an interesting alternative as it requires the insertion of only one gene for each locus characterisation.

Control strains with the acetone operon or *adh* gene inserted in the genome at traditional genome loci are necessary for comparison with the new loci to be tested.

5.2.3.4.1 Isopropanol production

The *purD* locus was selected for KI of the SADH in a strain already containing tcdR at *pyrE* and the acetone operon at *pheA*, an editing vector was designed according to Chapter 3. *A. woodii tcdR*/P_{tcdB}_acetone was transformed with the editing vector pMTL84151_pheA_KI_PtcdB_CbeSADH in one replicate but did not result in successful knock-in.

5.2.3.4.2 Acetone production

An editing vector (pMTL84151_pheA_KI_PtcdB_acetone) was designed to insert the acetone operon at the *pheA* locus in the same manner as described in section 5.2.3.1.2for the FAST cargo. *A. woodii tcdR* was transformed with pMTL84151_pheA_KI_PtcdB_acetone and knock-in colonies were successfully generated (Figure 5.17).

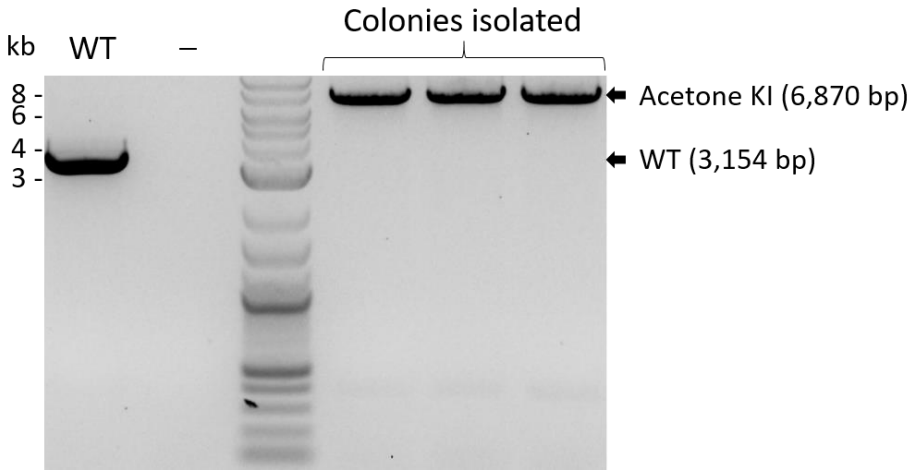


Figure 5-17: Knock-in of the acetone operon at the pheA locus of *A. woodii* using its endogenous CRISPR/Cas system. Colony PCR results obtained with primers FW_pheA_KI_seq and RV_pheA_KI_seq, WT amplicon is expected at 3,154 bp and KI at 6,870 bp.

Interestingly, colony PCRs were first performed on colonies directly from the transformation plates, and not after liquid cultures as was performed in all other experiments. Some colonies analysed presented a band corresponding to the WT locus and a band corresponding to the successful KI. The results presented in Figure 5.17 correspond to these colonies after liquid cultures with selection. This demonstrated that some cells with the WT locus might still be alive inside the colonies obtained on the selective plates; maintenance of the plasmid expressing the spacer targeting the WT locus with antibiotic selection allows for elimination of these WT cells.

The plasmid was lost successfully, and the strain is ready to function as a control for acetone production.

5.3 Discussion

5.3.1 Use of the endogenous CRISPR/Cas system for KO and KI

In this chapter, the tool developed in Chapter 3 was successfully used to generate various engineered strains. It allowed for KO of *gatD* and KI of both the P_{thl}_FAST cassette and the acetone operon at *pheA* locus. The successful KI at the *pheA* locus using a CCG PAM validated the hypothesis formulated in section 3.2.3.2 that only the double cytosine is essential for interference in *A. woodii*. This first KI confirmed the knock-in capabilities of *A. woodii* Type I-B CRISPR/Cas system and its flexibility. The subsequent KI of the acetone operon represented a 6-times increase in cargo size, demonstrating the robustness of the method developed. During the KI of the acetone operon, colony PCRs of the colonies directly from the plates and after liquid cultures, confirmed the selective pressure of Cas3 against the WT locus.

Disappointingly, the system failed to generate the knock-in of the SADH from *C. beijerinckii* at the *purD* locus. However, only one replicate of the transformation was performed, further experiments are necessary to confirm this result. If necessary, alternative spacer and homology arms could be designed.

5.3.2 Heterotrophic batch fermentation for acetone production

Unsurprisingly, heterotrophic batch fermentation of *A. woodii* resulted mainly in acetate formation in the absence of exogenous genes. However, a low amount of ethanol was produced. In the presence of high substrate concentrations (60 mM fructose), phosphate limitation (0.2 – 8.4 mM, the media utilized contained 5 mM) was shown to result in ethanol production (Buschhorn, Dürre, et al., 1989). In the same conditions as presented here (60 mM fructose, Na^+-rich, and 120 mM bicarbonate) Moon and Müller observed 6 ± 1 mM ethanol produced after 210 h (Moon & Müller, 2021). These concentrations

are comparable to the concentrations obtained in both section 5.2.1 and 5.2.2 with the same delayed onset of production. They demonstrated that ethanol formation starts at low pH in a similar manner as the shift from acidogenesis to solventogenesis in *C. acetobutylicum* (Monot et al., 1982; Moon & Müller, 2021). While elucidation of the responsible alcohol dehydrogenase was not successful, ethanol production was not observed during growth on H_2/CO_2 (Moon & Müller, 2021), limiting impacts on potential commercial strains development.

Addition of the acetone operon (Hoffmeister et al., 2016) controlled by the P_{tcdB} promoter in the *tcdR* background strain successfully resulted in acetone production. As previously observed, the yields obtained were lower when the operon was inserted in the genome compared to plasmid-based expression (Baker, 2020). During the development of this study, improvements to acetone production were attempted from the operon described in 2016 (Hoffmeister et al., 2016), two strategies were adopted (Arslan et al., 2022). First, the ribosome binding sites (RBS) of the acetoacetate decarboxylase (*adc*) and CoA transferase (*ctfA*) genes were changed to the native RBS of these genes. Second, the *ctfA/B* genes from *C. acetobutylicum* were replaced by genes from *C. aceticum* identified to have the same function but with a lower K_m. Both resulted in an increase of acetone production during autotrophic batch fermentation, the latter outperformed all the other strains in serum bottles and the former in stirred tank bioreactors.

At the time of this study, no acetone to isopropanol conversion had been reported (Hoffmeister et al., 2016) and isopropanol detected by HPLC had been omitted as artefacts (Baker, 2021). To confirm isopropanol conversion and validate HPLC isopropanol quantification, a strain containing the acetone production operon only was compared to a strain containing the acetone operon and the known SADH from

C. beijerinckii and *A. woodii* WT. While, addition of the SADH increased the isopropanol concentrations, both strains resulted in the production of isopropanol, not observed in the WT. This data was corroborated in 2022 by Arslan and co-workers (Arslan et al., 2022), in contradiction with previous literature (Hoffmeister et al., 2016; Kopke et al., 2014). They observed up to 14.5 mM isopropanol production in autotrophic batch fermentation in stirred tank bioreactors with the recombinant acetone-producing *A. woodii* containing the native RBS. In comparison, the highest isopropanol concentration achieved in the present study was 9.9 mM after 233 h of heterotrophic batch fermentation in serum bottles with the *tcdR*/P_{tcdb}_acetone strain containing *C. beijerinckii* SADH on an autonomous plasmid and under the control of the P_{tcdB} promoter.

The *gatD* gene was identified *in silico* as a candidate SADH in *A. woodii*, as also identified in the literature, because of its similarity to known SADH at the protein level (Arslan et al., 2022). Plasmid-based overexpression of *gatD* in *A. woodii* did not result in an increase of isopropanol compared to the strain containing genes for acetone production alone. Similarly, its plasmid-based expression in *C. acetobutylicum* did not result in isopropanol production (Arslan et al., 2022). Arslan and co-workers additionally tested gene Awo_c17360 with the same method and the same unsuccessful outcome.

In *C. autoethanogenum*, identification of the SADH responsible for acetone to isopropanol production enabled its KO and resulted in an increase in acetone yields (Liew et al., 2022). Accordingly, identification of the SADH in *A. woodii* remains an interesting avenue for future improvements of yields. Successive overexpression or KO of the alcohol dehydrogenases that remain uncharacterized and HPLC measurements could lead to its identification.

5.3.3 Fluorescence reporter genes

The proposed discovery of new genome loci at which inserted genes would exhibit improved expression levels relies on the selection of an appropriate reporter gene (Jeong et al., 2018; Saleski et al., 2021). The application of LOV-domain containing proteins for this purpose was compromised by their high toxicity to the cells, both *E. coli* and *A. woodii*, because of their requirement for riboflavin for chromophore formation. Riboflavin supplementation resulted in high background fluorescence. While this supplementation allowed for fluorescence measurements, the metabolic burden and resulting risk of mutations to the fluorescent genes ruled out the LOV-domain proteins as potential candidates for random genomic knock-in.

In contrast, FAST was found to be suitable for flow cytometry measurements, both with plasmid-based and genome-localized expression. It allowed differentiation between fluorescent and non-fluorescent populations by flow cytometry, the cells survived cell sorting and contamination was not observed. During the development of the present study, FAST was used in *E. limosum* and *A. woodii* as a protein tag for production optimisation of butanol and acetone (Flaiz et al., 2021), and lactate (Mook et al., 2022) respectively, confirming FAST as an efficient protein tag for anaerobes.

It is interesting to note that the *A. woodii* strain expressing the FAST reporter gene on an autonomous plasmid resulted in a heterogeneous fluorescent population. Previous studies had ruled out plasmid loss as the cause of this heterogeneity in *C. acetobutylicum* (Streett et al., 2019), it was also excluded in *A. woodii* by successful re-culturing of the non-fluorescence population in the presented thesis.

The second hypothesis for population heterogeneity involves cell morphogenesis differences in sporulating bacterial populations (Streett et al., 2019). During the

development of our study, another study in *A. woodii* (Mook et al., 2022) and one in the other non-sporulating Gram-positive *E. limosum* (Flaiz et al., 2021) also observed population heterogeneity, ruling out this hypothesis.

In *A. woodii*, up to 73.8% of cells were found to be non-fluorescent for a given time point (Mook et al., 2022). The percentage of highly fluorescent cells was dependent on the bacteria but also on the growth stage and promoter used for expression (Flaiz et al., 2021; Mook et al., 2022; Streett et al., 2019). It is also interesting to note that both for *A. woodii* and in the previous study in *C. acetobutylicum* (Streett et al., 2019), the non-fluorescent population of the plasmid-based expression strain had a higher median fluorescence than the WT strain.

The basis of this heterogeneity and higher background fluorescence remains elusive but the ability to knock-in the reporter gene cargo in the *A. woodii* genome allowed for confirmation that it is solely due to plasmid-based expression and is not inherent to the reporter gene. To avoid population heterogeneity, desirable metabolic pathways should be integrated into the genome.

While population heterogeneity confirmed the limitations of plasmid-based expression, lower fluorescence of the genome integrated FAST confirmed the limitations of genome integration. This was unsurprising as the experiments performed for acetone production demonstrated the reduced yield obtained with the genome integrated operon compared to the plasmid-based one.

Discovery of genome loci resulting in better production yields is crucial for circumventing the limitations of both plasmid-based and genome integration during the development of commercial strains.

5.3.4 Library analysis

After verification that *A. woodii* survived FACS and that no contamination could be detected in the resulting cultures, the pooled library was sorted. The library samples contained individual cells that displayed a fluorescence intensity similar, or higher, to the average fluorescence of the plasmid-based expression. Those cells constitute an interesting prospect as new genome loci for improved exogenous gene expression. The limitations of this technique and their potential resolution are discussed below.

Considering the library showed some plasmid retention during its construction (section 5.2.3.2.3), the main risk of the flow cytometry analysis was to obtain cells with high fluorescence only caused by the high copy number of the FAST gene from the plasmid. However, no colonies containing the plasmid could be found in the sorted population of highly fluorescent cells. Considering the library plasmids (pMTLGL15 and pMTLGL15-FAST) have no Gram-positive replicon, it is unlikely that the plasmid is maintained in the cells and it is likely integrated in the genome, as hypothesized for the library analysed in Chapter 4. In the case of the fluorescent library, the 0.0062% of cells still containing the plasmid would still only contain one copy of the FAST gene. The whole plasmid, also being randomly inserted, would not bias distribution of the plasmid containing cells depending on the fluorescence intensity. A hundred cells were sorted for each gate, the probability to obtain a cell containing the plasmid is thus the same as the general population (0.0062%), it is thus not surprising to not find clarithromycin resistant cells in the sorted populations.

Another risk is multiple insertions in the genome in each single mutant. The library built in this study was based on the library employed in Chapter 4, for which single insertions had been verified, however, verification that each mutant has a single transposon insertion remains crucial before further work can be performed. Unfortunately, the

inverse PCRs performed to check for single insertion in the mutants sorted were not successful. Verification of single insertions is essential before sequencing of the pooled library samples but also to enable sequencing of single mutants. This sequencing would allow the identification of the genome locus of insertion for each mutant. Sequencing of the transposon sequence inserted could be performed concomitantly with the sequencing of the insertion locus, to confirm that the increased fluorescence is not caused by mutations in the FAST gene. This sequencing procedure could be performed in both fluorescent and non-fluorescent populations to ensure that potential mutations observed in the fluorescent population are not present in the non-fluorescent population.

It is also important to check that the relative fluorescence of the sorted mutants is conserved when measuring fluorescence in plate readers. Sorting was done with different groups for both fluorescent and non-fluorescent populations to be able to isolate single mutants and assess their fluorescence using a plate reader. Comparison of the average fluorescence of the single mutants sorted through each gate will help confirm that the relative fluorescence of the different groups is maintained after re-culturing. Association of the fluorescence of single mutants with their respective genome locus would also inform on the correlation between genome locus insertion and fluorescence obtained.

It is possible that there is a bias on the mutants surviving the cell sorting protocol, as some beneficial loci might be more sensitive to the oxygen exposure they experience. The cells sorted from the non-fluorescent population of the library could serve as a test for survival rate. Assuming a similar number of single insertion mutant in the starting library pool as with the library in Chapter 4, it is unlikely that each gate from the non-fluorescent population contains multiple copies of the same single mutants. Accordingly, sequencing of the resulting population would allow for estimation of the

survival rate of the cells through the sorting protocol. Successful inverse PCR and subsequent sequencing of the populations gated and of the single mutants isolated would enable resolution to all the above limitations.

The only limitation that remains is the transferability of the higher expression of FAST to pathways of interest. Saleski and co-workers found that the transcript level of the genes did not always correlate with level of production of metabolites (Saleski et al., 2021). They hypothesized that there is a fine balance of cell burden, with higher expression resulting in a too high cell burden and decreased metabolite production. It might be interesting to measure mRNA levels of single colonies of the library to correlate them with their fluorescence. More importantly, since the production of a reporter gene likely exerts a very different burden on the cells compared to a complete metabolic pathway, it highlights that insertion sites found to be favorable for FAST expression might not be favorable for metabolic pathways of interest. Ultimately, only the insertion of gene(s) of interest in the loci identified could validate their benefits for improved metabolite production. An acetone knock-in at the *pheA* locus was generated to function as a control for comparison with expression of the operon from one of the potentially discovered genome loci.

The final objective of the study to use acetone and isopropanol production as an example for the characterisation of genome loci resulting in improved yields was not achieved, but the experiments are still very promising for the future. Confirmation that the FAST reporter gene can be used for flow cytometry in acetogens (Flaiz et al., 2021; Mook et al., 2022; Streett et al., 2019) and proof-of-concept that cell sorting is possible with subsequent re-culture of the cells opens new avenues for research in acetogens.

5.4 Key outcomes

Plasmid-based expression of the acetone operon resulted in relatively high yields but requires antibiotic selection for maintenance. Acetone yields when genes were localized to the *pheA* genome locus resulted in lower acetone levels. Expression of the FAST fluorescent reporter gene confirmed population heterogeneity linked to plasmid-based expression of genes. Overall, this study confirmed the need to identify new genome loci to improve the expression of pathways of interest. A technique was successfully devised to enable such discoveries. The FAST reporter gene combined to flow cytometry enabled differentiation between fluorescent and non-fluorescent cell populations and *A. woodii* survived FACS, enabling future characterization and sequencing of individual cells.

CHAPTER 6: General Discussion

6.1 Context

Acetogens represent excellent candidates for valorisation of C1 gases by solvent production using the Wood-Ljungdahl Pathway. While they are already used in industrial processes (Liew et al., 2016), improvements can be achieved through the application of synthetic biology for strain optimization. In *A. woodii*, attempts at gene overexpression to increase the carbon flow through the WLP showed a maximum acetate concentration of approximately 50 g/L (Kantzow & Weuster-Botz, 2016; Straub et al., 2014). The diversification of products has been achieved by expression of exogenous pathways for acetone and isopropanol production (Hoffmeister et al., 2016) and further studies have focused on promoters, RBS, and gene sequence for yield improvements (Arslan et al., 2022).

This study describes the development and application of synthetic biology tools for the better characterisation of *A. woodii* as an industrial chassis. As we move towards developing industrial strains, the requirement for genome insertion becomes critical, it was, however, shown to result in decreased target metabolite yields compared to plasmid-based expression. An understanding of the impact of genome position on expression could allow genome-integrated pathways to result in yields similar to plasmid-based expression (Saleski et al., 2021). Additionally, a better understanding of gene function and gene networks could allow for improvement of growth and yields based on direct modification of *A. woodii* genome.

6.2 Summary of findings

A Type I N^6-adenine methyltransferase system was hypothesized to impede transformation of plasmids containing *cas9*. The system comprises a specificity (*hsdS*), a methylation (*hsdM*), and a restriction (*hsdR*) subunit and recognizes the site

5′-TAAGN$_5$TCC-3′. The presence of the site significantly reduced transformation efficiencies of an empty backbone. Accordingly, a silent mutation was introduced in the *cas9* to study the isolated impact of the Cas9 toxicity. Unsurprisingly, Cas9 was found to be toxic to the cells even when non-induced and in the absence of sgRNA. Catalytically inactive Cas9 was similarly found to be toxic to cells but plasmid size might have participated to low transformation efficiencies.

Exploiting endogenous CRISPR systems offers an alternative solution to using the widely adopted *S. pyogenes* Cas9, which circumvents toxicity issues associated with using the heterologous nuclease while at the same time dramatically reducing the size of the genome editing vectors employed. At the time of the study, no tool was available for identification of functional PAM, relying instead on manual *in silico* analyses (Pyne et al., 2016) or tedious library construction and large-scale sequencing (Walker et al., 2020). The Python script developed to remedy this problem was successfully validated against the current literature of identified functional PAMs in *Clostridium*, demonstrating its utility. It subsequently enabled the characterisation of PAM sequences in a range of bacteria, including *A. woodii*. This tool is likely to prove useful for genetic engineering of novel microorganisms in which low transformation efficiencies hinder the use of Cas9 because of both its toxicity, its size, and the presence of potential R-M sites.

In *A. woodii*, the python script predictions were validated *in vivo,* and a double cytosine was found to be crucial for targeting of DNA by the Type I-B system. The system was initially validated by making a 300 bp in-frame deletion in the *pyrE* gene of *A. woodii*. Thereafter, the creation of a 354 bp deletion mutant of *pheA* enabled further characterisation of optimal parameters for genome editing with the system. These included demonstration that increasing the size of the homology arms used in the editing

template to 1.5 kb led to improved editing efficiencies and that mutants were generated after both a 5 h and 7.5 h recovery period. As expected, increasing the amounts of cells and DNA, led to increase in editing efficiencies when generating a 3.2 kb deletion of the *hsdR1* gene. Finally, utilization of the pCD6 Gram-positive replicon enabled rapid plasmid loss, which allowed for confirmation of the *pheA* and *hsdR1* deletion phenotypes. As expected, *hsdR1* was found responsible for the degradation of plasmids containing the recognition site 5′-TAAGN_5TCC-3′.

The pipeline developed was subsequently applied to *C. autoethanogenum* from *in silico* PAM identification to a 562 bp *pyrE* deletion. This represented the first successful use of *A. woodii* and *C. autoethanogenum* endogenous CRISPR systems for genetic engineering. Overall, these developments represent a significant improvement to genome editing in both acetogens, particularly *A. woodii* and showed great potential for application in other *Clostridium spp*.

After tool development, a transposon library was exploited to identify genes that are essential for heterotrophic and autotrophic fermentation as a first step towards making rational improvement to fermentation performance, particularly on C1 gas. Importantly, in addition to identifying essential genes, a study of read abundance at the gene-level and at the single mutant level was undertaken. This confirmed genes involved under both heterotrophic and autotrophic conditions. More interestingly, some of the data that emerged conflicted established views of the role of energy conservation mechanisms in *A. woodii*. The Rnf complex was not found to be essential for autotrophic growth and while the Hyd hydrogenase was found to be important for heterotrophic growth, it was found to be detrimental for autotrophic growth. On the other hand, the less studied Stn

complex was found to be essential in all the conditions tested. These observations open new avenues for future research.

Additionally, the study revealed the identity of genes detrimental to heterotrophic or autotrophic growth. The annotated alanine dehydrogenase pathway, and both *mttC7* and Awo_c03590, both of unknown function, were found to be detrimental in either growth condition. Detrimental genes with unknown functions represent an exciting prospect for genetic engineering to improve growth kinetics and product yields.

In *A. woodii* native conversion of acetone into isopropanol was hypothesized to reduce maximum acetone yields. While overexpression of the exogenous secondary alcohol dehydrogenase from *C. beijerinckii* in an acetone producing strain led to increase in isopropanol concentration, the native alcohol dehydrogenase responsible for the conversion was not successfully identified. Another crucial element for high yields was the position of the expressed gene, expression of the SADH at the end of the acetone operon was shown to result in lower yields than when expressed with a separate promoter. Similarly, plasmid-based expression of the acetone operon was also shown to result in significantly higher yields than when the operon was integrated into the chromosome. However, genome integration is required for industrial processes in the interest of cost and strain stability. To identify genome loci favourable to high gene expression, a strategy involving a transposon library containing a fluorescent reporter gene followed by FACS was developed. Reporter genes based on LOV-domains were ruled out as their requirement in riboflavin heightened the risk of mutation in the gene. The FAST system, on the contrary proved to be adequate and an initial cell sorting experiment demonstrated survival of *A. woodii* during the procedure and successful re-

culture. The transposon library was constructed following the strategy employed for building the transposon library used in Chapter 4 with the FAST reporter gene alongside the *catP* gene in the transposon cassette. Flow cytometry successfully revealed cells displaying fluorescence comparable to plasmid-based expression. Cells with high fluorescence were sorted and re-cultured. Unfortunately, single insertions could not be verified by inverse PCR and the samples were not sequenced due to time constraints.

6.3 Further studies

The high-throughput method of transposon libraries combined with essentiality and abundance studies provided a wealth of data that can be used to guide future work in *A. woodii*. The study of essential genes specific to growth in gas fermenters could provide valuable insight into the regulatory gene networks in place.

Results apparently contradictory to the current knowledge on *A. woodii* need to be investigated, namely the apparent non-essentiality of the HDCR, Rnf, and Hyd complexes. First, the separate KO of the two identified formate dehydrogenase complexes of the HDCR would allow confirmation that *fhdF1/hycB1* can compensate for the loss of *fdhF2/hycB2* (Ceccaldi et al., 2017). Second, study of compensation mechanisms in the absence of the Hyd hydrogenase (Δ*hydAB* strain from Wiechmann and co-workers (2020)) or with a downregulated Hyd, is crucial and could provide tremendous insight into hydrogen utilization and redox balancing during autotrophic fermentation. Lastly, population mechanisms and individual subunits deletion of the Rnf complex could potentially elucidate the apparent non-essentiality of the Rnf in the library population, as the deletion of the whole complex has been shown to abolish autotrophic growth (Westphal et al., 2018).

The Na^+ cycle for energy conservation in *A. woodii* is well studied, but a better understanding of proton transport and potential energetic implications is necessary, as demonstrated by the detrimental effect of the V-Type ATPase during autotrophic growth. Its deletion would enable validation of its impact on growth.

The highly detrimental genes identified are ideal deletion targets for increased growth kinetics and potential increased acetate yields. Among the genes identified, genes *mttC7* and Awo_c03590 represent very interesting candidates for future study as their detrimental effect was extreme and their functions are unknown. Additionally, alanine was only reported as a fermentation product in phosphate limited conditions (Buschhorn, Dürre, et al., 1989). A more in-depth study of the detrimental alanine dehydrogenase pathway would help better understand the conditions in which production is initiated and the reason for its detrimental effect on autotrophic fermentation.

Deletion of detrimental genes would also confirm the correlation between increasing number of reads and improved fitness in the condition studied. It would rule out the possibility of social cheating in the context of the transposon library and the study of their phenotype could also elucidate the reason for their detrimental effect on growth. From the transposon data it cannot be concluded whether they are detrimental solely because of the energy conserved by not producing the protein they encode or because of underlying metabolic imbalances. These metabolic imbalances would also be interesting from an evolutionary point of view. One would expect that detrimental genes would be tightly regulated to be expressed only in favourable conditions.

After confirmation of the phenotype of all the systems mentioned, it would be particularly interesting to feedback the obtained data into a metabolic model of *A. woodii* for a better understanding of the energetics and metabolic balance in *A. woodii*.

The remaining aliquots of the transposon library could be used to elucidate essential and detrimental genes in new conditions. They could be used to further the knowledge of metabolic networks on different substrates. This is particularly interesting for substrates that demonstrate optimal threshold for growth such as formate, methanol, or CO (co-utilized with H_2/CO_2). Syngas is a mixture of H_2 and CO_2 but also CO, *A. woodii* was shown to co-utilize CO when grown on H_2/CO_2, to levels of up to 15%, higher levels of CO were detrimental for growth (Bertsch & Muller, 2015b). To avoid the drawbacks of gas fermentation methanol and formate also represent attractive feedstocks as they are easy to store and avoid gas diffusion challenges. Formate concentrations of 200 to 300 mM led to optimal growth while a concentration of 1,000 mM had an apparent inhibitory effect on growth (Moon et al., 2021). Similarly, 60 mM methanol was optimal and higher concentrations decreased the growth rate with a limit of toxicity of 900 mM (Kremp et al., 2018). Concentration of these substrates could be incrementally increased during fermentation to allow adaption of the library. Evolution of reads abundance in genes would inform on the metabolic networks in operation at high concentrations.

A similar approach could be used for improved tolerance to the solvents produced. Increasing acetone concentrations of up to 500 mM resulted in slower cell growth but were tolerated by *A. woodii*, higher concentrations led to culture collapse (Baker, 2020). Preliminary work has established candidate gene networks responsible for an increased

tolerance. This was achieved by subculturing the transposon library twice in 0.2M, 0.3M, 0.4M, or 0.5M of acetone (Baker, 2020). However, no replicates were performed and the genes with increasing reads were not studied. Detrimental genes could be studied retrospectively but continuous fermentation with increasing amounts of acetone overtime represents a more interesting prospect for identification of essential and detrimental genes participating in tolerance mechanisms.

Expression of exogenous genes successfully led to acetone and isopropanol production. Attempts at elucidating the native alcohol dehydrogenase responsible for conversion of acetone into isopropanol were unsuccessful. *A. woodii* contains 11 alcohol dehydrogenases, successive KO of the ones that remain uncharacterized with the genomic engineering tool developed in Chapter 3 could lead to its successful identification and increase in acetone yields.

For the transposon library containing the FAST reporter gene and sorted by FACS, validation of the presence of unique and varied insertion sites is crucial before sequencing of the samples. New primers could be designed to resolve the inverse PCR issues encountered. As an alternative, the HindIII digestions could be used for Southern blotting with a probe specific to *catP* (Cartman & Minton, 2010; Liu et al., 2013). Another crucial verification is the correlation of higher fluorescence observed during flow cytometry to higher fluorescence in plate reader experiments. Plate reader assays need to be performed to verify that they, indeed, demonstrated higher fluorescence than the *pheA* KI strain. Finally, sequencing needs to be performed to potentially reveal favourable insertion sites for FAST expression. Both acetone and isopropanol production could be used to assess the loci identified. The acetone operon could be

inserted in these loci in a strain already containing *tcdR* at *pyrE* or the isopropanol gene could be inserted in the strain containing both *tcdR* and the acetone operon already inserted in the genome. The resulting strains could be analysed for improved yields. Better expression obtained with a reporter gene is likely to not directly correlate with better expression of more complex pathways because of increased metabolic burden on the cells. To circumvent this problem, the whole pathway could be inserted in the transposon element and a new library created. However, transposition frequency was shown to decrease with increasing transposon sequence length (Lampe et al., 1998) and sorting of the cells would require the creation of a sensor strain with acetone- or isopropanol-dependent fluorescence expression (Saleski et al., 2021).

Further studies discussed here rely heavily on the ability to knock-out the genes identified with the transposon library and knock-in of the FAST reporter gene, the acetone operon or the alcohol dehydrogenase at the loci identified by the FAST transposon library. While the endogenous CRISPR system developed in Chapter 3 allowed for the successful deletion of several genes and insertion of a FAST reporter cassette (614 bp) or the whole acetone operon (3,716 bp) at the *pheA* locus, the initial attempt at inserting the alcohol dehydrogenase from *C. beijerinckii* at the *purD* locus was unsuccessful. Alternative homology arm or spacer design is likely to resolve the issue, but this highlighted a need for consolidation of the endogenous CRISPR system. The main area that could benefit from more study is the impact of spacer sequence on interference. It was shown that the same PAM could lead to different outcomes in *E. coli* Type I-E system depending on the spacer (Xue et al., 2015). A more systematic study of the spacer sequences leading to successful interference in *A. woodii* is necessary.

References

Adhikari, S., & Curtis, P. D. (2016). DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiol Rev*, *40*(5), 575-591. https://doi.org/10.1093/femsre/fuw023

Ahmad, I., Anwar, Z., Ahmed, S., Sheraz, M. A., & Khattak, S. U. (2017). Metal ion mediated photolysis reactions of riboflavin: A kinetic study. *J Photochem Photobiol B*, *173*, 231-239. https://doi.org/10.1016/j.jphotobiol.2017.05.033

Alkhnbashi, O. S., Shah, S. A., Garrett, R. A., Saunders, S. J., Costa, F., & Backofen, R. (2016). Characterizing leader sequences of CRISPR loci. *Bioinformatics*, *32*(17), i576-i585. https://doi.org/10.1093/bioinformatics/btw454

Álvarez, A., Bansode, A., Urakawa, A., Bavykina, A. V., Wezendonk, T. A., Makkee, M., . . . Kapteijn, F. (2017). Challenges in the Greener Production of Formates/Formic Acid, Methanol, and DME by Heterogeneously Catalyzed CO. *Chem Rev*, *117*(14), 9804-9838. https://doi.org/10.1021/acs.chemrev.6b00816

Andreesen, J. R. (1994). Glycine metabolism in anaerobes. *Antonie Van Leeuwenhoek*, *66*(1-3), 223-237. https://doi.org/10.1007/BF00871641

Arcus, V. L., McKenzie, J. L., Robson, J., & Cook, G. M. (2011). The PIN-domain ribonucleases and the prokaryotic VapBC toxin-antitoxin array. *Protein Eng Des Sel*, *24*(1-2), 33-40. https://doi.org/10.1093/protein/gzq081

Arndt, D., Marcu, A., Liang, Y., & Wishart, D. S. (2019). PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes. *Brief Bioinform*, *20*(4), 1560-1567. https://doi.org/10.1093/bib/bbx121

Arslan, K., Schoch, T., Hofele, F., Herrschaft, S., Oberlies, C., Bengelsdorf, F., . . . Kennes, C. (2022). Engineering *Acetobacterium woodii* for the production of isopropanol and acetone from carbon dioxide and hydrogen. *Biotechnol J*, *17*(5), e2100515. https://doi.org/10.1002/biot.202100515

Atmadjaja, A. N., Holby, V., Harding, A. J., Krabben, P., Smith, H. K., & Jenkinson, E. R. (2019). CRISPR-Cas, a highly effective tool for genome editing in *Clostridium saccharoperbutylacetonicum* N1-4(HMT). *FEMS Microbiol Lett*, *366*(6). https://doi.org/10.1093/femsle/fnz059

Aufurth, S., Schägger, H., & Müller, V. (2000). Identification of subunits a, b, and c1 from *Acetobacterium woodii* Na+-F1F0-ATPase. Subunits c1, c2, AND c3 constitute a mixed c-oligomer. *J Biol Chem*, *275*(43), 33297-33301. https://doi.org/10.1074/jbc.M005134200

Bailey, R. (2008). *Another inconvenient truth: How biofuel policies are deepening poverty and accelerating climate change* (Vol. 114). Oxfam.

Baker, J. P. (2020). *ENGICOIN deliverable D6.2*. https://engicoin.eu/results/deliverables

Baker, J. P. (2021). *ENGICOIN deliverable D6.4*.

Baker, J. P., Saez-Saez, J., Jensen, S. I., Nielsen, A. T., & Minton, N. P. (2022). A clean in-frame knockout system for gene deletion in *Acetobacterium woodii*. *J Biotechnol*, *353*, 9-18. https://doi.org/10.1016/j.jbiotec.2022.05.013

Balch, W. E., Schoberth, S., Tanner, R. S., & Wolfe, R. S. (1977). *Acetobacterium*, a New Genus of Hydrogen-Oxidizing, Carbon Dioxide-Reducing, Anaerobic Bacteria. *Int J Syst Bacteriol*, *27*(4), 355-361. https://doi.org/Doi 10.1099/00207713-27-4-355

Bannam, T. L., & Rood, J. I. (1993). *Clostridium perfringens-Escherichia coli* shuttle vectors that carry single antibiotic resistance determinants. *Plasmid*, *229*, 233-235.

Bar-Even, A., Noor, E., Flamholz, A., & Milo, R. (2013). Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. *Biochim Biophys Acta*, *1827*(8-9), 1039-1047. https://doi.org/10.1016/j.bbabio.2012.10.013

Barquist, L., Mayho, M., Cummins, C., Cain, A. K., Boinett, C. J., Page, A. J., . . . Parkhill, J. (2016). The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics*, *32*(7), 1109-1111. https://doi.org/10.1093/bioinformatics/btw022

Becher, B., & Müller, V. (1994). Delta mu Na+ drives the synthesis of ATP via an delta mu Na(+)-translocating F1F0-ATP synthase in membrane vesicles of the archaeon *Methanosarcina mazei* Gö1. *J Bacteriol*, *176*(9), 2543-2550. https://doi.org/10.1128/jb.176.9.2543-2550.1994

Beck, M. H., Flaiz, M., Bengelsdorf, F. R., & Dürre, P. (2020). Induced heterologous expression of the arginine deiminase pathway promotes growth advantages in the strict anaerobe Acetobacterium woodii. *Appl Microbiol Biotechnol*, *104*(2), 687-699. https://doi.org/10.1007/s00253-019-10248-9

Bellais, S., Nehlich, M., Ania, M., Duquenoy, A., Mazier, W., van den Engh, G., . . . Thomas, V. (2022). Species-targeted sorting and cultivation of commensal bacteria from the gut microbiome using flow cytometry under anaerobic conditions. *Microbiome*, *10*(1), 24. https://doi.org/10.1186/s40168-021-01206-7

Bengelsdorf, F. R., Poehlein, A., Linder, S., Erz, C., Hummel, T., Hoffmeister, S., . . . Durre, P. (2016). Industrial Acetogenic Biocatalysts: A Comparative Metabolic and Genomic Analysis. *Front Microbiol*, *7*, 1036. https://doi.org/10.3389/fmicb.2016.01036

Berg, I. A., Kockelkorn, D., Ramos-Vera, W. H., Say, R. F., Zarzycki, J., Hügler, M., . . . Fuchs, G. (2010). Autotrophic carbon fixation in archaea. *Nat Rev Microbiol*, *8*(6), 447-460. https://doi.org/10.1038/nrmicro2365

Bertsch, J., & Muller, V. (2015a). Bioenergetic constraints for conversion of syngas to biofuels in acetogenic bacteria. *Biotechnol Biofuels*, *8*, 210. https://doi.org/10.1186/s13068-015-0393-x

Bertsch, J., & Muller, V. (2015b). CO Metabolism in the Acetogen Acetobacterium woodii. *Appl Environ Microbiol*, *81*(17), 5949-5956. https://doi.org/10.1128/AEM.01772-15

Bertsch, J., Siemund, A. L., Kremp, F., & Muller, V. (2016). A novel route for ethanol oxidation in the acetogenic bacterium *Acetobacterium woodii*: the acetaldehyde/ethanol dehydrogenase pathway. *Environ Microbiol*, *18*(9), 2913-2922. https://doi.org/10.1111/1462-2920.13082

Bertsch, J., Öppinger, C., Hess, V., Langer, J. D., & Müller, V. (2015). Heterotrimeric NADH-oxidizing methylenetetrahydrofolate reductase from the acetogenic bacterium *Acetobacterium woodii*. *J Bacteriol*, *197*(9), 1681-1689. https://doi.org/10.1128/jb.00048-15

Biegel, E., & Müller, V. (2010). Bacterial Na+-translocating ferredoxin:NAD+ oxidoreductase. *Proc Natl Acad Sci*, *107*(42), 18138-18142. https://doi.org/10.1073/pnas.1010318107

Biegel, E., Schmidt, S., & Müller, V. (2009). Genetic, immunological and biochemical evidence for a Rnf complex in the acetogen *Acetobacterium woodii*

[https://doi.org/10.1111/j.1462-2920.2009.01871.x]. *Environ Microbiol*, *11*(6), 1438-1443. https://doi.org/https://doi.org/10.1111/j.1462-2920.2009.01871.x
Biswas, A., Fineran, P. C., & Brown, C. M. (2015). Computational Detection of CRISPR/crRNA Targets. *Methods Mol Biol*, *1311*, 77-89. https://doi.org/10.1007/978-1-4939-2687-9_5
Biswas, A., Gagnon, J. N., Brouns, S. J., Fineran, P. C., & Brown, C. M. (2013). CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA Biol*, *10*(5), 817-827. https://doi.org/10.4161/rna.24046
Blow, M. J., Clark, T. A., Daum, C. G., Deutschbauer, A. M., Fomenkov, A., Fries, R., . . . Roberts, R. J. (2016). The Epigenomic Landscape of Prokaryotes. *PLoS Genet*, *12*(2), e1005854. https://doi.org/10.1371/journal.pgen.1005854
Boudry, P., Semenova, E., Monot, M., Datsenko, K. A., Lopatina, A., Sekulovic, O., . . . Soutourina, O. (2015). Function of the CRISPR-Cas System of the Human Pathogen *Clostridium difficile*. *mBio*, *6*(5), e01112-01115. https://doi.org/10.1128/mBio.01112-15
Bourgeois, J., Lazinski, D. W., & Camilli, A. (2020). Identification of Spacer and Protospacer Sequence Requirements in the *Vibrio cholerae* Type I-E CRISPR/Cas System. *mSphere*, *5*(6). https://doi.org/10.1128/mSphere.00813-20
Brandt, K., Müller, D. B., Hoffmann, J., Hübert, C., Brutschy, B., Deckers-Hebestreit, G., & Müller, V. (2013). Functional production of the Na+ F1FO ATP synthase from *Acetobacterium woodii* in *Escherichia coli* requires the native AtpI. *J Bioenerg Biomembr*, *45*(1), 15-23. https://doi.org/10.1007/s10863-012-9474-8
Breitkopf, R., Uhlig, R., Drenckhan, T., & Fischer, R. J. (2016). Two propanediol utilization-like proteins of *Moorella thermoacetica* with phosphotransacetylase activity. *Extremophiles*, *20*(5), 653-661. https://doi.org/10.1007/s00792-016-0854-6
Bruder, M. R., Pyne, M. E., Moo-Young, M., Chung, D. A., & Chou, C. P. (2016). Extending CRISPR-Cas9 Technology from Genome Editing to Transcriptional Engineering in the Genus *Clostridium*. *Appl Environ Microbiol*, *82*(20), 6109-6119. https://doi.org/10.1128/AEM.02128-16
Bryant, J. A., Sellars, L. E., Busby, S. J., & Lee, D. J. (2014). Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res*, *42*(18), 11383-11392. https://doi.org/10.1093/nar/gku828
Buschhorn, H., Durre, P., & Gottschalk, G. (1989). Production and Utilization of Ethanol by the Homoacetogen *Acetobacterium woodii*. *Appl Environ Microbiol*, *55*(7), 1835-1840.
Buschhorn, H., Dürre, P., & Gottschalk, G. (1989). Production and Utilization of Ethanol by the Homoacetogen *Acetobacterium woodii*. *Appl Environ Microbiol*, *55*(7), 1835-1840. https://doi.org/10.1128/aem.55.7.1835-1840.1989
Canadas, I. C., Groothuis, D., Zygouropoulou, M., Rodrigues, R., & Minton, N. P. (2019). RiboCas: A Universal CRISPR-Based Editing Tool for *Clostridium*. *ACS Synth Biol*, *8*(6), 1379-1390. https://doi.org/10.1021/acssynbio.9b00075
Canez, C., Selle, K., Goh, Y. J., & Barrangou, R. (2019). Outcomes and characterization of chromosomal self-targeting by native CRISPR-Cas systems in *Streptococcus thermophilus*. *FEMS Microbiol Lett*, *366*(9). https://doi.org/10.1093/femsle/fnz105

Carius, Y., Christian, H., Faust, A., Zander, U., Klink, B. U., Kornberger, P., . . . Scheidig, A. J. (2010). Structural insight into substrate differentiation of the sugar-metabolizing enzyme galactitol dehydrogenase from *Rhodobacter sphaeroides* D. *J Biol Chem*, *285*(26), 20006-20014. https://doi.org/10.1074/jbc.M110.113738

Cartman, S. T., & Minton, N. P. (2010). A mariner-based transposon system for in vivo random mutagenesis of *Clostridium difficile*. *Appl Environ Microbiol*, *76*(4), 1103-1109. https://doi.org/10.1128/AEM.02525-09

Ceccaldi, P., Schuchmann, K., Müller, V., & Elliott, S. J. (2017). The hydrogen dependent CO2 reductase: the first completely CO tolerant FeFe-hydrogenase [10.1039/C6EE02494G]. *Energy Environ Sci*, *10*(2), 503-508. https://doi.org/10.1039/C6EE02494G

Chapman, S., Faulkner, C., Kaiserli E., Garcia-Mata C., Savenkov E. I., Roberts A. G. Oparka K. J., Christie J. M. (2008). The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *PNAS*, *105*(50), 20038-20043.

Chari, R., Mali, P., Moosburner, M., & Church, G. M. (2015). Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods*, *12*(9), 823-826. https://doi.org/10.1038/nmeth.3473

Cho, C., & Lee, S. Y. (2017). Efficient gene knockdown in *Clostridium acetobutylicum* by synthetic small regulatory RNAs. *Biotechnol Bioeng*, *114*(2), 374-383. https://doi.org/10.1002/bit.26077

Chowdhury, N. P., Alberti, L., Linder, M., & Muller, V. (2020). Exploring Bacterial Microcompartments in the Acetogenic Bacterium *Acetobacterium woodii*. *Front Microbiol*, *11*, 593467. https://doi.org/10.3389/fmicb.2020.593467

Chowdhury, N. P., Alberti, L., Linder, M., & Müller, V. (2020). Exploring Bacterial Microcompartments in the Acetogenic Bacterium. *Front Microbiol*, *11*, 593467. https://doi.org/10.3389/fmicb.2020.593467

Chowdhury, N. P., Litty, D., & Müller, V. (2022). Biosynthesis of butyrate from methanol and carbon monoxide by recombinant *Acetobacterium woodii*. *Int Microbiol*, *25*(3), 551-560. https://doi.org/10.1007/s10123-022-00234-z

Chowdhury, N. P., Moon, J., & Müller, V. (2021). Adh4, an alcohol dehydrogenase controls alcohol formation within bacterial microcompartments in the acetogenic bacterium *Acetobacterium woodii*. *Environ Microbiol*, *23*(1), 499-511. https://doi.org/10.1111/1462-2920.15340

Christie, J. M., Hitomi, K., Arvai, A. S., Hartfield, K. A., Mettlen, M., Pratt, A. J., . . . Getzoff, E. D. (2012). Structural tuning of the fluorescent protein iLOV for improved photostability. *J Biol Chem*, *287*(26), 22295-22304. https://doi.org/10.1074/jbc.M111.318881

Claudel-Renard, C., Chevalet, C., Faraut, T., & Kahn, D. (2003). Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res*, *31*(22), 6633-6639. https://doi.org/10.1093/nar/gkg847

Cotter, P. D., & Hill, C. (2003). Surviving the acid test: responses of gram-positive bacteria to low pH. *Microbiol Mol Biol Rev*, *67*(3), 429-453, table of contents. https://doi.org/10.1128/MMBR.67.3.429-453.2003

Cotton, C. A., Claassens, N. J., Benito-Vaquerizo, S., & Bar-Even, A. (2020). Renewable methanol and formate as microbial feedstocks. *Curr Opin Biotechnol*, *62*, 168-180. https://doi.org/10.1016/j.copbio.2019.10.002

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Neron, B., . . . Pourcel, C. (2018). CRISPRCasFinder, an update of CRISRFinder, includes

a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res*, *46*(W1), W246-W251. https://doi.org/10.1093/nar/gky425

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res*, *14*(6), 1188-1190. https://doi.org/10.1101/gr.849004

Dembek, M., Barquist, L., Boinett, C. J., Cain, A. K., Mayho, M., Lawley, T. D., . . . Fagan, R. P. (2015). High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. *mBio*, *6*(2), e02383. https://doi.org/10.1128/mBio.02383-14

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., . . . Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol*, *32*(12), 1262-1267. https://doi.org/10.1038/nbt.3026

Donig, J., & Muller, V. (2018). Alanine, a Novel Growth Substrate for the Acetogenic Bacterium Acetobacterium woodii. *Appl Environ Microbiol*, *84*(23). https://doi.org/10.1128/AEM.02023-18

Drake, H. L., Gössner, A. S., & Daniel, S. L. (2008). Old acetogens, new light. *Ann N Y Acad Sci*, *1125*, 100-128. https://doi.org/10.1196/annals.1419.016

Drepper, T., Eggert, T., Circolone, F., Heck, A., Krauss, U., Guterl, J. K., . . . Jaeger, K. E. (2007). Reporter proteins for in vivo fluorescence without oxygen. *Nat Biotechnol*, *25*(4), 443-445. https://doi.org/10.1038/nbt1293

Eden, G., & Fuchs, G. (1983). Autotrophic CO_2 fixation in *Acetobacterium woodii*. In (Vol. 135, pp. 68-73). Arch. Microbiol.

Eichler, B., & Schink, B. (1984). Oxidation of primary aliphatic alcohols by *Acetobacterium* carbinolicum so. nov., a homoacetogenic anaerobe. *Arch. Microbiol.*, 147-152. https://doi.org/10.1007/BF00454917

Ellis, D. J., Dryden, D. T., Berge, T., Edwardson, J. M., & Henderson, R. M. (1999). Direct observation of DNA translocation and cleavage by the EcoKI endonuclease using atomic force microscopy. *Nat Struct Biol*, *6*(1), 15-17. https://doi.org/10.1038/4882

Fagan, R. P., & Fairweather, N. F. (2011). Clostridium difficile has two parallel and essential Sec secretion systems. *J Biol Chem*, *286*(31), 27483-27493. https://doi.org/10.1074/jbc.M111.263889

Fast, A. G., & Papoutsakis, E. T. (2012). Stoichiometric and energetic analyses of non-photosynthetic CO2-fixation pathways to support synthetic biology strategies for production of fuels and chemicals. *Curr Opin Chem Eng*, *1*(4), 380-395. https://doi.org/https://doi.org/10.1016/j.coche.2012.07.005

Fischer, S., Maier, L. K., Stoll, B., Brendel, J., Fischer, E., Pfeiffer, F., . . . Marchfelder, A. (2012). An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. *J Biol Chem*, *287*(40), 33351-33363. https://doi.org/10.1074/jbc.M112.377002

Flaiz, M., Ludwig, G., Bengelsdorf, F. R., & Durre, P. (2021). Production of the biocommodities butanol and acetone from methanol with fluorescent FAST-tagged proteins using metabolically engineered strains of *Eubacterium limosum*. *Biotechnol Biofuels*, *14*(1), 117. https://doi.org/10.1186/s13068-021-01966-2

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., . . . Turner, S. W. (2010). Direct detection of DNA methylation during

single-molecule, real-time sequencing. *Nat Methods*, *7*(6), 461-465. https://doi.org/10.1038/nmeth.1459
Forgac, M. (1999). Structure and properties of the vacuolar (H+)-ATPases. *J Biol Chem*, *274*(19), 12951-12954. https://doi.org/10.1074/jbc.274.19.12951
Fowler, C. C., & Galán, J. E. (2018). Decoding a *Salmonella Typhi* Regulatory Network that Controls Typhoid Toxin Expression within Human Cells. *Cell Host Microbe*, *23*(1), 65-76.e66. https://doi.org/10.1016/j.chom.2017.12.001
Fritz, M., & Müller, V. (2007). An intermediate step in the evolution of ATPases--the F1F0-ATPase from *Acetobacterium woodii* contains F-type and V-type rotor subunits and is capable of ATP synthesis. *FEBS J*, *274*(13), 3421-3428. https://doi.org/10.1111/j.1742-4658.2007.05874.x
Furdui, C., & Ragsdale, S. W. (2000). The role of pyruvate ferredoxin oxidoreductase in pyruvate synthesis during autotrophic growth by the Wood-Ljungdahl pathway. *J Biol Chem*, *275*(37), 28494-28499. https://doi.org/10.1074/jbc.M003291200
García-Contreras, R., & Loarca, D. (2020). The bright side of social cheaters: potential beneficial roles of "social cheaters" in microbial communities. *FEMS Microbiol Ecol*, *97*(1). https://doi.org/10.1093/femsec/fiaa239
Gawronski, J. D., Wong, S. M., Giannoukos, G., Ward, D. V., & Akerley, B. J. (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc Natl Acad Sci U S A*, *106*(38), 16422-16427. https://doi.org/10.1073/pnas.0906627106
Geddes, C. C., Mullinnix, M. T., Nieves, I. U., Peterson, J. J., Hoffman, R. W., York, S. W., . . . Ingram, L. O. (2011). Simplified process for ethanol production from sugarcane bagasse using hydrolysate-resistant Escherichia coli strain MM160. *Bioresour Technol*, *102*(3), 2702-2711. https://doi.org/10.1016/j.biortech.2010.10.143
Gomaa, A. A., Klumpe, H. E., Luo, M. L., Selle, K., Barrangou, R., & Beisel, C. L. (2014). Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *mBio*, *5*(1), e00928-00913. https://doi.org/10.1128/mBio.00928-13
Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., . . . Gordon, J. I. (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe*, *6*(3), 279-289. https://doi.org/10.1016/j.chom.2009.08.003
Grissa, I., Vergnaud, G., & Pourcel, C. (2007a). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*, *35*(Web Server issue), W52-57. https://doi.org/10.1093/nar/gkm360
Grissa, I., Vergnaud, G., & Pourcel, C. (2007b). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*, *8*, 172. https://doi.org/10.1186/1471-2105-8-172
Grüber, G., Manimekalai, M. S., Mayer, F., & Müller, V. (2014). ATP synthases from archaea: the beauty of a molecular motor. *Biochim Biophys Acta*, *1837*(6), 940-952. https://doi.org/10.1016/j.bbabio.2014.03.004
Guan, N., & Liu, L. (2020). Microbial response to acid stress: mechanisms and applications. *Appl Microbiol Biotechnol*, *104*(1), 51-65. https://doi.org/10.1007/s00253-019-10226-1
Guan, N., Liu, L., Shin, H. D., Chen, R. R., Zhang, J., Li, J., . . . Chen, J. (2013). Systems-level understanding of how *Propionibacterium acidipropionici* respond to propionic acid stress at the microenvironment levels: mechanism

and application. *J Biotechnol*, *167*(1), 56-63. https://doi.org/10.1016/j.jbiotec.2013.06.008

Guan, N., Shin, H. D., Chen, R. R., Li, J., Liu, L., Du, G., & Chen, J. (2014). Understanding of how *Propionibacterium acidipropionici* respond to propionic acid stress at the level of proteomics. *Sci Rep*, *4*, 6951. https://doi.org/10.1038/srep06951

Gyulev, I. S., Willson, B. J., Hennessy, R. C., Krabben, P., Jenkinson, E. R., & Thomas, G. H. (2018). Part by Part: Synthetic Biology Parts Used in Solventogenic Clostridia. *ACS Synth Biol*, *7*(2), 311-327. https://doi.org/10.1021/acssynbio.7b00327

Hanai, T., Atsumi, S., & Liao, J. C. (2007). Engineered synthetic pathway for isopropanol production in *Escherichia coli*. *Appl Environ Microbiol*, *73*(24), 7814-7818. https://doi.org/10.1128/AEM.01140-07

Handler, R. M., Shonnard, D. R., Griffing, E. M., Lai, A., & Palou-Rivera, I. (2016). Life Cycle Assessments of Ethanol Production via Gas Fermentation: Anticipated Greenhouse Gas Emissions for Cellulosic and Waste Gas Feedstocks. *Ind Eng Chem Res*, *55*(12), 3253-3261. https://doi.org/10.1021/acs.iecr.5b03215

Hassan, K. A., Cain, A. K., Huang, T., Liu, Q., Elbourne, L. D., Boinett, C. J., . . . Paulsen, I. T. (2016). Fluorescence-Based Flow Sorting in Parallel with Transposon Insertion Site Sequencing Identifies Multidrug Efflux Systems in *Acinetobacter baumannii*. *mBio*, *7*(5). https://doi.org/10.1128/mBio.01200-16

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

Heise, R., Muller, V., & Gottschalk, G. (1993). Acetogenesis and Atp Synthesis in *Acetobacterium woodii* Are Coupled Via a Transmembrane Primary Sodium-Ion Gradient. *Fems Microbiology Letters*, *112*(3), 261-268. https://doi.org/Doi 10.1016/0378-1097(93)90610-E

Hess, V., González, J. M., Parthasarathy, A., Buckel, W., & Müller, V. (2013). Caffeate respiration in the acetogenic bacterium *Acetobacterium woodii*: a coenzyme A loop saves energy for caffeate activation. *Appl Environ Microbiol*, *79*(6), 1942-1947. https://doi.org/10.1128/AEM.03604-12

Hess, V., Oyrik, O., Trifunovic, D., & Muller, V. (2015). 2,3-Butanediol Metabolism in the Acetogen *Acetobacterium woodii*. *Appl Environ Microbiol*, *81*(14), 4711-4719. https://doi.org/10.1128/AEM.00960-15

Hess, V., Schuchmann, K., & Muller, V. (2013). The ferredoxin:NAD+ oxidoreductase (Rnf) from the acetogen *Acetobacterium woodii* requires Na+ and is reversibly coupled to the membrane potential. *J Biol Chem*, *288*(44), 31496-31502. https://doi.org/10.1074/jbc.M113.510255

Hidalgo-Cantabrana, C., Goh, Y. J., Pan, M., Sanozky-Dawes, R., & Barrangou, R. (2019). Genome editing using the endogenous type I CRISPR-Cas system in *Lactobacillus crispatus*. *Proc Natl Acad Sci U S A*, *116*(32), 15774-15783. https://doi.org/10.1073/pnas.1905421116

Hoffmeister, S., Gerdom, M., Bengelsdorf, F. R., Linder, S., Fluchter, S., Ozturk, H., . . . Durre, P. (2016). Acetone production with metabolically engineered strains of *Acetobacterium woodii*. *Metab Eng*, *36*, 37-47. https://doi.org/10.1016/j.ymben.2016.03.001

Hong, W., Zhang, J., Cui, G., Wang, L., & Wang, Y. (2018). Multiplexed CRISPR-Cpf1-Mediated Genome Editing in *Clostridium difficile* toward the

Understanding of Pathogenesis of C. difficile Infection. *ACS Synth Biol*, *7*(6), 1588-1600. https://doi.org/10.1021/acssynbio.8b00087

Huang, H., Chai, C., Li, N., Rowe, P., Minton, N. P., Yang, S., . . . Gu, Y. (2016). CRISPR/Cas9-Based Efficient Genome Editing in *Clostridium ljungdahlii*, an Autotrophic Gas-Fermenting Bacterium. *ACS Synth Biol*, *5*(12), 1355-1361. https://doi.org/10.1021/acssynbio.6b00044

Höner zu Bentrup, K., Ubbink-Kok, T., Lolkema, J. S., & Konings, W. N. (1997). An Na+-pumping V1V0-ATPase complex in the thermophilic bacterium *Clostridium fervidus*. *J Bacteriol*, *179*(4), 1274-1279. https://doi.org/10.1128/jb.179.4.1274-1279.1997

Imkamp, F., & Müller, V. (2007). Acetogenic Bacteria. In *Encyclopedia of Life Sciences*. https://doi.org/https://doi.org/10.1002/9780470015902.a0020086

Ingle, P., Groothuis, D., Rowe, P., Huang, H., Cockayne, A., Kuehne, S. A., . . . Minton, N. P. (2019). Generation of a fully erythromycin-sensitive strain of *Clostridioides difficile* using a novel CRISPR-Cas9 genome editing system. *Sci Rep*, *9*(1), 8123. https://doi.org/10.1038/s41598-019-44458-y

Ismaiel, A. A., Zhu, C. X., Colby, G. D., & Chen, J. S. (1993). Purification and characterization of a primary-secondary alcohol dehydrogenase from two strains of *Clostridium beijerinckii*. *J Bacteriol*, *175*(16), 5097-5105. https://doi.org/10.1128/jb.175.16.5097-5105.1993

Iwasaki, Y., Kita, A., Yoshida, K., Tajima, T., Yano, S., Shou, T., . . . Nakashimada, Y. (2017). Homolactic Acid Fermentation by the Genetically Engineered Thermophilic Homoacetogen *Moorella thermoacetica* ATCC 39073. *Appl Environ Microbiol*, *83*(8). https://doi.org/10.1128/AEM.00247-17

Janscak, P., MacWilliams, M. P., Sandmeier, U., Nagaraja, V., & Bickle, T. A. (1999). DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. *EMBO J*, *18*(9), 2638-2647. https://doi.org/10.1093/emboj/18.9.2638

Javaid, N., & Choi, S. (2021). CRISPR/Cas System and Factors Affecting Its Precision and Efficiency. *Front Cell Dev Biol*, *9*, 761709. https://doi.org/10.3389/fcell.2021.761709

Jensen, T. O., Tellgren-Roth, C., Redl, S., Maury, J., Jacobsen, S. A. B., Pedersen, L. E., & Nielsen, A. T. (2019). Genome-wide systematic identification of methyltransferase recognition and modification patterns. *Nat Commun*, *10*(1), 3311. https://doi.org/10.1038/s41467-019-11179-9

Jeong, D. E., So, Y., Park, S. Y., Park, S. H., & Choi, S. K. (2018). Random knock-in expression system for high yield production of heterologous protein in **Bacillus subtilis**. *J Biotechnol*, *266*, 50-58. https://doi.org/10.1016/j.jbiotec.2017.12.007

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

Jones, D. L., Leroy, P., Unoson, C., Fange, D., Curic, V., Lawson, M. J., & Elf, J. (2017). Kinetics of dCas9 target search in *Escherichia coli*. *Science*, *357*(6358), 1420-1424. https://doi.org/10.1126/science.aah7084

Joseph, R. C., Kim, N. M., & Sandoval, N. R. (2018). Recent Developments of the Synthetic Biology Toolkit for *Clostridium*. *Front Microbiol*, *9*, 154. https://doi.org/10.3389/fmicb.2018.00154

Kaiser, J. C., Sen, S., Sinha, A., Wilkinson, B. J., & Heinrichs, D. E. (2016). The role of two branched-chain amino acid transporters in *Staphylococcus aureus*

growth, membrane fatty acid composition and virulence. *Mol Microbiol*, *102*(5), 850-864. https://doi.org/10.1111/mmi.13495
Kakinuma, Y., Yamato, I., & Murata, T. (1999). Structure and function of vacuolar Na+-translocating ATPase in *Enterococcus hirae*. *J Bioenerg Biomembr*, *31*(1), 7-14. https://doi.org/10.1023/a:1005499126939
Kantzow, C., Mayer, A., & Weuster-Botz, D. (2015). Continuous gas fermentation by *Acetobacterium woodii* in a submerged membrane reactor with full cell retention. *J Biotechnol*, *212*, 11-18. https://doi.org/10.1016/j.jbiotec.2015.07.020
Kantzow, C., & Weuster-Botz, D. (2016). Effects of hydrogen partial pressure on autotrophic growth and product formation of *Acetobacterium woodii*. *Bioprocess Biosyst Eng*, *39*(8), 1325-1330. https://doi.org/10.1007/s00449-016-1600-2
Karimi, K., Tabatabaei, M., Horvath, I. S., & Kumar, R. (2015). Recent trends in acetone, butanol, and ethanol (ABE) production. *Biofuel Res J*, *2*(4), 301-308. https://doi.org/Doi 10.18331/Brj2015.2.4.4
Katsyv, A., & Müller, V. (2020). Overcoming Energetic Barriers in Acetogenic C1 Conversion. *Front Bioeng Biotechnol*, *8*, 621166. https://doi.org/10.3389/fbioe.2020.621166
Kelleher, J. E., Daniel, A. S., & Murray, N. E. (1991). Mutations the confer *de Novo* activity upon a maintenance methyltransferase. *J Mol Bio*, *221*(2), 431-440.
Kopke, M., Gerth, M. L., Maddock, D. J., Mueller, A. P., Liew, F., Simpson, S. D., & Patrick, W. M. (2014). Reconstruction of an acetogenic 2,3-butanediol pathway involving a novel NADPH-dependent primary-secondary alcohol dehydrogenase. *Appl Environ Microbiol*, *80*(11), 3394-3403. https://doi.org/10.1128/AEM.00301-14
Kopke, M., Held, C., Hujer, S., Liesegang, H., Wiezer, A., Wollherr, A., . . . Durre, P. (2010). *Clostridium ljungdahlii* represents a microbial production platform based on syngas. *Proc Natl Acad Sci U S A*, *107*(29), 13087-13092. https://doi.org/10.1073/pnas.1004716107
Krauss, U., Minh, B. Q., Losi, A., Gartner, W., Eggert, T., von Haeseler, A., & Jaeger, K. E. (2009). Distribution and phylogeny of light-oxygen-voltage-blue-light-signaling proteins in the three kingdoms of life. *J Bacteriol*, *191*(23), 7234-7242. https://doi.org/10.1128/JB.00923-09
Kremp, F., Poehlein, A., Daniel, R., & Muller, V. (2018). Methanol metabolism in the acetogenic bacterium *Acetobacterium woodii*. *Environ Microbiol*, *20*(12), 4369-4384. https://doi.org/10.1111/1462-2920.14356
Kremp, F., Roth, J., & Müller, V. (2020). The Sporomusa type Nfn is a novel type of electron-bifurcating transhydrogenase that links the redox pools in acetogenic bacteria. *Sci Rep*, *10*(1), 14872. https://doi.org/10.1038/s41598-020-71038-2
Kremp, F., Roth, J., & Müller, V. (2022). A Third Way of Energy Conservation in Acetogenic Bacteria. *Microbiol Spectr*, *10*(4), e0138522. https://doi.org/10.1128/spectrum.01385-22
Kuehne, S. A., & Minton, N. P. (2012). ClosTron-mediated engineering of *Clostridium*. *Bioengineered*, *3*(4), 247-254. https://doi.org/10.4161/bioe.21004
Köpke, M., & Simpson, S. D. (2020). Pollution to products: recycling of 'above ground' carbon by gas fermentation. *Curr Opin Biotechnol*, *65*, 180-189. https://doi.org/10.1016/j.copbio.2020.02.017

Lampe, D. J., Grant, T. E., & Robertson, H. M. (1998). Factors affecting transposition of the Himar1 mariner transposon in vitro. *Genetics*, *149*(1), 179-187. https://doi.org/10.1093/genetics/149.1.179

Landete, J. M., Langa, S., Revilla, C., Margolles, A., Medina, M., & Arques, J. L. (2015). Use of anaerobic green fluorescent protein versus green fluorescent protein as reporter in lactic acid bacteria. *Appl Microbiol Biotechnol*, *99*(16), 6865-6877. https://doi.org/10.1007/s00253-015-6770-3

Landete, J. M., Peiroten, A., Medina, M., & Arques, J. L. (2017). Short communication: Labeling Listeria with anaerobic fluorescent protein for food safety studies. *J Dairy Sci*, *100*(1), 113-117. https://doi.org/10.3168/jds.2016-11226

Langridge, G. C., Phan, M. D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., . . . Turner, A. K. (2009). Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. *Genome Res*, *19*(12), 2308-2316. https://doi.org/10.1101/gr.097097.109

Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., & Qi, L. S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc*, *8*(11), 2180-2196. https://doi.org/10.1038/nprot.2013.132

Leang, C., Ueki, T., Nevin, K. P., & Lovley, D. R. (2013). A genetic system for *Clostridium ljungdahlii*: a chassis for autotrophic production of biocommodities and a model homoacetogen. *Appl Environ Microbiol*, *79*(4), 1102-1109. https://doi.org/10.1128/AEM.02891-12

Lechtenfeld, M., Heine, J., Sameith, J., Kremp, F., & Muller, V. (2018). Glycine betaine metabolism in the acetogenic bacterium *Acetobacterium woodii*. *Environ Microbiol*, *20*(12), 4512-4525. https://doi.org/10.1111/1462-2920.14389

Leenay, R. T., Maksimchuk, K. R., Slotkowski, R. A., Agrawal, R. N., Gomaa, A. A., Briner, A. E., . . . Beisel, C. L. (2016). Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Mol Cell*, *62*(1), 137-147. https://doi.org/10.1016/j.molcel.2016.02.031

Li, Q., Chen, J., Minton, N. P., Zhang, Y., Wen, Z., Liu, J., . . . Yang, S. (2016). CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *Biotechnol J*, *11*(7), 961-972. https://doi.org/10.1002/biot.201600053

Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., . . . She, Q. (2016). Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic Acids Res*, *44*(4), e34. https://doi.org/10.1093/nar/gkv1044

Li, Y., Xu, T., Tschaplinski, T. J., Engle, N. L., Yang, Y., Graham, D. E., . . . Zhou, J. (2014). Improvement of cellulose catabolism in *Clostridium cellulolyticum* by sporulation abolishment and carbon alleviation. *Biotechnol Biofuels*, *7*(1), 25. https://doi.org/10.1186/1754-6834-7-25

Liew, F., Kopke, M., & Simpson, S. D. (2013). Gas Fermentation for Commercial Biofuels Production.

Liew, F., Martin, M. E., Tappel, R. C., Heijstra, B. D., Mihalcea, C., & Kopke, M. (2016). Gas Fermentation-A Flexible Platform for Commercial Scale Production of Low-Carbon-Fuels and Chemicals from Waste and Renewable Feedstocks. *Front Microbiol*, *7*, 694. https://doi.org/10.3389/fmicb.2016.00694

Liew, F., Nogle, R., Abdalla, T., Rasor, B. J., Canter, C., & Jensen, R. O. (2022). Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. *Nat Biotechnol*.

Little, G. T., Willson, B. J., Heap, J. T., Winzer, K., & Minton, N. P. (2018). The Butanol Producing Microbe *Clostridium beijerinckii* NCIMB 14988 Manipulated Using Forward and Reverse Genetic Tools. *Biotechnol J*, *13*(11), e1700711. https://doi.org/10.1002/biot.201700711

Liu, H., Bouillaut, L., Sonenshein, A. L., & Melville, S. B. (2013). Use of a mariner-based transposon mutagenesis system to isolate *Clostridium perfringens* mutants deficient in gliding motility. *J Bacteriol*, *195*(3), 629-636. https://doi.org/10.1128/JB.01288-12

Ljungdahl, L. G. (1986). The autotrophic pathway of acetate synthesis in acetogenic bacteria. *Annu Rev Microbiol*, *40*, 415-450. https://doi.org/10.1146/annurev.mi.40.100186.002215

Loenen, W. A., & Murray, N. E. (1986). Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. *J Mol Biol*, *190*(1), 11-22. https://doi.org/10.1016/0022-2836(86)90071-9

Low, D. A., & Casadesus, J. (2008). Clocks and switches: bacterial gene regulation by DNA adenine methylation. *Curr Opin Microbiol*, *11*(2), 106-112. https://doi.org/10.1016/j.mib.2008.02.012

Maikova, A., Kreis, V., Boutserin, A., Severinov, K., & Soutourina, O. (2019). Using an Endogenous CRISPR-Cas System for Genome Editing in the Human Pathogen *Clostridium difficile*. *Appl Environ Microbiol*, *85*(20). https://doi.org/10.1128/AEM.01416-19

Makarova, K. S., & Koonin, E. V. (2015). Annotation and Classification of CRISPR-Cas Systems. *Methods Mol Biol*, *1311*, 47-75. https://doi.org/10.1007/978-1-4939-2687-9_4

Marcellin, E., Behrendorff, J. B., Nagaraju, S., DeTissera, S., Segovia, S., Palfreyman, R. W., . . . Nielsen, L. K. (2016). Low carbon fuels and commodity chemicals from waste gases – systematic approach to understand energy metabolism in a model acetogen [10.1039/C5GC02708J]. *Green Chem*, *18*(10), 3020-3028. https://doi.org/10.1039/C5GC02708J

Matelska, D., Steczkiewicz, K., & Ginalski, K. (2017). Comprehensive classification of the PIN domain-like superfamily. *Nucleic Acids Res*, *45*(12), 6995-7020. https://doi.org/10.1093/nar/gkx494

Mayer, C., & Boos, W. (2005). Hexose/Pentose and Hexitol/Pentitol Metabolism. *EcoSal Plus*, *1*(2). https://doi.org/10.1128/ecosalplus.3.4.1

Mesfin, N. A. (2020). *Structural Metabolic Modelling of the Acetogen Acetobacterium woodii* Oxford Brookes University].

Mignon, C., Mariano, N., Stadthagen, G., Lugari, A., Lagoutte, P., Donnat, S., . . . Werle, B. (2018). Codon harmonization - going beyond the speed limit for protein expression. *FEBS Lett*, *592*(9), 1554-1564. https://doi.org/10.1002/1873-3468.13046

Minton, N. P., Ehsaan, M., Humphreys, C. M., Little, G. T., Baker, J., Henstra, A. M., . . . Zhang, Y. (2016). A roadmap for gene system development in Clostridium. *Anaerobe*, *41*, 104-112. https://doi.org/10.1016/j.anaerobe.2016.05.011

Mock, J., Zheng, Y., Mueller, A. P., Ly, S., Tran, L., Segovia, S., . . . Thauer, R. K. (2015). Energy Conservation Associated with Ethanol Formation from H2 and CO2 in *Clostridium autoethanogenum* Involving Electron Bifurcation. *J Bacteriol*, *197*(18), 2965-2980. https://doi.org/10.1128/JB.00399-15

Molitor, B., Kirchner, K., Henrich, A. W., Schmitz, S., & Rosenbaum, M. A. (2016). Expanding the molecular toolkit for the homoacetogen *Clostridium ljungdahlii*. *Sci Rep*, *6*, 31518. https://doi.org/10.1038/srep31518

Moller, B., Oßmer, R., Howard, B. H., Gottschalk, G., & Hippe, H. (1984). *Sporomusa*, a new genus of gram-negative anaerobic bacteria including S*poromusa sphaeroides* spec. nov. and Sporomusa ovata spec. nov.. *Arch. Microbiol.*, *139*, 388-396.

Monot, F., Martin, J. R., Petitdemange, H., & Gay, R. (1982). Acetone and Butanol Production by *Clostridium acetobutylicum* in a Synthetic Medium. *Appl Environ Microbiol*, *44*(6), 1318-1324. https://doi.org/10.1128/aem.44.6.1318-1324.1982

Montenegro-Rodriguez, C., Peiroten, A., Sanchez-Jimenez, A., Arques, J. L., & Landete, J. M. (2015). Analysis of gene expression of bifidobacteria using as the reporter an anaerobic fluorescent protein. *Biotechnol Lett*, *37*(7), 1405-1413. https://doi.org/10.1007/s10529-015-1802-8

Mook, A., Beck, M. H., Baker, J. P., Minton, N. P., Durre, P., & Bengelsdorf, F. R. (2022). Autotrophic lactate production from H2 + CO2 using recombinant and fluorescent FAST-tagged *Acetobacterium woodii* strains. *Appl Microbiol Biotechnol*, *106*(4), 1447-1458. https://doi.org/10.1007/s00253-022-11770-z

Moon, H. G., Jang, Y. S., Cho, C., Lee, J., Binkley, R., & Lee, S. Y. (2016). One hundred years of clostridial butanol fermentation. *FEMS Microbiol Lett*, *363*(3). https://doi.org/10.1093/femsle/fnw001

Moon, J., Dönig, J., Kramer, S., Poehlein, A., Daniel, R., & Müller, V. (2021). Formate metabolism in the acetogenic bacterium *Acetobacterium woodii*. *Environ Microbiol*, *23*(8), 4214-4227. https://doi.org/10.1111/1462-2920.15598

Moon, J., & Müller, V. (2021). Physiology and genetics of ethanologenesis in the acetogenic bacterium *Acetobacterium woodii*. *Environ Microbiol*, *23*(11), 6953-6964. https://doi.org/10.1111/1462-2920.15739

Moon, J., Schubert, A., Waschinger, L. M., & Müller, V. (2023). Reprogramming the metabolism of an acetogenic bacterium to homoformatogenesis. *ISME J*, *17*(7), 984-992. https://doi.org/10.1038/s41396-023-01411-2

Moon, S. B., Kim, D. Y., Ko, J. H., & Kim, Y. S. (2019). Recent advances in the CRISPR genome editing tool set. *Exp Mol Med*, *51*(11), 1-11. https://doi.org/10.1038/s12276-019-0339-7

Mordaka, P. M., & Heap, J. T. (2018). Stringency of Synthetic Promoter Sequences in *Clostridium* Revealed and Circumvented by Tuning Promoter Library Mutation Rates. *ACS Synth Biol*, *7*(2), 672-681. https://doi.org/10.1021/acssynbio.7b00398

Mruk, I., & Kobayashi, I. (2014). To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Res*, *42*(1), 70-86. https://doi.org/10.1093/nar/gkt711

Mukherjee, A., Weyant, K. B., Agrawal, U., Walker, J., Cann, I. K., & Schroeder, C. M. (2015). Engineering and characterization of new LOV-based fluorescent proteins from *Chlamydomonas reinhardtii* and Vaucheria frigida. *ACS Synth Biol*, *4*(4), 371-377. https://doi.org/10.1021/sb500237x

Mukherjee, A., Weyant, K. B., Walker, J., & Schroeder, C. M. (2012). Directed evolution of bright mutants of an oxygen-independent flavin-binding fluorescent protein from *Pseudomonas putida*. *J Biol Eng*, *6*(1), 20. https://doi.org/10.1186/1754-1611-6-20

Mulepati, S., & Bailey, S. (2013). In vitro reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J Biol Chem*, *288*(31), 22184-22192. https://doi.org/10.1074/jbc.M113.472233

Muller, E., Fahlbusch, K., Walther, R., & Gottschalk, G. (1981). Formation of N,N-Dimethylglycine, Acetic Acid, and Butyric Acid from Betaine by *Eubacterium limosum*. *Appl Environ Microbiol*, *42*(3), 439-445. https://doi.org/10.1128/aem.42.3.439-445.1981

Murphy, J., Mahony, J., Ainsworth, S., Nauta, A., & van Sinderen, D. (2013). Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. *Appl Environ Microbiol*, *79*(24), 7547-7555. https://doi.org/10.1128/AEM.02229-13

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

Müller, V., Aufurth, S., & Rahlfs, S. (2001). The Na(+) cycle in *Acetobacterium woodii*: identification and characterization of a Na(+) translocating F(1)F(0)-ATPase with a mixed oligomer of 8 and 16 kDa proteolipids. *Biochim Biophys Acta*, *1505*(1), 108-120. https://doi.org/10.1016/s0005-2728(00)00281-4

Müller, V., Imkamp, F., Biegel, E., Schmidt, S., & Dilling, S. (2008). Discovery of a ferredoxin:NAD+-oxidoreductase (Rnf) in *Acetobacterium woodii*: a novel potential coupling site in acetogens. *Ann N Y Acad Sci*, *1125*, 137-146. https://doi.org/10.1196/annals.1419.011

Nagaraju, S., Davies, N. K., Walker, D. J., Kopke, M., & Simpson, S. D. (2016). Genome editing of *Clostridium autoethanogenum* using CRISPR/Cas9. *Biotechnol Biofuels*, *9*, 219. https://doi.org/10.1186/s13068-016-0638-3

Negahdaripour, M., Nezafat, N., Hajighahramani, N., Rahmatabadi, S. S., & Ghasemi, Y. (2017). Investigating CRISPR-Cas systems in *Clostridium botulinum* via bioinformatics tools. *Infect Genet Evol*, *54*, 355-373. https://doi.org/10.1016/j.meegid.2017.06.027

Neuendorf, C. S., Vignolle, G. A., Derntl, C., Tomin, T., Novak, K., Mach, R. L., . . . Pflügl, S. (2021). A quantitative metabolic analysis reveals *Acetobacterium woodii* as a flexible and robust host for formate-based bioproduction. *Metab Eng*, *68*, 68-85. https://doi.org/10.1016/j.ymben.2021.09.004

Ng, Y. K., Ehsaan, M., Philip, S., Collery, M. M., Janoir, C., Collignon, A., . . . Minton, N. P. (2013). Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using pyrE alleles. *PLoS One*, *8*(2), e56051. https://doi.org/10.1371/journal.pone.0056051

Nishimasu, H., & Nureki, O. (2017). Structures and mechanisms of CRISPR RNA-guided effector nucleases. *Curr Opin Struct Biol*, *43*, 68-78. https://doi.org/10.1016/j.sbi.2016.11.013

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., . . . Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, *156*(5), 935-949. https://doi.org/10.1016/j.cell.2014.02.001

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*, *302*(1), 205-217. https://doi.org/10.1006/jmbi.2000.4042

Olson, D. G., McBride, J. E., Shaw, A. J., & Lynd, L. R. (2012). Recent progress in consolidated bioprocessing. *Curr Opin Biotechnol*, *23*(3), 396-405. https://doi.org/10.1016/j.copbio.2011.11.026

Peng, Y., Luo, Y., Yu, T., Xu, X., Fan, K., Zhao, Y., & Yang, K. (2011). A blue native-PAGE analysis of membrane protein complexes in *Clostridium thermocellum*. *BMC Microbiol*, *11*(1), 22. https://doi.org/10.1186/1471-2180-11-22

Peters, V., Janssen, P. H., & Conrad, R. (1999). Transient Production of Formate During Chemolithotrophic Growth of Anaerobic Microorganisms on Hydrogen. *Curr Microbiol*, *38*, 285-289.

Pezacka, E., & Wood, H. G. (1984). The synthesis of acetyl-CoA by *Clostridium thermoaceticum* from carbon dioxide, hydrogen, coenzyme A and methyltetrahydrofolate. *Arch Microbiol*, *137*(1), 63-69. https://doi.org/10.1007/BF00425809

Philipps, G., de Vries, S., & Jennewein, S. (2019). Development of a metabolic pathway transfer and genomic integration system for the syngas-fermenting bacterium *Clostridium ljungdahlii*. *Biotechnol Biofuels*, *12*, 112. https://doi.org/10.1186/s13068-019-1448-1

Pinhal, S., Ropers, D., Geiselmann, J., & de Jong, H. (2019). Acetate Metabolism and the Inhibition of Bacterial Growth by Acetate. *J Bacteriol*, *201*(13). https://doi.org/10.1128/JB.00147-19

Plamont, M. A., Billon-Denis, E., Maurin, S., Gauron, C., Pimenta, F. M., Specht, C. G., . . . Gautier, A. (2016). Small fluorescence-activating and absorption-shifting tag for tunable protein imaging in vivo. *Proc Natl Acad Sci U S A*, *113*(3), 497-502. https://doi.org/10.1073/pnas.1513094113

Poehlein, A., Cebulla, M., Ilg, M. M., Bengelsdorf, F. R., Schiel-Bengelsdorf, B., Whited, G., . . . Dürre, P. (2015). The Complete Genome Sequence of *Clostridium aceticum*: a Missing Link between Rnf- and Cytochrome-Containing Autotrophic Acetogens. *mBio*, *6*(5), e01168-01115. https://doi.org/10.1128/mBio.01168-15

Poehlein, A., Schmidt, S., Kaster, A. K., Goenrich, M., Vollmers, J., Thürmer, A., . . . Müller, V. (2012). An ancient pathway combining carbon dioxide fixation with the generation and utilization of a sodium ion gradient for ATP synthesis. *PLoS One*, *7*(3), e33439. https://doi.org/10.1371/journal.pone.0033439

Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A., & Chou, C. P. (2016). Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Sci Rep*, *6*, 25666. https://doi.org/10.1038/srep25666

Ragsdale, S. W., & Ljungdahl, L. G. (1984). Purification and properties of NAD-dependent 5,10-methylenetetrahydrofolate dehydrogenase from *Acetobacterium woodii*. *J Biol Chem*, *259*(6), 3499-3503.

Ragsdale, S. W., & Pierce, E. (2008). Acetogenesis and the Wood-Ljungdahl pathway of CO2 fixation. *Biochim Biophys Acta*, *1784*(12), 1873-1898. https://doi.org/10.1016/j.bbapap.2008.08.012

Razin, A., & Riggs, A. D. (1980). DNA methylation and gene function. *Science*, *210*(4470), 604-610. https://doi.org/10.1126/science.6254144

Redfern, E. R. M. (2021). *Developing Synthetic Biology Tools for Acetogens* University of Nottingham].

Reidlinger, J., & Müller, V. (1994). Purification of ATP synthase from *Acetobacterium woodii* and identification as a Na(+)-translocating F1F0-type enzyme. *Eur J Biochem*, *223*(1), 275-283. https://doi.org/10.1111/j.1432-1033.1994.tb18992.x

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., . . . Xu, S. Y. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res*, *31*(7), 1805-1812. https://doi.org/10.1093/nar/gkg274

Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2010). REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res*, *38*(Database issue), D234-236. https://doi.org/10.1093/nar/gkp874

Roe, A. J., O'Byrne, C., McLaggan, D., & Booth, I. R. (2002). Inhibition of Escherichia coli growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity. *Microbiology (Reading)*, *148*(Pt 7), 2215-2222. https://doi.org/10.1099/00221287-148-7-2215

Roer, L., Aarestrup, F. M., & Hasman, H. (2015). The EcoKI type I restriction-modification system in *Escherichia coli* affects but is not an absolute barrier for conjugation. *J Bacteriol*, *197*(2), 337-342. https://doi.org/10.1128/JB.02418-14

Ronzon, T., Piotrowski, S., Tamosiunas, S., Dammer, L., Carus, M., & M'barek, R. (2020). Developments of Economic Growth and Employment in Bioeconomy Sectors across the EU. *Sustainability*, *12*(11).

Ross, D. E., Marshall, C. W., Gulliver, D., May, H. D., & Norman, R. S. (2020). Defining Genomic and Predicted Metabolic Features of the. *mSystems*, *5*(5). https://doi.org/10.1128/mSystems.00277-20

Rousseau, C., Gonnet, M., Le Romancer, M., & Nicolas, J. (2009). CRISPI: a CRISPR interactive database. *Bioinformatics*, *25*(24), 3317-3318. https://doi.org/10.1093/bioinformatics/btp586

Rybnicky, G. A., Fackler, N. A., Karim, A. S., Kopke, M., & Jewett, M. C. (2022). Spacer2PAM: A computational framework to guide experimental determination of functional CRISPR-Cas system PAM sequences. *Nucleic Acids Res*, *50*(6), 3523-3534. https://doi.org/10.1093/nar/gkac142

Saleski, T. E., Chung, M. T., Carruthers, D. N., Khasbaatar, A., Kurabayashi, K., & Lin, X. N. (2021). Optimized gene expression from bacterial chromosome by high-throughput integration and screening. *Sci Adv*, *7*(7). https://doi.org/10.1126/sciadv.abe1767

Sampson, E. M., & Bobik, T. A. (2008). Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J Bacteriol*, *190*(8), 2966-2971. https://doi.org/10.1128/JB.01925-07

Sauer, U., & Eikmanns, B. J. (2005). The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol Rev*, *29*(4), 765-794. https://doi.org/10.1016/j.femsre.2004.11.002

Schoelmerich, M. C., Katsyv, A., Sung, W., Mijic, V., Wiechmann, A., Kottenhahn, P., . . . Müller, V. (2018). Regulation of lactate metabolism in the acetogenic bacterium *Acetobacterium woodii*. *Environ Microbiol*, *20*(12), 4587-4595. https://doi.org/10.1111/1462-2920.14412

Schoelmerich, M. C., & Müller, V. (2019). Energy conservation by a hydrogenase-dependent chemiosmotic mechanism in an ancient metabolic pathway. *Proc Natl Acad Sci U S A*, *116*(13), 6329-6334. https://doi.org/10.1073/pnas.1818580116

Scholz, S. A., Diao, R., Wolfe, M. B., Fivenson, E. M., Lin, X. N., & Freddolino, P. L. (2019). High-Resolution Mapping of the Escherichia coli Chromosome

Reveals Positions of High and Low Transcription. *Cell Syst*, *8*(3), 212-225 e219. https://doi.org/10.1016/j.cels.2019.02.004
Schuchmann, K., & Muller, V. (2016). Energetics and Application of Heterotrophy in Acetogenic Bacteria. *Appl Environ Microbiol*, *82*(14), 4056-4069. https://doi.org/10.1128/AEM.00882-16
Schuchmann, K., & Müller, V. (2012). A bacterial electron-bifurcating hydrogenase. *J Biol Chem*, *287*(37), 31165-31171. https://doi.org/10.1074/jbc.M112.395038
Schuchmann, K., & Müller, V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat Rev Microbiol*, *12*(12), 809-821. https://doi.org/10.1038/nrmicro3365
Schuchmann, K., Schmidt, S., Martinez Lopez, A., Kaberline, C., Kuhns, M., Lorenzen, W., . . . Müller, V. (2015). Nonacetogenic growth of the acetogen *Acetobacterium woodii* on 1,2-propanediol. *J Bacteriol*, *197*(2), 382-391. https://doi.org/10.1128/JB.02383-14
Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., . . . Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A*, *108*(25), 10098-10103. https://doi.org/10.1073/pnas.1104144108
Sen, S., Sirobhushanam, S., Hantak, M. P., Lawrence, P., Brenna, J. T., Gatto, C., & Wilkinson, B. J. (2015). Short branched-chain C6 carboxylic acids result in increased growth, novel 'unnatural' fatty acids and increased membrane fluidity in a *Listeria monocytogenes* branched-chain fatty acid-deficient mutant. *Biochim Biophys Acta*, *1851*(10), 1406-1415. https://doi.org/10.1016/j.bbalip.2015.07.006
Serganov, A., & Nudler, E. (2013). A decade of riboswitches. *Cell*, *152*(1-2), 17-24. https://doi.org/10.1016/j.cell.2012.12.024
Shin, J., Song, Y., Jeong, Y., & Cho, B. K. (2016). Analysis of the Core Genome and Pan-Genome of Autotrophic Acetogenic Bacteria. *Front Microbiol*, *7*, 1531. https://doi.org/10.3389/fmicb.2016.01531
Shin, J., Song, Y., Kang, S., Jin, S., Lee, J. K., Kim, D. R., . . . Cho, B. K. (2021). Genome-Scale Analysis of Acetobacterium woodii Identifies Translational Regulation of Acetogenesis. *mSystems*, *6*(4), e0069621. https://doi.org/10.1128/mSystems.00696-21
Smith, P., & Schuster, M. (2019). Public goods and cheating in microbes. *Curr Biol*, *29*(11), R442-R447. https://doi.org/10.1016/j.cub.2019.03.001
Solovyev, V., & Salamov, A. (2010). Automatic annotation of microbial genomes and metagnomic sequences. In p. Robert W. Li (Ed.), *Metagenomics and Its Applications in Agriculture, Biomedicine and Environmental Studies*. Nova Science Publishers, Inc.
Song, Y., Lee, J. S., Shin, J., Lee, G. M., Jin, S., Kang, S., . . . Cho, B. K. (2020). Functional cooperation of the glycine synthase-reductase and Wood-Ljungdahl pathways for autotrophic growth of. *Proc Natl Acad Sci U S A*, *117*(13), 7516-7523. https://doi.org/10.1073/pnas.1912289117
Song, Y., Shin, J., Jin, S., Lee, J. K., Kim, D. R., Kim, S. C., . . . Cho, B. K. (2018). Genome-scale analysis of syngas fermenting acetogenic bacteria reveals the translational regulation for its autotrophic growth. *BMC Genomics*, *19*(1), 837. https://doi.org/10.1186/s12864-018-5238-0
Spruth, M., Reidlinger, J., & Müller, V. (1995). Sodium ion dependence of inhibition of the Na+-translocating F1F0-ATPase from *Acetobacterium woodii*. Probing

the site(s) involved in ion transport. *Biochem Biophys Acta Bioenerg*, *1229*(1), 96-102. https://doi.org/https://doi.org/10.1016/0005-2728(94)00192-8

Steger, F., Rachbauer, L., Windhagauer, M., Montgomery, L. F. R., & Bochmann, G. (2017). Optimisation of continuous gas fermentation by immobilisation of acetate-producing *Acetobacterium woodii*. *Anaerobe*, *46*, 96-103. https://doi.org/10.1016/j.anaerobe.2017.06.010

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, *507*(7490), 62-67. https://doi.org/10.1038/nature13011

Stratz, M., Sauer, U., Kuhn, A., & Durre, P. (1994). Plasmid transfer into the homoacetogen *Acetobacterium woodii* by electroporation and conjugation. *Appl Environ Microbiol*, *60*(3), 1033-1037.

Straub, M., Demler, M., Weuster-Botz, D., & Durre, P. (2014). Selective enhancement of autotrophic acetate production with genetically modified *Acetobacterium woodii*. *J Biotechnol*, *178*, 67-72. https://doi.org/10.1016/j.jbiotec.2014.03.005

Streett, H. E., Kalis, K. M., & Papoutsakis, E. T. (2019). A strongly fluorescing anaerobic reporter and protein-tagging system for Clostridium organisms based on the Fluorescence-Activating and Absorption-Shifting Tag (FAST) protein. *Appl Environ Microbiol*. https://doi.org/10.1128/AEM.00622-19

Strätz, M., Gottschalk, G., & Dürre, P. (1990). Transfer and expression of the tetracycline resistance transposon Tn925 in *Acetobacterium woodii*. *FEMS Microbiol Lett*, *56*(1-2), 171-176. https://doi.org/10.1111/j.1574-6968.1990.tb04144.x

Studier, F. W., & Bandyopadhyay, P. K. (1988). Model for how type I restriction enzymes select cleavage sites in DNA. *Proc. Natl. Acad. Sci.*, *85*, 4677-4681.

Sun, Y., Wilkinson, B. J., Standiford, T. J., Akinbi, H. T., & O'Riordan, M. X. (2012). Fatty acids regulate stress resistance and virulence factor production for Listeria monocytogenes. *J Bacteriol*, *194*(19), 5274-5284. https://doi.org/10.1128/JB.00045-12

Szima, S., & Cormos, C.-C. (2018). Improving methanol synthesis from carbon-free H2 and captured CO2: A techno-economic and environmental evaluation. *J CO2 Util*, *24*, 555-563. https://doi.org/https://doi.org/10.1016/j.jcou.2018.02.007

Takors, R., Kopf, M., Mampel, J., Bluemke, W., Blombach, B., Eikmanns, B., . . . Dürre, P. (2018). Using gas mixtures of CO, CO. *Microb Biotechnol*, *11*(4), 606-625. https://doi.org/10.1111/1751-7915.13270

Tebo, A. G., Pimenta, F. M., Zhang, Y., & Gautier, A. (2018). Improved Chemical-Genetic Fluorescent Markers for Live Cell Microscopy. *Biochem*, *57*(39), 5648-5653. https://doi.org/10.1021/acs.biochem.8b00649

Thompson, A. W., Crow, M. J., Wadey, B., Arens, C., Turkarslan, S., Stolyar, S., . . . Baliga, N. S. (2015). A method to analyze, sort, and retain viability of obligate anaerobic microorganisms from complex microbial communities. *J Microbiol Methods*, *117*, 74-77. https://doi.org/10.1016/j.mimet.2015.07.009

Topp, S., Reynoso, C. M., Seeliger, J. C., Goldlust, I. S., Desai, S. K., Murat, D., . . . Gallivan, J. P. (2010). Synthetic riboswitches that induce gene expression in diverse bacterial species. *Appl Environ Microbiol*, *76*(23), 7881-7884. https://doi.org/10.1128/AEM.01537-10

Tummala, S. B., Welker, N. E., & Papoutsakis, E. T. (1999). Development and characterization of a gene expression reporter system for *Clostridium acetobutylicum* ATCC 824. *Appl Environ Microbiol*, *65*(9), 3793-3799.

UNFCC. (2015). *The Paris agreement*.
Valgepea, K., de Souza Pinto Lemgruber, R., Meaghan, K., Palfreyman, R. W., Abdalla, T., Heijstra, B. D., . . . Marcellin, E. (2017). Maintenance of ATP Homeostasis Triggers Metabolic Shifts in Gas-Fermenting Acetogens. *Cell Syst*, *4*(5), 505-515.e505. https://doi.org/10.1016/j.cels.2017.04.008
Van Melderen, L., & Saavedra De Bast, M. (2009). Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS Genet*, *5*(3), e1000437. https://doi.org/10.1371/journal.pgen.1000437
van Opijnen, T., Bodi, K. L., & Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods*, *6*(10), 767-772. https://doi.org/10.1038/nmeth.1377
van Opijnen, T., & Camilli, A. (2012). A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res*, *22*(12), 2541-2551. https://doi.org/10.1101/gr.137430.112
van Opijnen, T., & Camilli, A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol*, *11*(7), 435-442. https://doi.org/10.1038/nrmicro3033
Vasu, K., & Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev*, *77*(1), 53-72. https://doi.org/10.1128/MMBR.00044-12
Vink, J. N. A., Baijens, J. H. L., & Brouns, S. J. J. (2021). PAM-repeat associations and spacer selection preferences in single and co-occurring CRISPR-Cas systems. *Genome Biol*, *22*(1), 281. https://doi.org/10.1186/s13059-021-02495-9
Vrancken, G., Rimaux, T., Weckx, S., De Vuyst, L., & Leroy, F. (2009). Environmental pH determines citrulline and ornithine release through the arginine deiminase pathway in *Lactobacillus fermentum* IMDO 130101. *Int J Food Microbiol*, *135*(3), 216-222. https://doi.org/10.1016/j.ijfoodmicro.2009.07.035
Vrancken, G., Rimaux, T., Wouters, D., Leroy, F., & De Vuyst, L. (2009). The arginine deiminase pathway of *Lactobacillus fermentum* IMDO 130101 responds to growth under stress conditions of both temperature and salt. *Food Microbiol*, *26*(7), 720-727. https://doi.org/10.1016/j.fm.2009.07.006
Wakabayashi, Y., Nariya, H., Yasugi, M., Kuwahara, T., Sarker, M. R., & Miyake, M. (2019). An enhanced green fluorescence protein (EGFP)-based reporter assay for quantitative detection of sporulation in *Clostridium perfringens* SM101. *Int J Food Microbiol*, *291*, 144-150. https://doi.org/10.1016/j.ijfoodmicro.2018.11.015
Walker, J. E., Lanahan, A. A., Zheng, T., Toruno, C., Lynd, L. R., Cameron, J. C., . . . Eckert, C. A. (2020). Development of both type I-B and type II CRISPR/Cas genome editing systems in the cellulolytic bacterium *Clostridium thermocellum*. *Metab Eng Commun*, *10*, e00116. https://doi.org/10.1016/j.mec.2019.e00116
Wang, S., Huang, H., Kahnt, J., Mueller, A. P., Köpke, M., & Thauer, R. K. (2013). NADP-specific electron-bifurcating [FeFe]-hydrogenase in a functional complex with formate dehydrogenase in *Clostridium autoethanogenum* grown on CO. *J Bacteriol*, *195*(19), 4373-4386. https://doi.org/10.1128/JB.00678-13
Wang, S., Huang, H., Moll, J., & Thauer, R. K. (2010). NADP+ reduction with reduced ferredoxin and NADP+ reduction with NADH are coupled via an

electron-bifurcating enzyme complex in *Clostridium kluyveri*. *J Bacteriol*, *192*(19), 5115-5123. https://doi.org/10.1128/JB.00612-10
Wang, Y., Zhang, Z. T., Seo, S. O., Choi, K., Lu, T., Jin, Y. S., & Blaschek, H. P. (2015). Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *J Biotechnol*, *200*, 1-5. https://doi.org/10.1016/j.jbiotec.2015.02.005
Wang, Y., Zhang, Z. T., Seo, S. O., Lynn, P., Lu, T., Jin, Y. S., & Blaschek, H. P. (2016). Bacterial Genome Editing with CRISPR-Cas9: Deletion, Integration, Single Nucleotide Modification, and Desirable "Clean" Mutant Selection in *Clostridium beijerinckii* as an Example. *ACS Synth Biol*, *5*(7), 721-732. https://doi.org/10.1021/acssynbio.6b00060
Weghoff, M. C., Bertsch, J., & Müller, V. (2015). A novel mode of lactate metabolism in strictly anaerobic bacteria. *Environ Microbiol*, *17*(3), 670-677. https://doi.org/10.1111/1462-2920.12493
Westphal, L., Wiechmann, A., Baker, J., Minton, N. P., & Müller, V. (2018). The Rnf Complex Is an Energy-Coupled Transhydrogenase Essential To Reversibly Link Cellular NADH and Ferredoxin Pools in the Acetogen *Acetobacterium woodii*. *J Bacteriol*, *200*(21). https://doi.org/10.1128/JB.00357-18
Wiechmann, A., Ciurus, S., Oswald, F., Seiler, V. N., & Müller, V. (2020). It does not always take two to tango: "Syntrophy" via hydrogen cycling in one bacterial cell. *ISME J*, *14*(6), 1561-1570. https://doi.org/10.1038/s41396-020-0627-1
Wiedenheft, B., van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A., . . . Doudna, J. A. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A*, *108*(25), 10092-10097. https://doi.org/10.1073/pnas.1102716108
Wieringa, K. T. (1939). The formation of acetic acid from carbon dioxide and hydrogen by anaerobic spore-forming bacteria. *Antonie van Leeuwenhoek*, *6*(1), 251-262. https://doi.org/10.1007/BF02146190
Wimmer, F., & Beisel, C. L. (2019). CRISPR-Cas Systems and the Paradox of Self-Targeting Spacers. *Front Microbiol*, *10*, 3078. https://doi.org/10.3389/fmicb.2019.03078
Winther, K. S., & Gerdes, K. (2011). Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. *Proc Natl Acad Sci U S A*, *108*(18), 7403-7407. https://doi.org/10.1073/pnas.1019587108
Wion, D., & Casadesus, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat Rev Microbiol*, *4*(3), 183-192. https://doi.org/10.1038/nrmicro1350
Wood, H. G. (1991). Life with CO or CO2 and H2 as a source of carbon and energy. *FASEB J*, *5*(2), 156-163. https://doi.org/10.1096/fasebj.5.2.1900793
Wood, H. G., Ragsdale, S. W., & Pezacka, E. (1986). The acetyl-CoA pathway of autotrophic growth. *FEMS Microbiology Reviews*, *2*(4), 345-362. https://doi.org/10.1111/j.1574-6968.1986.tb01865.x
Woods, C., Humphreys, C. M., Rodrigues, R. M., Ingle, P., Rowe, P., Henstra, A. M., . . . Minton, N. P. (2019). A novel conjugal donor strain for improved DNA transfer into *Clostridium* spp. *Anaerobe*, *59*, 184-191. https://doi.org/10.1016/j.anaerobe.2019.06.020
Woods, C., Humphreys, C. M., Tomi-Andrino, C., Henstra, A. M., Köpke, M., Simpson, S. D., . . . Minton, N. P. (2022). Required Gene Set for Autotrophic

Growth of. *Appl Environ Microbiol*, *88*(7), e0247921. https://doi.org/10.1128/aem.02479-21
Woolston, B. M., Emerson, D. F., Currie, D. H., & Stephanopoulos, G. (2018). Rediverting carbon flux in *Clostridium ljungdahlii* using CRISPR interference (CRISPRi). *Metab Eng*, *48*, 243-253. https://doi.org/10.1016/j.ymben.2018.06.006
Wu, C., Zhang, J., Wang, M., Du, G., & Chen, J. (2012). *Lactobacillus casei* combats acid stress by maintaining cell membrane functionality. *J Ind Microbiol Biotechnol*, *39*(7), 1031-1039. https://doi.org/10.1007/s10295-012-1104-2
Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C., Hsu, P. D., Dadon, D. B., . . . Sharp, P. A. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol*, *32*(7), 670-676. https://doi.org/10.1038/nbt.2889
Xu, T., Li, Y., Shi, Z., Hemme, C. L., Li, Y., Zhu, Y., . . . Zhou, J. (2015). Efficient Genome Editing in *Clostridium cellulolyticum* via CRISPR-Cas9 Nickase. *Appl Environ Microbiol*, *81*(13), 4423-4431. https://doi.org/10.1128/AEM.00873-15
Xu, Z., Li, Y., & Yan, A. (2020). Repurposing the Native Type I-F CRISPR-Cas System in *Pseudomonas aeruginosa* for Genome Editing. *STAR Protocols*, *1*(1). https://doi.org/10.1016/j.xpro.2020.100039
Xue, C., Seetharam, A. S., Musharova, O., Severinov, K., Brouns, S. J., Severin, A. J., & Sashital, D. G. (2015). CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res*, *43*(22), 10831-10847. https://doi.org/10.1093/nar/gkv1259
Yishai, O., Lindner, S. N., Gonzalez de la Cruz, J., Tenenboim, H., & Bar-Even, A. (2016). The formate bio-economy. *Curr Opin Chem Biol*, *35*, 1-9. https://doi.org/10.1016/j.cbpa.2016.07.005
Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., . . . Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, *163*(3), 759-771. https://doi.org/10.1016/j.cell.2015.09.038
Zhang, J., Hong, W., Zong, W., Wang, P., & Wang, Y. (2018). Markerless genome editing in *Clostridium beijerinckii* using the CRISPR-Cpf1 system. *J Biotechnol*, *284*, 27-30. https://doi.org/10.1016/j.jbiotec.2018.07.040
Zhang, J., Zong, W., Hong, W., Zhang, Z. T., & Wang, Y. (2018). Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. *Metab Eng*, *47*, 49-59. https://doi.org/10.1016/j.ymben.2018.03.007
Zhang, S., & Voigt, C. A. (2018). Engineered dCas9 with reduced toxicity in bacteria: implications for genetic circuit design. *Nucleic Acids Res*, *46*(20), 11115-11125. https://doi.org/10.1093/nar/gky884
Zhang, Y., Grosse-Honebrink, A., & Minton, N. P. (2015). A universal mariner transposon system for forward genetic studies in the genus Clostridium. *PLoS One*, *10*(4), e0122411. https://doi.org/10.1371/journal.pone.0122411
Zhang, Y., Sukthankar, P., Tomich, J. M., & Conrad, G. W. (2012). Effect of the synthetic NC-1059 peptide on diffusion of riboflavin across an intact corneal epithelium. *Invest Ophthalmol Vis Sci*, *53*(6), 2620-2629. https://doi.org/10.1167/iovs.12-9537
Zhao, R., Liu, Y., Zhang, H., Chai, C., Wang, J., Jiang, W., & Gu, Y. (2019). CRISPR-Cas12a-Mediated Gene Deletion and Regulation in *Clostridium*

ljungdahlii and Its Application in Carbon Flux Redirection in Synthesis Gas Fermentation. *ACS Synth Biol*, *8*(10), 2270-2279. https://doi.org/10.1021/acssynbio.9b00033

Zheng, Y., Han, J., Wang, B., Hu, X., Li, R., Shen, W., . . . Peng, W. (2019). Characterization and repurposing of the endogenous Type I-F CRISPR-Cas system of *Zymomonas mobilis* for genome engineering. *Nucleic Acids Res*, *47*(21), 11461-11475. https://doi.org/10.1093/nar/gkz940

Zhou, X., Wang, X., Luo, H., Wang, Y., Wang, Y., Tu, T., . . . Zhang, J. (2021). Exploiting heterologous and endogenous CRISPR-Cas systems for genome editing in the probiotic *Clostridium butyricum*. *Biotechnol Bioeng*, *118*(7), 2448-2459. https://doi.org/10.1002/bit.27753

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., & Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res*, *39*(Web Server issue), W347-352. https://doi.org/10.1093/nar/gkr485

Zhu, Y., Liu, X., & Yang, S. T. (2005). Construction and characterization of pta gene-deleted mutant of *Clostridium tyrobutyricum* for enhanced butyric acid fermentation. *Biotechnol Bioeng*, *90*(2), 154-166. https://doi.org/10.1002/bit.20354

Appendices

List of supplementary figures

Supplementary Figure 1: Schematic representation of all the steps of the Python script as described in this study. 303

Supplementary Figure 2: Fluorescence and corresponding mutations for *E. coli* strains containing the LOV-domain proteins creiLOV, Pp1, and Bs2 under the thiolase promoter (P_{thl}). 327

Supplementary Figure 3: InterPro data for *A. woodii* gene *mttC7*, protein H6LB49 . 338

Supplementary Figure 4: InterPro data for *A. woodii* gene Awo_c03590, protein H6LH22 339

Supplementary Figure 5: InterPro data for *C. autoethanogenum* SADH protein, U5RQB1 340

Supplementary Figure 6: InterPro data for *C. beijerinckii* SADH protein, P25984... 341

Supplementary Figure 7: Emission and excitation scan of *A. woodii* strains containing the Pp1 protein expressed with the P_{thl} promoter. 342

Supplementary Figure 8: Standard curve of fluorescence in relation to the percentage of vitamin solution present in the AWM medium. 343

List of supplementary tables

Supplementary Table 1: Attributes kept in the Spacer object for each hit during the Python script run. 303

Supplementary Table 2: Bacterial strains 304

Supplementary Table 3: Rich *A. woodii* media: liquid *Acetobacterium woodii* liquid medium (AWM), modified agar ATCC medium 1019, and buffered fermentation medium. 305

Supplementary Table 4: Rich *C. autoethanogenum* YTF media 307

Supplementary Table 5: Minimal *C. autoethanogenum* PETC MES media 310

Supplementary Table 6: Primers 312

Supplementary Table 7: Plasmids 321

Supplementary Table 8: GeneBlocks ordered to Integrated DNA Technologies (IDT) 325

Supplementary Table 9: Primers for RNA quality control and RT-qPCR (Chapter 3 and 5) 327

Supplementary Table 10: Primers used for sample preparation and sequencing of the TraDis library. 328

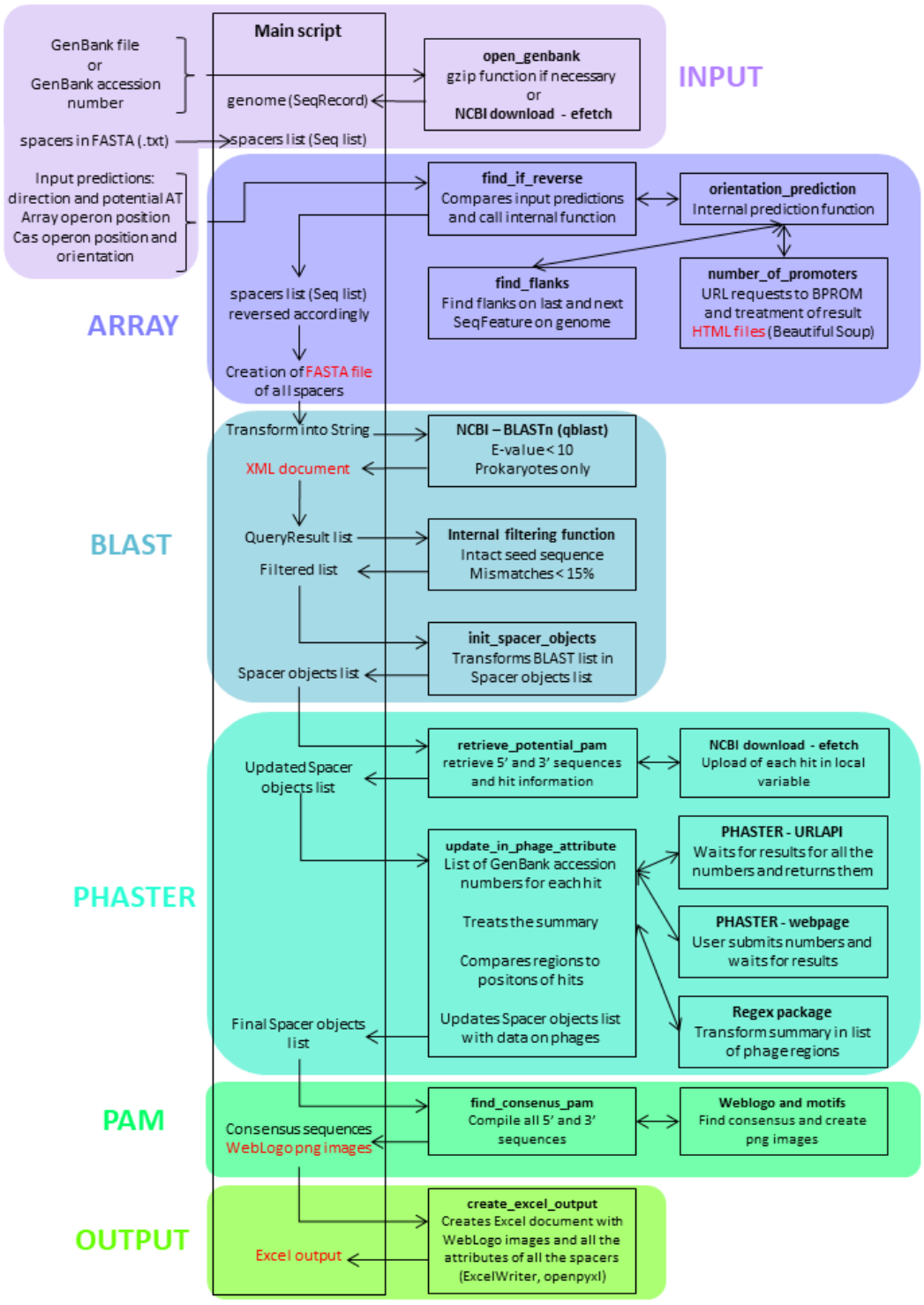
Supplementary Table 11: List of protospacer matches in invading elements from *A. woodii* spacers obtained with the Python script described. 329

Supplementary Table 12: List of protospacer matches in invading elements from *C. autoethanogenum* spacers obtained with the Python script described in increasing mismatch percentage (up to 15%). 330

Supplementary Table 13: List of protospacer matches in invading elements from *C. autoethanogenum* spacers obtained with the Python script described in increasing mismatch percentage (between 15% and 20%). 332

Supplementary Table 14: Genes with highest fold change of abundance superior during fructose fermentation. 334

Supplementary Table 15: Genes with highest fold change of abundance superior during H_2/CO_2 fermentation. 335



Supplementary Figure 1: Schematic representation of all the steps of the Python script as described in this study. The colour code for each block references Figure 6. The outputs saved to the working directory are written in red, they allow the user to restart the script without submitting again to the different databases or tools, the functions and objects taken from external packages are indicated.

Supplementary Table 1: Attributes kept in the Spacer object for each hit during the Python script run.

Column name	Details
number	Spacer identification number
query	Query number from the original spacer list
iterations	Number of duplicate matches with the same attributes
sequence	Full sequence of the query
target_acc	GenBank accession number of the hit organism found with BLAST
target_org	NCBI name of the hit organism
query_pos	Position of the hit in the query sequence
hit_pos	position of the query in the hit sequence
hit_direc	Direction of the hit, 1 for forward and -1 for reverse
hit_sequence	Sequence of the hit
similarity	Similarity between hit and query sequences, each bar represents a nucleotide match
mismatch	Mismatch percentage between query and hit
hit_locus	Locus in the target genome where the hit was found
hit_product	Product from the gene where the hit was found
in_phage	Indicates if the match was found to be in a phage sequence by PHAST
phage_completeness	Completeness of the phage as indicated by PHAST
phage_name	Phage name as indicated by PHAST
fivep	5′ sequence retrieved from the target organism
threep	3′ sequence retrieved from the target organism

Supplementary Table 2: Bacterial strains

Bacterial strain	Genotype	Source
General		
***Escherichia coli* TOP10**	*mcrA, Δ(mrr-hsdRMS-mcrBC), Phi80lacZ(del)M15, ΔlacX74, deoR, recA1, araD139, Δ(ara-leu)7697, galU, galK, rpsL(Sm^R), endA1, nupG*	Invitrogen
***Acetobacterium woodii* DSM1030**	Wild type strain	DSMZ
***Clostridium autoethanogenum* DMS10061**	Wild type strain	DSMZ
Acetobacterium woodii tcdR	*tcdR* gene inserted at *pyrE*	Baker, 2020
***Acetobacterium woodii tcdR*/P$_{tcdB}$_acetone**	*tcdR* gene inserted at *pyrE*, acetone operon under P$_{tcdB}$ at *pheA*.	Baker, 2020
Chapter 3		
***Acetobacterium woodii* Δ*pheA* 1.0 kb 5 h**	Δ*pheA*, obtained with 1.0 kb HAs and 5 h recovery	This study
***Acetobacterium woodii* Δ*pheA* 1.0 kb 7.5 h**	Δ*pheA*, obtained with 1.0 kb HAs and 7.5 h recovery	This study
***Acetobacterium woodii* Δ*pheA* 1.5 kb 5 h**	Δ*pheA*, obtained with 1.5 kb HAs and 5 h recovery	This study
***Acetobacterium woodii* Δ*pheA* 1.5 kb 7.5 h**	Δ*pheA*, obtained with 1.5 kb HAs and 7.5 h recovery	This study
***Acetobacterium woodii* Δ*hsdR1* colonies 1 to 3**	Δ*hsdR1*	This study
***Clostridium autoethanogenum* Δ*pyrE* col. 1 to 6**	Δ*pyrE*	This study
Chapter 5		
Acetobacterium woodii* Δ*gatD	Δ*gatD*	This study
***Acetobacterium woodii* FAST KI**	P$_{thl}$_FAST inserted at *pheA* locus	This study
***Acetobacterium woodii* Acetone KI**	P$_{tcdB}$_acetone inserted at *pheA* locus	This study

Supplementary Table 3: Rich *A. woodii* media: liquid *Acetobacterium woodii* liquid medium (AWM), modified agar ATCC medium 1019, and buffered fermentation medium.

	AWM, for 1 L	**Modified ATCC 1019, for 1 L**	**Buffered *A. woodii* medium, for 1 L**
K_2HPO_4	0.45 g	0.4 g	8.44 g
KH_2PO_4	0.33 g	0.4 g	1.76 g
NH_4Cl	1.0 g	1.0 g	1.0 g
NaCl	20 mg	0.1 g	2.9 g
$MgSO_4.7H_2O$	0.16 g	0.1 g	N/A (See notes below)
$CaCl_2$	20 mg	N/A	N/A
yeast extract	2.0 g	1.0 g	2.0 g
SL9	1 mL	1 mL	1 mL
selenite-tungstate	1 mL	1 mL	1 mL
vitamin solution	20 mL	2 mL	2.0 mL
$NaHCO_3$	10.0 g	3.0 g	$KHCO_3$ 6.0 g
L-cysteine*	0.04% (w/v)	40 µL / plate	Cysteine hydrochloride 0.5 g
Fructose*	20 mM	20 mM	20 mM

Adjust pH to 7.2 with HCl after autoclaving and filter sterilise before adding fructose and L-cysteine. Asterisks indicate elements added after autoclaving.	Add 15 g/L of agar to make plates. Asterisks indicate elements added after autoclaving. Fructose is added when pouring the plates and L-cysteine is spread on the plates before use.	Add 1.5 mM $MgSO_4$ and 20 mM fructose after autoclaving. In this medium L-cysteine is replaced by cysteine hydrochloride added before autoclaving

Trace element solution SL9, for 1 L		**Vitamin solution, for 1 L**		**Selenite-tungstate solution, for 1 L**	
nitrilotriacetic acid	12.8 g	Biotin	2.00 mg	NaOH	0.5 g
$FeCl_2.4H_2O$	2.0 g	Folic acid	2.00 mg	$Na_2SeO_3.5H_2O$	3 mg
$ZnCl_2$	0.070 g	Pyridoxine-HCl	10.00 mg	$Na_2WO_4.2H_2O$	4 mg
$MnCl_2.4H_2O$	0.1 g	Thiamine-HCl.$2H_2O$	5.00 mg	Filter sterilise and store at 4°C.	
H_3BO_3	0.006 g	Riboflavin	5.00 mg		
$CoCl_2.6H_2O$	0.19 g	Nicotinic acid	5.00 mg		
$CuCl_2.2H_2O$	0.002 g	D-Ca-pantothenate	5.00 mg		
$NiCl_2.6H_2O$	0.024 g	Vitamin B12	0.10 mg		
$Na_2MoO_4.2H_2O$	0.036 g	p-Aminobenzoic acid	5.00 mg		
First dissolve nitrilotriacetic acid and adjust pH to 6.0 with NaOH, then add minerals. Filter sterilise and store at 4°C.		Lipoic acid	5.00 mg		
		Filter sterilise and store at 4°C.			

Supplementary Table 4: Rich *C. autoethanogenum* YTF media

YTF, for 1 L	
Yeast extract	10 g
Tryptone	16 g
Fructose	10 g
NaCl	0.2 g
Acidic trace elements solution	1 ml
Basic trace elements solution	1 ml
Vitamin solution	1 ml
Adjust pH to 5.8 with HCl and autoclave. For agar media, add 15 g of bacteriological agar.	

Vitamin solution, for 1 L		**Acidic trace elements solution, for 1 L**		**Basic trace elements solution, for 1 L**	
p-aminobenzoate	100 mg	HCl	50 mM	NaOH	10 mM
riboflavin	100 mg	H_3BO_3	100 mg	Na_2SeO_3	58 mg
thiamine	200 mg	$MnCl_2.4H_2O$	230 mg	Na_2WO_4	53 mg
nicotinate	200 mg	$FeCl_2.4H_2O$	780 mg	$Na_2MbO_4.2H_2O$	52 mg
pyridoxin	500 mg	$CoCl_2.6H_2O$	103 mg	Filter sterilise and store at 4°C.	
pantothenate (calcium)	100 mg	$NiCl_2.6H_2O$	602 mg		
cyanocobalamin	100 mg	$ZnCl_2$	78 mg		
d-biotin	20 mg	$CuSO_4.5H_2O$	50 mg		
folate	50 mg	$AlK(SO_4)_2.12H_2O$	50 mg		
lipoate/thioctic acid	50 mg	Filter sterilise and store at 4°C.			
2-mercapto sulfonic acid	-				
Dissolve p-aminobenzoate separately					

by titrating with 5 M NaOH until dissolved. Dissolve riboflavin separately at 50ºC (0.012% solubility). Dissolve lipoate separately by titrating with 5 M NaOH until dissolved. Final pH 7.8; filter sterilise and store at 4°C

Supplementary Table 5: Minimal *C. autoethanogenum* PETC MES media

PETC MES minimal, for 1 L	
NH_4Cl	1.00 g
KCl	0.10 g
$MgSO_4.7H_2O$	0.20 g
KH_2PO_4	0.20 g
$CaCl_2$	0.02 g
Nitrilotriacetic Acid	0.05 g
$Fe(SO_4)_2(NH_4)_2.6H2O$	0.05 g
CH_3COONa	0.25 g
MES buffer	20.00 g
Trace metals (100X)	10 ml
Wolfe's vitamins (100X)	10 ml
Resazurin (2 $g.L^{-1}$) (2000X)	0.5 ml
Adjust pH to 5.8 with NaOH and autoclave. For agar media, add 15 g of bacteriological agar.	

Wolfe's vitamin solution, for 1 L		**100X Trace metal solution, for 1 L**	
Biotin	0.004 g	Nitrilotriacetic Acid	2.00 g
Folic acid	0.004 g	$MnSO_4.H_2O$	1.00 g
Pyridoxine hydrochloride	0.002 g	$Fe\ (SO_4)_2(NH_4)_2.6H_2O$	0.80 g
Thiamine.HCl	0.010 g	$CoCl_2.6H_2O$	0.20 g
Riboflavin	0.010 g	$ZnSO_4.7H_2O$	0.0002 g

Nicotinic acid	0.010 g	$CuCl_2.2H_2O$	0.02 g
Calcium D-(+)-pantothenate	0.010 g	$NaMoO_4.2H_2O$	0.02 g
Vitamin B_{12}	0.0002 g	Na_2SeO_3	0.02 g
p-Aminobenzoic acid	0.010 g	$NiCl_2.6H_2O$	0.02 g
Thioctic acid	0.010 g	$Na_2WO_4.2H_2O$	0.02 g
Filter sterilise and store at 4°C.			

Supplementary Table 6: Primers

Primer name	Sequence	Description
GENERAL - pMTL80000 series		
ColE1+tra-F2	CCATCAAGAAGAGCGAC	Colony PCRs and Sanger sequencing of vectors
pBP1-R1	CTTCATTAAATGCCTTAGAATC	
pCB102-R1	CTGTTATGCCTTTTGACTATC	
pCD6-R1	GACTTTAAGCCTACGAATACC	
catPrep-F1	GGAAACAATCATAGAATGGAAG	
catP-R1	CAAGTTTATCGCTCTAATGAAC	
R-M recognition site		
FW_targetIIP	CAGGAAACAGCTATGACCGCGGCCGCAGACAAACTCCGATTCAAGTTTTGGATATTTCTTAATC	R-M site generation
RV_targetIIP	GACGCGTGACGTCGACTCTAGAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAA	
R-M recognition site and *cas9* nuclease toxicity		
FW_IIPmut	GGGTATTTCTTAATCAAAGCAGTTCC	Amplification of *cas9* in fragments with R-M site mutates / SNPs for abolition of nucleolytic activity
RV_IIPmut	GCTTTGATTAAGAAATACCCAAAACTTG	
FW_NotI_cas9	CAGGAAACAGCTATGACCGCGGCCGCTGTATCCATATGAC	
RV_Pfdx_XbaI	AAGGGACTAAATATAAATCTAGAGTGTAGTAGCCTGTGAAATAAG	
FW_D10A	CGATAGCTAAGCCTATTGAGTATTTCT	
RV_D10A	CAATAGGCTTAGCTATCGGCACAAA	
FW_H840A	ACAATTGCATCGACATCATAATCACTTAAAC	
RV_H840A	TATGATGTCGATGCAATTGTTCCACAAAG	

***A. woodii* interference assay**		
FW_proto8	AAACAGCTATGACCGCGGCCCTAGACAAATGAATCAAAATCATTAATAGCTTTAATTAAAT	Assembly of each of the 3 protospacer with and without 5′ PAM
RV_proto8	GCGCGCCATTGACGTCTATGTCGACTAAATGACGATTTAATTAAAGCTATTAATGATTTTGA	
FW_proto8-5′	AAACAGCTATGACCGCGGCCCTAGAATGCCGAATCAAAATCATTAATAGCTTTAATTAAAT	
FW_proto20	AAACAGCTATGACCGCGGCCCTAGACAAATTTATTCTATGAGGTGATGAGATGAATGTT	
RV_proto20	CGCCATTGACGTCTATGTCGACTAAATTCTTAATAACATTCATCTCATCACCTCAT	
FW_proto20-5′	AAACAGCTATGACCGCGGCCCTAGAGGGTGTTATTCTATGAGGTGATGAGATGAATGTT	
FW_proto22	AAACAGCTATGACCGCGGCCCTAGACAAATAATAACTATCAGGCTATCAGCTTCGGGG	
RV_proto22	CCATTGACGTCTATGTCGACTAAATAAAGTCTCCCCGAAGCTGATAGCCTGATA	
FW_proto22-5′	AAACAGCTATGACCGCGGCCCTAGAAACCAAATAACTATCAGGCTATCAGCTTCGGGG	
***A. woodii* leader characterisation**		
FW_leader	CAGGAAACAGCTATGACCGCGGTTGTCGACCTGTGAT	Assembly of leader_MCS
RV_leader_SacI	ATATGGATCTACTCAAATGAGCTCCCGTGCTTGTTAGATACC	
FW_catP_SacI	GCACGGGAGCTCATGGTATTTGAAAAAATTG	Assembly of leader sequence with catP or FAST gene
RV_catP_NheI	GCAGGAATTGATAAATAGTTAAGCTAGCATAAAA	
FW_FAST_SacI	GCATAAGAGCTCATGGAACACGTAGC	
RV_FAST_NheI	ATTTTTATGCTAGCTCATACCCTCTTAACG	
***A. woodii pyrE* KO**		

FW_LHA_pyrE	GGCTTCTTATTTTTATGACGTCTAGCAACTGGGAAGATGCG	Amplification of HAS from gDNA
RV_LHA_pyrE+BM4	CCTTCCGTACAACCCACAACCCTTTTTCGTTCCGTAAAAATGG	
FW_BM4+1+RHA_pyrE	AGGGTTGTGGGTTGTACGGAAGGATAAGTTAAAGAGCACGTAAA	
RV_RHA_pyrE	AATAATGGCGGCGCGCCGCTGGTAACATGAATGC	
FW_leader	CAGGAAACAGCTATGACCGCGGTTGTCGACCTGTGAT	Amplification of leader sequence
RV_leader	ATCTACTCAAATGTCGACCCGTGCTTGTTAGATACCTAT	
RV_RHA_pyrE+MCS	TCTTATTTTTATGCTAGCGCTGGTAACATGAATGC	RV editing cassette
FW_pyrE_endo2	TCTAACAAGCACGGGTCGACATTTGAGTAGATCCATATTGGAATGTAAATAGGC GATCTTACCCGCTTGACAAACCGGGCAATCGT	spacer assembly with flanking DR
RV_pyrE_endo2	CGCATCTTCCCAGTTGCTAGACGTCATTTACATTCCAATATGGATCTACTCAAAT ACGATTGCCCGGTTTGTCAAGCGGGTAAGAT	
AW_CRISPR_pyrE_armF	TATACGCGCGCGGATATCCCGGTTATGGCGAT	Colony PCR for *pyrE* locus
AW_pyrEcomp_RHAR	TATAGGCGCGCCCCATCCATTTTCATCTGATT	
***A. woodii pheA* KO**		
FW_pheA_spacer	TAACAAGCACGGGAGCTCATTTGAGTAGATCCATATTGGAATGTAAATAACCAA TGTCACCCGATTTGTGATTATTTCCCGCA	spacer assembly with flanking DR
RV_pheA_spacer	CCATGGACGCGTGACGTCATTTACATTCCAATATGGATCTACTCAAATTGCGGGA AATAATCACAAATCG	
RV_LHA_pheA	TAATCTGGAAATTCTTTGACTTAAAATCTTCTC	LHA and RHA amplification, lengths of 0.5kb, 1.0kb and 1.5kb
FW_RHA_pheA	TAAGTCAAAGAATTTCCAGATTATAGATTTCCCCGGC	
FW_L0.5	AAATGACGTCCGGCACAGTCGATGTTAG	
RV_R0.5	TATGCTAGCTCTATCCGCAAACGCTT	
FW_L1.0	AAATGACGTCATTGACGGCTGGGG	

RV_R1.0	ATGCTAGCGAGATTGGCCTGCTC	
FW_L1.5	TAAATGACGTCCAGCCGACTGATACTGA	
RV_R1.5	TTTTATGCTAGCAAGCCTCGTCATCGC	
FW_pheAHA1.5_out	CCGGAAAGGGATCAAATTTG	Colony PCR primers
RV_pheA HA1.5_out	TGGTTTCGTCTTTTTTAGGTAG	
A. woodii *hsdR1* KO		
FW_hsdR1_spacer	AACAAGCACGGGAGCTCATTTGAGTAGATCCATATTGGAATGTAAATTCGAAATACTGACCGGCCTGGGCTAC	spacer assembly with flanking DR
RV_hsdR1_spacer	CTGTATTTGAATTTCATTGACGTCATTTACATTCCAATATGGATCTACTCAAATAGGCATACGCGTAGCCCAGGCCGGTCAGTATT	
FW_LHA_hsdR1	GTAAATGACGTCGGTTGTGATAGAGCTT	LHA and RHA amplification
RV_LHA_hsdR1	CTGCTACACCCGTAAACTCATAACGTGTCCC	
FW_RHA_hsdR1	GAGTTTACGGGTGTAGCAGGTTG	
RV_RHA_hsdR1	TATTTTTATGCTAGCAAGTGTTCACC	
FW_hsdR1_screen_1.5	GCTGTTGATGATGGGATG	Colony PCR primers
RV_hsdR1_screen_1.5	CATTGGCATAGGGATTTTTC	
FW_hsdR1_del_screen	CAATAAAGTAACGGGTTGATTA	Deletion sequencing
RV_hsdR1_del_screen	CAAGTATGCTATGATTGTATC	
***C. autoethanogenum pyrE* KO**		
FW_leader4_CLAU	CTATGACCGCGGAATACACATTCTCCCTCAT	Leader of CRISPR array4
RV_leader4_CLAU	GGTACCGAGCTCCACAGTAAAATAGCCATTTCT	
FW_CLAUpyrE_HA	CCGGGGATCCGTACCGCTG	HAs from vFS67
RV_CLAUpyrE_HA	TTTATGCTAGCCCTGTAATCGGAGCATC	

FW_CLAUpyrE_spacer1	ATGGCTATTTTACTGTGGAGCTCGTTGAACCTCAACATGAGATGTATTTAAATCTGTGTTTTCCTGAAGAAAGTAGAAAATG	spacer assembly with flanking DR
RV_CLAUpyrE_spacer1	CGCGTGACGTCGACTCTAGAGGATCCATTTAAATACATCTCATGTTGAGGTTCAACGAAGGGCATTTTCTACTTTCTTCAGGAAAA	
FW_pyrECLAU_screening	GTTAATTGAGAATCTTTTGCTATTAC	Colony PCR primers
RV_pyrECLAU_screening	GTCTTGTATGTACTTGTAGTAC	
CHAPTER 5		
Plasmid-based production of isopropanol with *C. beijerinckii* SADH		
FW_hifi_adc_Cbe_adh	GTTATATATGATTATCTTAAGTAGAGGAGGAACATATTTTATGAAAGGTTTTGCAATG	Vector assembly
FW_hifi_Pfdx_Cbe_adh	TTTTAAGGAGGTGTGTTACATATGAAAGGTTTTGCAATG	
RV_hifi_Cbe_adh	GCCTGCAAATGCAGGCTTCTTATTTTTATGCTAGCTTATAATATAACTACTGCTTTA	
FW_seq_thlA	TTTCTTATGGTTCAGCAGG	Col. PCRs and seq.
FW_seq_Cbe_adh	ATCAGATATACATACTGTTTT	
Overexpression and *gatD* KO		
FW_AWO_gatD	GTAAAGGAGAAAATCAtatgaaagcggtatgtatg	*gatD* amplification
RV_AWO_gatD	cttatttttatgctagcttaattctccccccatg	
FW_gatD_spacer	gcacgggagctcatttgagtagatccatattggaatgtaaatgtcttccccgggataaggtttggtaaac	Spacer assembly
RV_gatD_spacer	ctggcgggacGTCatttacattccaatatggatctactcaaattggaattcgtttaccaaaccttatcccg	
FW_gatD_LHA	gaatgtaaatGACgtcccgccagtaaaatatttaag	HAs assembly
RV_gatD_LHA	cattgccttattcatttaatttttcatcttttcctccttaatg	
FW_gatD_RHA	cattaaggaggaaaagatgaaaaattaaatgaataaggcaatg	
RV_gatD_RHA	caggcttcttatttttatgctagcgttcattcatataaacatagg	
FW_gatD_seq	gtgtttaaaacttcaaggc	

RV_gatD_seq	tgtcagtgatgacatcac	Genome locus sequencing
creiLOV, Bs2, and Pp1 fluorescent proteins		
FW_Pthl_backbone	CTATGACCGCGGCCGCtttttaacaaaatatattgataaaaataataatag	Assembly of the 3 genes in pMTL82221
RV_Pthl	atgaactaacctcctaaattttg	
FW_creiLOV_Pthl	aatttaggaggttagttcatATGGCTGGATTAAGGCATACTTTC	
RV_creiLOV_backbone	CATGTCTGCAGGCCTCGAGTTAAGCCAAAGCCTTTCCTTCAGTTTTG	
FW_evoglow-Pp1_Pthl	ttaggaggttagttcatATGATTAATGCTAAATTATTGCAGTTG	
RV_evoglow-Pp1_backbone	CATGTCTGCAGGCCTCGAGTTAGTGCTTAGCTTGTCCTTGCTGTC	
FW_evoglow-Bs2_Pthl	aatttaggaggttagttcatATGGCAAGTTTTCAAAGTTTTG	
RV_evoglow-Bs2_backbone	GTCTGCAGGCCTCGAGTTATTCTAATAGTTTCTCATATTC	
FW_NdeI_creiLOV_corrected	gttacatatgGCTGGATTAAGGCATA	Assembly of the 3 genes in pMTL84152
RV_AatII_creiLOV	gcgtgacgtcTTAAGCCAAAGCCTTTC	
FW_NdeI_Pp1_corrected	gtgtgttacatatgATTAATGCTAAATTATTGCAGT	
RV_AatII_Pp1	gcgtgacgtcTTAGTGCTTAGCTTG	
FW_NdeI_Bs2_corrected	tgtgttacatatgGCAAGTTTTCAAAGTTTTG	
RV_AatII_Bs2	acgcgtgacgtcTTATTCTAATAGTTTCTCATAT	
FAST KI at the *pheA* locus		
FW_AWO_pheA_CCG_KI_spacer	AAGCACGGGAGCTCATTTGAGTAGATCCATATTGGAATGTAAATCTTTTTTGACCTTAAGCAGTCGTCGTTAAAGCGGGGATTTGAGTA	spacer assembly with flanking DR
RV_AWO_pheA_CCG_KI_spacer	GCAAAAATAACCCTCGAGACGTCATTTACATTCCAATATGGATCTACTCAAATCCCCGCTTTAACGACGACTGCTTAAGGTCAA	
FW_AWO_LHA_pheA_cargo	GTAAATGACGTCTCGAGGGTTATTTTTGCAA	

RV_AWO_LHA_pheA_CCG_cargo_MCS	GAATCGTCATATGGACAGCGGCCGCGACTCGGATCCGGGGTTATCTGTCATG	LHA and RHA amplification
FW_AWO_RHA_pheA_CCG_cargo_MCS	GCCGCTGTCCATATGACGATTCTAGACGGCTTTTTTGACCTTAAGCAGTCGTC	
RV_AWO_RHA_pheA_cargo	TTTATGCTAGCCAGCCGACTGATACTGAA	
FW_BamHI_FAST_cassette	GATAACCCCGGATCCTTTTTAACAAAATATATTGATAAAAAT	Cargo amplification P_{thl}_FAST_T
RV_FAST_NotI_AscI	TTAAGGGCGGGCGCGCCAGCGGCCGCTCATACCCTCTTAACGAA	
FW_pheA_KI_seq	GTCAATGATTCATCGAGTTC	Colony PCR and Sanger sequencing
RV_pheA_KI_seq	gataattatcagaatcttagtaaatc	
FW_pheA_cargo_seq	CTTCACTGGTAATATTTCAC	
RV_pheA_cargo_seq	TGATTATAGCCCGGTAGTAGA	
T and CT terminator characterisation		
FW_catP_NdeI	gttagaCAtatggtatttgaaaaaattg	*catP* amplification
RV_catP_NdeI	GACTAGTCATaTGttaactatttatcaattcctg	
FW_RBS_FAST_AscI	ATATATggcgcgcctaaatttaggaaggtgacttttATGGAACACGTAGCATTTG	FAST amplification
RV_FAST_AscI	cgggcgcgccTCATACCCTCTTAACGAAAACC	
FW_T	gttagaCAtATGGTAAAATAAATGTGCCTCAA	Terminators amplification
FW_CT	gttagaCAtATGACTAGTCTGCCAATAGA	
FAST transposon library		
FW_Pthl_ITR_GL15	GATTACGAATTCGAGGTATAACAGGTTGGCTGATAAGTCCCCGGTCTGACAccgcGctttttaacaaaatatattgat	FAST cassette amplification
RV_pyrEhydA_GL15	CTTCTTGAACACTTGCCGAAAAACAATTGggcgcgccaATATAT	
FW_catp_cassette_corrected	gccCAATTGTTTTTCGGCAAGTGTTCAAGAAGTTATT	

RV_XbaI_CatP	GTGACGTCGACTCTAGAGGATCCCCGGG	*catP* cassette amplification
FW_cassette_GL15	GTCGACTCTAGAGGATC	SOEing PCR FAST-catP
RV_cassette_GL15	TATACCTCGAATTCGTAATC	
catP_INV_F1_bis	TGAAAGTGATACGCAACGGTATGG	Inverse PCR
catP_INV_R1_bis	CATCTCATCATATATCCCCAATTCACC	
catP_INV_F1	TAAATCATTTTTAGCAGATTATGAAAGTGATACGCAACGGTATGG	
catP_INV_R1	TATTGTATAGCTTGGTATCATCTCATCATATATCCCCAATTCACC	
C. beijerinckii* SADH KI at *purD		
FW_purD_spacer	aagcacgggagctcatttgagtagatccatattggaatgtaaatatgtaaacagagtccgttactactaaatag	*purD* spacer assembly
RV_purD_spacer	gttcccgtccGACGTCatttacattccaatatggatctactcaaatgccaatctatttagtagtaacggactctg	
FW_purD_LHA	gtaaatGACGTCggacgggaacatg	HAs assembly
FW_purD_LHA	gaatcgtcatatggacagcggccgcgactcGGATCCattgctaattaaacttagc	
FW_purD_RHA	gccgctgtccatatgacgattctagaatgtaaacagagtccgttactactaaat	
RV_purD_RHA	gcttcttatttttatgctagcgtccgcacagcggc	
FW_NotI_PtcdB	CCgagtcGCGGCCGCTTAATGAATTTAAAG	PtcdB_Cbe_adh amplification
RV_XbaI_adh	atctgtatctagagcgggcgcgcc	
FW_purD cargo seq	gcgttctttgtgtgacag	Cargo sequencing
RV_purD cargo seq	ctatcaacattgaaatgattcaac	
FW_purD_KI_seq	gagcttgatgaaagtattaatgattggatc	Genome locus sequencing
RV_purD_KI_seq	cgatagcaacggatcttc	
Acetone operon KI at *pheA* locus		
FW_seq_ctfA	ttctataaaggtactactaaaaac	
RV_seq_ctfA	aactattacggttttagctg	

FW_seq_thlA	TTTCTTATGGTTCAGCAGG	Sequencing of the acetone operon
JPB_475	tgaactggttaatgtcgatg	
JPB_476	gtctttttcaatccaagttggg	
JPB_657	GACTTCATTAACATCCTCTG	
JPB_693	gtctaaagttcctactaaagt	
JPB_694	GATGTACCAGCAGTAGATTT	

Supplementary Table 7: Plasmids

Plasmid name	Description	Source
General backbones		
pMTL82151	Cm^R; pBP1	SBRC Nottingham culture collection
pMTL84151	Cm^R; pCD6	
pMTL84152_P_{thl}_FAST	Cm^R; pCD6; P_{thl}_FAST	
pTML83251	Erm^R; pBP1; ColE1; MCS	
pMTL84422	tet^R; pBP1; p15a; P_{thl}_MCS	
pMTL82221	Erm^R; pBP1; p15a; MCS	
pMTL84153	Cm^R; pCD6; ColE1; P_{fdx}_MCS	
pMTL84252	Erm^R; pCD6; ColE1; P_{thl}_MCS	
pMTL84152	Cm^R ; pCD6; ColE1; P_{thl}_MCS	
pRECas1_MCS	Cm^R; $P_{fdx\text{-}RbE}$-Cas9; P_{araE}-MCS; pCB102	Canadas et al. 2019
pMTL-GL15	Erm^R; *himarI* C9 under the control of *C. difficile* P_{tcdB}; *catP* gene in transposon cassette	Little et al. 2018
pMTL-MPD15	Cm^R; leader sequence upstream a MCS; pCD6	This study
pMTL-MPD23	Cm^R; leader sequence upstream *pheA* intergenic spacer; *pheA* HAs flanking a MCS; pCD6	This study
pMTL-MPD25	Cm^R; *C. autoethanogenum* leader3 sequence upstream a MCS; pCB102	This study
Chapter 3		
R-M recognition site and *cas9* nuclease toxicity		
pMTL8215_targetIIP	Cm^R; R-M site; pBP1	This study
pMTL8215_pREdCas1_MCS	Cm^R; $P_{fdx\text{-}RbE}$-dCas9; P_{araE}-MCS; pBP1	This study

pMTL8215_pREdCas1_IIPmut	Cm^R; $P_{fdx\text{-}RbE}$-dCas9-mut; P_{araE}-MCS; pBP1	This study
pMTL8215_P1339_MCS	Cm^R; P_{araE}-MCS; pBP1	This study
***A. woodii* interference assay**		
pMTL-MPD1	Cm^R; pBP1; *A. woodii* protospacer 8	This study
pMTL-MPD2	Cm^R; pBP1; *A. woodii* protospacer 8 with 5′ PAM candidate	This study
pMTL-MPD3	Cm^R; pBP1; *A. woodii* protospacer 20	This study
pMTL-MPD4	Cm^R; pBP1; *A. woodii* protospacer 20 with 5′ PAM candidate	This study
pMTL-MPD5	Cm^R; pBP1; *A. woodii* protospacer 22	This study
pMTL-MPD6	Cm^R; pBP1; *A. woodii* protospacer 22 with 5′ PAM candidate	This study
***A. woodii* leader characterisation**		
pMTL-MPD29	Cm^R; pCD6; leader_catP	This study
pMTL-MPD30	Cm^R; pCD6; leader_FAST	This study
***A. woodii pyrE* KO**		
pMTL-MPD7	Cm^R; P_{araE}-MCS; *pyrE* HAs; pBP1	This study
pMTL-MPD8	Cm^R; leader sequence upstream *pyrE* spacer2; *pyrE* HAs; pBP1	This study
pMTL-MPD9	Cm^R; *pyrE* HAs; pBP1	This study
pMTL-MPD10	Cm^R; leader sequence upstream *pyrE* spacer2; pBP1	This study
pMTL-MPD11	Cm^R; leader sequence upstream *pyrE* spacer2; *pyrE* HAs; pBP1	This study
pMTL-MPD12	Cm^R; *pyrE* HAs; pCD6	This study
pMTL-MPD13	Cm^R; leader sequence upstream *pyrE* spacer2; pCD6	This study
pMTL-MPD14	Cm^R; leader sequence upstream *pyrE* spacer2; *pyrE* HAs; pCD6	This study
***A. woodii pheA* KO**		
pMTL-MPD16	Cm^R; *pheA* 0.5kb HAs; pCD6	This study
pMTL-MPD17	Cm^R; *pheA* 1.0kb HAs; pCD6	This study

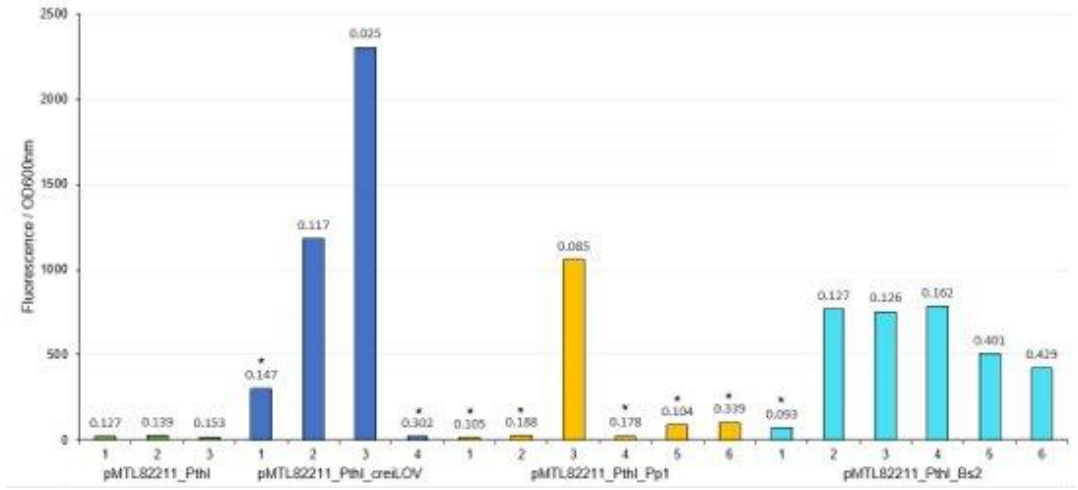
pMTL-MPD18	Cm^R; *pheA* 1.5kb HAs; pCD6	This study
pMTL-MPD19	Cm^R; leader sequence upstream *pheA* spacer; *pheA* 0.5kb HAs; pCD6	This study
pMTL-MPD20	Cm^R; leader sequence upstream *pheA* spacer; *pheA* 1.0kb HAs; pCD6	This study
pMTL-MPD21	Cm^R; leader sequence upstream *pheA* spacer; *pheA* 1.5kb HAs; pCD6	This study
***A. woodii* editing vector**		
pMTL-MPD22	Cm^R; leader sequence upstream *hsdR1* spacer; *hsdR1* HAs; pCD6	This study
***C. autoethanogenum pyrE* KO**		
pMTL-MPD26	Cm^R; *pyrE* HAs; pCB102	This study
pMTL-MPD27	Cm^R; *C. autoethanogenum* leader3 sequence upstream *pyrE* spacer; pCB102	This study
pMTL-MPD28	Cm^R; *C. autoethanogenum* leader3 sequence upstream *pyrE* spacer; *pyrE* HAs; pCB102	This study
CHAPTER 5		
pMTL24-1 or pAcetone	Cm^R; pCD6; acetone operon under the control of *C. difficile* P_{tcdB}	Baker, 2020
pMTL24-1+Adh or pAcetone+Adh	Cm^R; pCD6; acetone operon + *C. beijerinckii* SADH with its native RBS under the control of *C. difficile* P_{tcdB}	This study
pMTL84151_PtcdB_CbeSADH or pAdh	Cm^R; pCD6; *C. beijerinckii* SADH under the control of *C. difficile* P_{tcdB}	This study
pMTL84151_PtcdB_AwoGatD or pGatD	Cm^R; pCD6; *A. woodii gatD* under the control of *C. difficile* P_{tcdB}	This study
pMTL8415_GatD_KO	Cm^R; pCD6; leader sequence upstream *gatD* spacer; *gatD* 1.5kb HAs	This study
pMTL82222_creiLOV	Erm^R; pBP1; p15a; P_{thl}_creiLOV	This study
pMTL82222_Pp1	Erm^R; pBP1; p15a; P_{thl}_Pp1	This study
pMTL82222_Bs2	Erm^R; pBP1; p15a; P_{thl}_BS2	This study
pMTL84153_creiLOV	Erm^R; pBP1; ColE1; P_{fdx}_creiLOV	This study
pMTL84153_Pp1	Erm^R; pBP1; ColE1; P_{fdx}_Pp1	This study
pMTL84153_Bs2	Erm^R; pBP1; ColE1; P_{fdx}_BS2	This study

pMTL84152_Pp1	Erm^R; pBP1; ColE1; P_{thl}_Pp1	This study
pMTL-MPD24	Cm^R; leader sequence with *pheA* intergenic spacer; *pheA* HAs flanking P_{thl}_FAST; pCD6	This study
pMTL84151_Pthl_FAST_CT	Cm^R; pCD6; FAST under the control of P_{thl} and with *C. tetani* terminator	This study
pMTL84252_catP_T	Erm^R; pCD6; ColE1; FAST under the control of P_{thl} and with *C. acetobutylicum pyrE-hydA* terminator	This study
pMTL84252_catP_CT	Erm^R; pCD6; ColE1; FAST under the control of P_{thl} and with *C. tetani* terminator	This study
pMTL84252_catP_T_FAST	Erm^R; pCD6; ColE1; FAST under the control of P_{thl} and with *C. acetobutylicum pyrE-hydA* terminator followed by FAST gene	This study
pMTL84252_catP_CT_FAST	Erm^R; pCD6; ColE1; FAST under the control of P_{thl} and with *C. tetani* terminator followed by FAST gene	This study
pMTL-GL15-FAST	Erm^R; Himar1 C9 transposase under the control of *C. difficile* P_{tcdB}; *catP* and P_{thl}_FAST between ITR	This study
pMTL84151_pheA_KI_PtcdB_CbeSADH	Cm^R; leader sequence with *pheA* intergenic spacer; *pheA* HAs flanking P_{tcdB}_CbeSADH; pCD6	This study
pMTL84151_pheA_KI_PtcdB_acetone	Cm^R; leader sequence with *pheA* intergenic spacer; *pheA* HAs flanking P_{thl}_FAST; pCD6	This study

Supplementary Table 8: GeneBlocks ordered to Integrated DNA Technologies (IDT)

Name	Sequence
creiLOV	ATGGCTGGATTAAGGCATACTTTCGTTGTAGCAGACGCTACA CTACCTGACTGTCCATTGGTGTACGCTAGTGAAGGATTCTATG CAATGACTGGATACGGACCAGACGAGGTGCTAGGTCATAACG CAAGGTTCTTACAGGGTGAGGGAACTGATCCAAAGGAAGTAC AAAAAATTAGAGACGCAATAAAAAAGGGAGAGGCATGTTCA GTGAGATTGTTAAATTATAGAAAGGATGGAACACCATTCTGG AATCTTCTTACAGTGACACCAATAAAAACACCAGACGGAAGA GTTTCTAAATTCGTTGGAGTACAGGTGGACGTAACAAGCAAA ACTGAAGGAAAGGCTTTGGCTTAA
Pp1	ATGATTAATGCTAAATTATTGCAGTTGATGGTTGAACATTCTA ACGATGGAATAGTAGTGGCTGAGCAGGAGGGAAATGAAAGC ATACTTATTTATGTGAATCCTGCTTCTGAAAGATTAACTGGAT ATTGCGCTGATGATATATTGTACCAGGACGCAAGGTTTCTTCA AGGAGAAGACCACGACCAACCTGGTATTGCTATTATTAGAGA GGCTATAAGAGAGGGAAGGCCATGTTGTCAAGTTCTTAGGAA CTATAGAAAAGATGGTTCATTATTCTGGAATGAATTAAGCAT AACTCCAGTTCACAATGAAGCTGACCAGCTAACTTACTACAT AGGAATTCAGAGGGACGTAACTGCACAAGTGTTTGCTGAAGA GAGAGTTAGGGAATTAGAAGCAGAAGTGGCAGAGCTAAGGA GACAGCAAGGACAAGCTAAGCACTAA
Bs2	ATGGCAAGTTTTCAAAGTTTTGGAATTCCTGGTCAGCTTGAAG TAATTAAGAAGGCTTTGGACCACGTTAGGGTTGGAGTAGTGA TAACAGATCCTGCTTTAGAGGACAATCCTATAGTATATGTTAA CCAAGGTTTCGTTCAAATGACAGGATATGAAACAGAGGAGAT TTTAGGAAAGAACGCAAGATTCTTGCAAGGTAAACATACTGA TCCTGCTGAAGTAGATAATATAAGGACTGCACTTCAGAACAA AGAGCCTGTTACTGTACAAATACAAAATTACAAGAAGGACGG AACTATGTTCTGGAATGAGTTGAACATTGACCCTATGGAGAT AGAAGACAAAACATATTTCGTGGGAATTCAGAACGACATAAC AAAGCAAAAGGAATATGAGAAACTATTAGAATAA

(A)



(B)

Construction	Mutation position	Nucleotide	AA impact
Pthl_creiLOV 1	150/360	A deletion	FS glycine
Pthl_creiLOV 4	189/360	A deletion	FS isoleucine
Pthl_Pp1 1	7/447	A deletion	FS asparagine
Pthl_Pp1 2	19/447	T deletion	FS leucine
Pthl_Pp1 4	196/447	A deletion	FS isoleucine
Pthl_Pp1 5	430/447	G deletion	FS glycine
Pthl_Pp1 6	439/447	A deletion	FS lysine
Pthl_Bs2 1	367/414	C → A mutation	glutamine → lysine

Supplementary Figure 2: Fluorescence and corresponding mutations for *E. coli* strains containing the LOV-domain proteins creiLOV, Pp1, and Bs2 under the thiolase promoter (P_{thl}). (A) Fluorescence normalized to OD_{600nm} for multiple *E. coli* colonies for each different construction. Samples collected after 4 h of subcultures, the OD_{600nm} values at the time of sampling were added to the top of each bar. Asterisks indicate strains harbouring a mutation in the fluorescence gene. (B) Mutations in the genes encoding the LOV-domain proteins.

Supplementary Table 9: Primers for RNA quality control and RT-qPCR (Chapter 3 and 5)

Primer name	Sequence	Product size	Comments
gDNA_FW	TATACGCGCGCGGATATCCCGGTTATGGCGAT	1997 bp	gDNA contamination control
gDNA_RV	TATAGGCGCGCCCCATCCATTTTCATCTGATT		
FWD 4	GGAAGGAGTAGCTAGTGGTAAT	145 bp	FAST reporter gene
REV4	CCTCTTAACGAAAACCCAGTAG		
Chow_gyrA_FW	GTAAGTCGGCCCGTATTGTTG	108 bp	*A. woodii* gyrase, internal reference gene (Chowdhury et al, 2020; Shin et al., 2021)
Chow_gyrA_RV	AACGGATCGACCATTCCTGAG		
orfB_FWD	GTATTACACAGGAGTTTGAACC	74 bp	Plasmid backbone.
orfB_REV	AAATACTCTGCTTGGTCGGCTT		
qPCR_catP_FWD	aatggaaggaaagccaaatgctccg	78 bp	*catP* gene
qPCR_catP_REV	ttaaagccatcgaaggttgaccacg		

Supplementary Table 10: Primers used for sample preparation and sequencing of the TraDis library.

Primer name	Sequence
SplA5_top	G*AGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
SplA5_bottom	/5Phos/G*ATCGGAAGAGCGGTTCAGCAGGttttttttttcaaaaaaa*a
5′ Tn1	AATGATACGGCGACCACCGAGATCTACACttcgactgcactcccgacttaataacttc
SplAP5.1	C*AAGCAGAAGACGGCATACGAGATAACGTGATGAGATCGGTCTCGGCATTC*C
SplAP5.2	C*AAGCAGAAGACGGCATACGAGATAAACATCGGAGATCGGTCTCGGCATTC*C
SplAP5.7	C*AAGCAGAAGACGGCATACGAGATCAGATCTGGAGATCGGTCTCGGCATTC*C
SplAP5.8	C*AAGCAGAAGACGGCATACGAGATCATCAAGTGAGATCGGTCTCGGCATTC*C
SplAP5.9	C*AAGCAGAAGACGGCATACGAGATCGCTGATCGAGATCGGTCTCGGCATTC*C
SplAP5.10	C*AAGCAGAAGACGGCATACGAGATACAAGCTAGAGATCGGTCTCGGCATTC*C
SplAP5.11	C*AAGCAGAAGACGGCATACGAGATCTGTAGCCGAGATCGGTCTCGGCATTC*C
SplAP5.12	C*AAGCAGAAGACGGCATACGAGATAGTACAAGGAGATCGGTCTCGGCATTC*C
SplAP5.14	C*AAGCAGAAGACGGCATACGAGATAACCGAGAGAGATCGGTCTCGGCATTC*C
SplAP5.15	C*AAGCAGAAGACGGCATACGAGATAACGCTTAGAGATCGGTCTCGGCATTC*C
SplAP5.16	C*AAGCAGAAGACGGCATACGAGATAAGACGGAGAGATCGGTCTCGGCATTC*C
SplAP5.17	C*AAGCAGAAGACGGCATACGAGATAAGGTACAGAGATCGGTCTCGGCATTC*C
SplAP5.18	C*AAGCAGAAGACGGCATACGAGATACACAGAAGAGATCGGTCTCGGCATTC*C
5′ SeqLong	cttgaacacttgccgaaaaacaattgtgtcagaccggggacttatc

Supplementary Table 11: List of protospacer matches in invading elements from *A. woodii* spacers obtained with the Python script described. # is the position of the spacer in the array; the spacer and protospacer sequences are aligned and mismatches underlined; % is the percentage of mismatch between spacer and protospacer sequences; the invading elements are indicated and position of the hit in the genome indicated if relevant in parentheses; potential 5′ PAM are indicated in the last column based on each match.

#	Spacer – protospacer sequences	%	Invading element	5′ PAM
5	GGGAGGAATAAAAATGAACTTAAAAGATGTAAAAAG -GGAGGAATCAAAATGAACTTAAAAAATTTAAA---	19	*Fusobacterium varium* Fv113-g1 (questionable prophage)	AAATA
5	GGGAGGAATAAAAATGAACTTAAAAGATGTAAAAAG -GGAGGACTAAAAATGAACTTAAATGATGAACAAA-	17	Bacteriophage sp. isolate 150	CTATA
5	GGGAGGAATAAAAATGAACTTAAAAGATGTAAAAAG -GGAGGATTAAAAATGAACTGTAAAGAATTTAAAAG	19	*Fusobacterium* sp. oral taxon 203 (incomplete prophage)	AATAA
8	GAATCAAAATCATTAATAGCTTTAATTAAATCGTC GAATCTAAATTTTTAATAGCTTTTATCACATCGTC	17	*Yersinia* phage phiR1-RT	ATGCC
15	TTGAAGACATGAGAAAAG---CAAAAGAAACAATTCTT TTGAAGAAATGAGTAACGTTCCAAAAGAAACAATT---	17	*Listeria monocytogenes* FDAARGOS_57(unnamed plasmid)	AACTT
29	TTATTCTATGA--GGTGATGAGATGAATGTTATTAAGA TTATTCTATCATTGGTGATAAGATGAATGTTACTA---	17	*Aeromonas* phage AS-zj	GGGTG
34	AATAACTATCAGGCTATCAGCTTCGGGGAGACTTT AATAACTATTATGCTATCAGCTTCAGGCATACT--	20	*Hathewaya histolytica* NCTC503, (phage protein)	AACCA
45	ACCACCAACGCCGGGATTACCCTTGCAAGCGCCATC -CCACCAGGGCCTGGATTACCCTTGCACGCACCATC	17	*Pseudomonas* sp. Leaf58 pBASL58	CCCAG

Supplementary Table 12: List of protospacer matches in invading elements from *C. autoethanogenum* spacers obtained with the Python script described in increasing mismatch percentage (up to 15%). # refers to the array number and the position of the spacer in the array; the spacer and protospacer sequences are aligned and mismatches underlined; % is the percentage of mismatch between spacer and protospacer sequences; the invading elements are indicated and position of the hit in the genome indicated if relevant in parentheses; potential 5′ PAM are indicated in the last column based on each match.

#	Spacer – protospacer sequences	%	Invading element	5′ PAM
4-27	GCACCTAAGTAATCTATAAGTAAAGTATCAACTCTT GCACCTAAGTAATCTATAAGTAAAGTATCAACTCTT	0	*Clostridium ljungdahlii* DSM 13528 (phage protein)	CTTCA
4-27	GCACCTAAGTAATCTATAAGTAAAGTATCAACTCTT GCACCTAAATAATCTATAAATAAAGTTTCAACTCTT	8	incomplete phage PHAGE_Staphy_PT1028_NC_007045	CATCA
4-8	CCTCATATTCTTTTAATAATTTATAAAATGTATTT - CTCATATTCTTTTATTATTTTATAAAATGTATTT	9	*Bacillus cereus* strain FORC087 plasmid pFORC087.2	GTGTT
4-7	TTTCTCTTAT-AGATCCTTTAATTCTTCTATAGCTTC TTTCTCTTATTAAACCCTTTAATTCTTCTAAATCTTC	11	intact phage PHAGE_Clostr_phiCT19406C_NC_029006	TTCCC
4-8	CCTCATATTCTTTTAATAATTTATAAAATGTATTT CCTCATATTCTTTTATTATTTTATAGAAAGTATTT	11	*Bacillus thuringiensis* strain YGd22-03 plasmid pYGD30 (recombinase gene)	TTTTA
4-17	TCTTTTGATATGGCCCCTGCATGAGAAAGTTGTTG TCTTTTGAAATGACACCTGCATGAGAAAGCTGTTG	11	intact phage PHAGE_Clostr_phiCT9441A_NC_029022	TTGCA
4-10	AATACATCTATATGCTTAATGCATGTAGTATCTATAC AATATATCTATATGCTTAATACATGTAGTATCTA---	14	*Clostridium pasteurianum* BC1 (phage portal protein)	TAGAG
3-21	TTTATACATTCCTCCATTTTTATACTAATTATTTAA TTTTTACATTCCTCCAATTTTTTATTTATTATTTAA	14	intact phage PHAGE_Clostr_phiCTC2B_NC_030951	ACTCA
3-22	TGGAAAGGAAGTATA-ATATGATAGATGAGATATTAA -GGAAAGGATGTATAAATATGATAGATAAAATAATAA	14	intact phage PHAGE_Clostr_phiCD27_NC_011398	ATAAA

3-25	CTTACAGGCGTATGTCCAAGATCATTTACTGCCTCC CTTACAGGAGCATGTCCAAGGTCATTCACTGCTTCC	14	incomplete phage PHAGE_Clostr_phiCTC2A_NC_030949	CTTCA
4-27	GCACCTAAGTAATCTATAAGTAAAGTATCAACTCTT GCACCTAAATAATCTATAAATAATGTGTCAACTCT-	14	questionable phage PHAGE_Clostr_vB_CpeS_CP51_NC_021325 (virulence-associated protein E)	CTTCA
4-28	GAATTTA-TATAAACCCAATCAGAATATATATAATTA GAATATACTATAAAACCAATTACAATAGATATAATTA	14	*Clostridium taeniosporum* strain 1/k plasmid pCt3	CTGAT
4-28	GAATTTA-TATAAACCCAATCAGAATATATATAATTA GAATATACTATAAAACCAATTACAATAGATATAATTA	14	*Clostridium taeniosporum* strain 1/k plasmid pCt3	CTGAT
4-6	TATATATCTATACCTTTTAACAATTTAAATACCTC - ATATATCTATTCCTTTTAAC-ATTTACATTCCTC	14	questionable phage PHAGE_Clostr_phiMMP04_NC_019422	ATTTG
4-8	CCTCATATTCTTTTAATAATTTATAAAATGTATTT CCTTATATTCTTTCAATAATTTATCATATGTTTTT	14	*Borrelia turicatae* strain BTE5EL plasmid lp159	ATATT

Supplementary Table 13: List of protospacer matches in invading elements from *C. autoethanogenum* spacers obtained with the Python script described in increasing mismatch percentage (between 15% and 20%). # refers to the array number and the position of the spacer in the array; the spacer and protospacer sequences are aligned and mismatches underlined; % is the percentage of mismatch between spacer and protospacer sequences; the invading elements are indicated and position of the hit in the genome indicated if relevant in parentheses; potential 5′ PAM are indicated in the last column based on each match.

#	Spacer – protospacer sequences	%	Invading element	5′ PAM
3-37	GTGGATTTAGGATCTTTTGTGAAGTTTCTATTTATAA -TGGATT-AGGCTATTTTGT--AGTTTCTATTTATAA	16	incomplete phage PHAGE_Staphy_PT1028_NC_007045	CAAAT
4-10	AATACATCTATATGCTTAATGCATGTAGTATCTATAC AATATATCTAAATGTTTTATACATGTAGTATCTATA-	16	*Clostridium botulinum* CDC_53174, (phage portal protein)	CTGCA
2-9	GAATACAGATACCAATCTCGGAACTGGAATAACTAA - AATACAGATACTAATATTGGAAGTGGAATAAGTAA	17	*Clostridium kluyveri* NBRC 12016 plasmid pCKL1	GATTA
2-20	GAAATTAAAAAGC---TTGAAGAACAAATTCTTGATAGC GAAAGTAAAAAGCCACTTGAAGAACAAATTATTGA----	17	*Mycoplasma salivarium* NCTC10113 plasmid 2	TTCTA
2-20	GAAATTAAAAAGCTT---GAAGAACAAATTCTTGATAGC GAAATTAAAAAGAAATATGAAGAAAAAATTCTTGAT---	17	*Spiroplasma citri* C5 plasmid pScpC5-3	GTACA
3-3	TATATGCCTGTAATTAATAAAATTACTGCTATTATT TATAT-CCT-TAATTAAAAAAATTAGTGCTATTA--	17	incomplete phage PHAGE_Lister_B054_NC_009813	ATTAT
3-7	AAAAAATGACTAATATAATATTATCTTATCCATTAT AAAAAATGACGAATATCATTTTAT-TTATACATTA-	17	*Candidatus Profftella armatura* YCPA plasmid	AATCA
3-10	TGATGTAATTATTCCTAATCCAGTATCTTTAGCACC TGATGTAAAAGTGCCTAATCCAGTAGTTTTAGCACC	17	intact phage PHAGE_Clostr_vB_CpeS_CP51_NC_021325	GTTAA
4-27	GCACCTAAGTAATCTATAAGTAAAGTATCAACTCTT GCACCTAGGTAATCTATTAATAAAGTGTTTACTCTT	17	Clostridium phage phiCT453A (virulence-associated protein E)	CTTCA

2-20	GTTTCTT--TTGTAGGAGTATTCTTTGTAGTGGTTGT -TTTCTTGGTTGTAGTAGTCTTCTTTGTAGTAGTT--	17	uncultured phage genome assembly, NCBI LR745210	GTAGA
4-6	TATATATCTATACCTTTTAACAATTTAAATACCTC -ATATATCTATTCCTTTTATCATTTTACATTCCTC	17	questionable phage PHAGE_Clostr_PhiS63_NC_017978	ATTTG
4-8	CCTCATATTCTTTTAATAATTTATAAAATGTATTT -CTCATATTCTTTCATTACCTTGTAAAATGTATTT	17	*Bacillus pseudomycoides* BTZ plasmid pBTZ_1	TGATT
4-10	AATACATCTATATGCTTAATGCATGTAGTATCTATAC AATAAATCAATGTGCTTAATGGCTTTAGTATCTATA	19	intact phage PHAGE_Lister_2389_NC_003291	CATCT
3-1	GCGAATAGTACCTTTGGTTTTTGCTCCTGGAGCTGC GCAAATATTACTTTTGGCTTTTGCTCCTGAAGTTG-	19	*Clostridium kluyveri* DSM 555 (predicted prophage)	TTTCT
4-7	TTTCTCTTATAGATCCTTTAATTCTTCTATAGCTTC TTTGTCTTAATGATCCTTTAATTCTTTTGTAGCT--	19	intact phage PHAGE_Clostr_vB_CpeS_CP51_NC_021325	TTATT
4-27	GCACCTAAGTAATCTATAAGTAAAGTATCAACTCTT GCGCCTAAGTAATCTATAAGTAAGGTATCAA-----	19	*Clostridium scatologenes* ATCC 25775 (phage-like protein)	CCTCA
4-27	GCACCTAAGTAATCTATAAGTAAAGTATCAACTCTT GCACCTAAATAGTCTATAAATAAAGTACTTACTCT-	19	intact phage PHAGE_Clostr_vB_CpeS_CP51_NC_021325	CTTCC

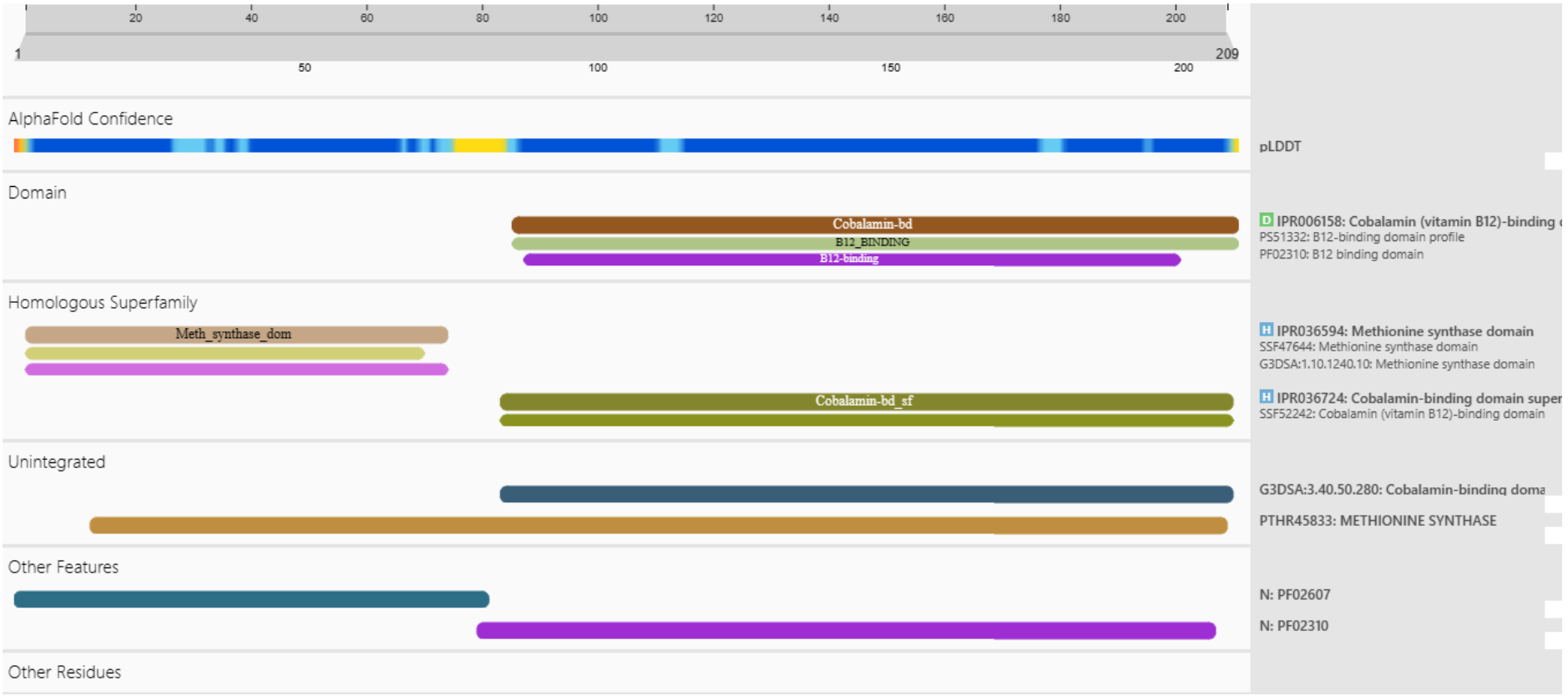
Supplementary Table 14: Genes with highest fold change of abundance superior during fructose fermentation. Fold change of abundance was calculated between the first and last sample of the fructose fermenter, genes with fold change above 4 are listed.

Locus tag	Gene	Description	Fold change
Awo_c08080	-	transcriptional regulator GntR family	61.70
Awo_c08070	*mttC7*	corrinoid protein MttC7	45.79
Awo_c26530	-	putative HIT family protein	13.41
Awo_c07710	-	serine protease	12.84
Awo_c35480	-	signal transduction phosphoesterase YybT-like protein	12.00
Awo_c24550	*rpoN*	RNA polymerase sigma-54 factor RpoN	11.97
Awo_c07560	*opuD3*	choline/carnitine/betaine transporter OpuD3	11.75
Awo_c29720	-	transcriptional regulator	10.88
Awo_c35490	-	hypothetical protein	10.65
Awo_c07510	-	sigma-54 dependent DNA-binding response regulator	9.60
Awo_c09360	-	hypothetical protein UPF0182	8.77
Awo_c07550	*mttC6*	corrinoid protein MttC6	8.48
Awo_c07540	*mttB10*	methyltransferase 1 MttB10	6.97
Awo_c09860	-	hypothetical protein DUF1292	6.91
Awo_c07520	*mttA2*	methyltransferase 2 MttA2	6.31
Awo_c34330	-	transcriptional regulator MarR family	6.02
Awo_c10230	*cbiJ1*	precorrin-6A reductase CbiJ1	4.55
Awo_c14610	-	hydroxyindole O-methyltransferase	4.54
Awo_c13070	-	methyltransferase	4.51
Awo_c07530	*opuD2*	choline/carnitine/betaine transporter OpuD2	4.50
Awo_c19890	*trkH2*	potassium uptake protein TrkH2	4.33
Awo_c02510	-	putative membrane protein	4.16
Awo_c29970	-	putative peptidyl-prolyl cis-trans isomerase	4.06
Awo_c12800	*pyk*	pyruvate kinase Pyk	4.05

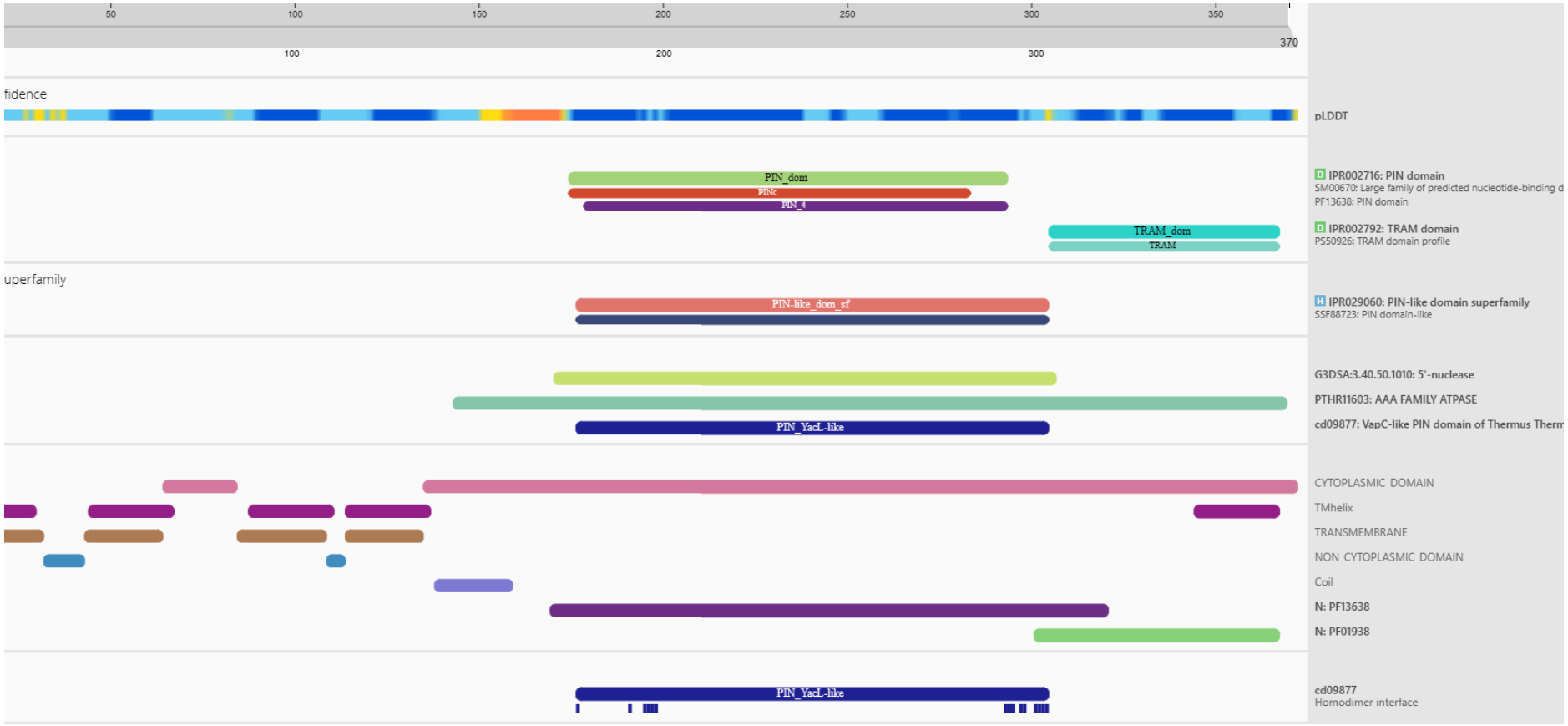
Supplementary Table 15: Genes with highest fold change of abundance superior during H_2/CO_2 fermentation. Fold change of abundance was calculated between the first and last sample of the H_2/CO_2 fermenter, genes with fold change above 4 are listed.

Locus tag	Gene	Description	Fold change
Awo_c03590	-	PilT-like protein	91.41
Awo_c00050	*gyrB*	DNA gyrase subunit B	70.64
Awo_c00040	*recF*	DNA replication and repair protein RecF	49.60
Awo_c14190	-	glutaredoxin	41.23
Awo_c24360	*aldR*	AsnC family transcriptional regulator AldR	35.85
Awo_c24350	*ald*	alanine dehydrogenase Ald	33.31
Awo_c14200	-	hypothetical protein	33.05
Awo_c14460	-	DNA modification methyltransferase	25.08
Awo_c09920	-	peptidase U32	23.75
Awo_c28760	-	putative membrane protein	23.45
Awo_c18460	-	putative methyl-accepting chemotaxis sensory transducer	22.27
Awo_c23890	*vatD*	V-ATPase D-subunit VatD	22.23
Awo_c24370	-	hypothetical protein	20.03
Awo_c23930	*vatF*	V-ATPase F-subunit VatF	19.97
Awo_c23910	*vatA*	V-ATPase A-subunit VatA	15.54
Awo_c27000	*hydE*	sensory transduction histidine kinase HydE	14.43
Awo_c23900	*vatB*	V-ATPase B-subunit VatB	14.20
Awo_c23950	*vatI*	V-ATPase a-subunit VatI	13.77
Awo_c23970	*vatG*	V-ATPase G-subunit VatG	13.25
Awo_c23940	*vatK*	V-ATPase c-subunit VatK	12.20
Awo_c23920	*vatE*	V-ATPase E-subunit VatE	12.10
Awo_c23960	*vatC*	V-ATPase C-subunit VatC	11.94
Awo_c24830	*neuC*	N-acetylglucosamine-6-phosphate 2-epimerase/N-acetylglucosamine-6-phosphatase NeuC	11.36
Awo_c35490	-	hypothetical protein	10.96
Awo_c27010	*hydC*	iron hydrogenase HydC	10.59
Awo_c26510	-	cation-transporting ATPase	10.50
Awo_c26980	*hydB*	iron hydrogenase HydB	9.30
Awo_c26990	*hydD*	iron hydrogenase HydD	8.63

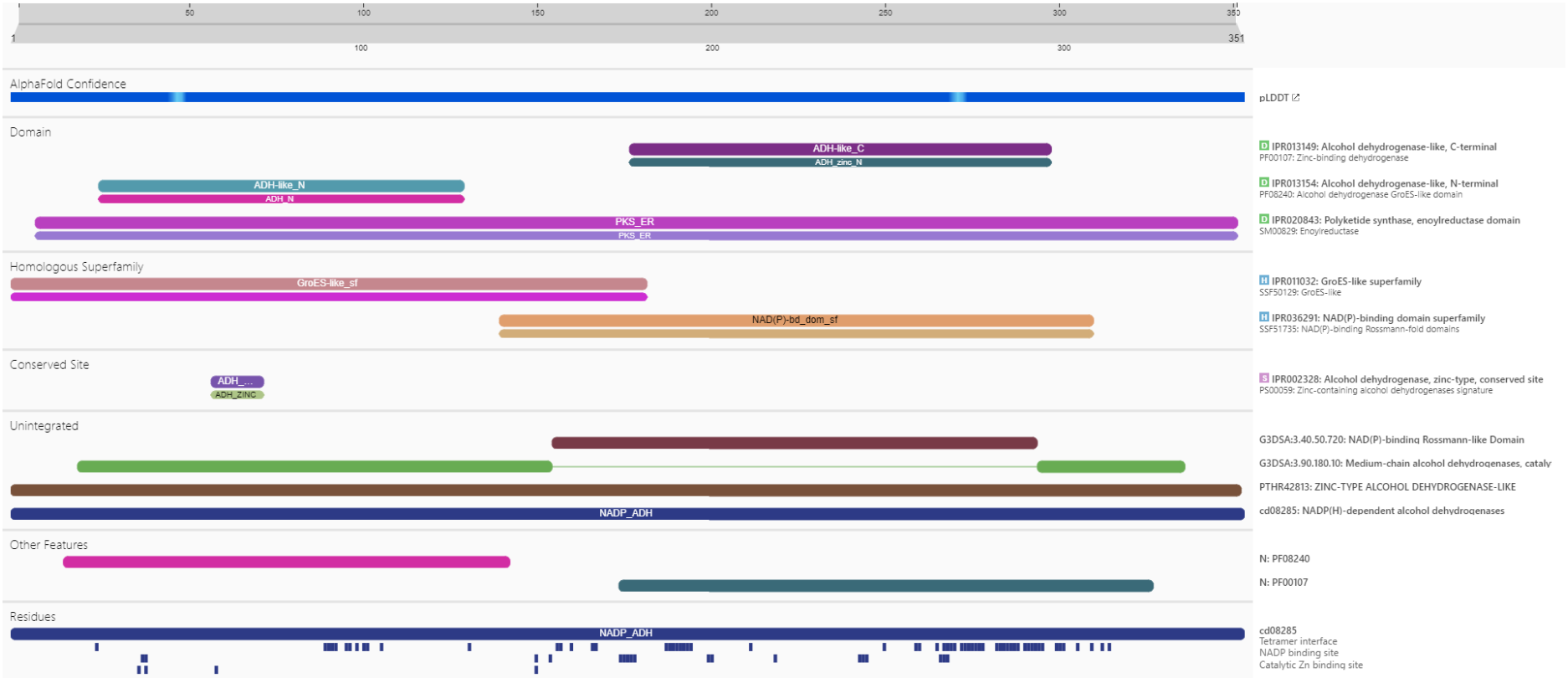
Awo_c20000	-	putative RNA-binding protein	8.26
Awo_c26970	*hydA1*	iron hydrogenase HydA1	8.16
Awo_c21740	-	putative sigma E factor regulator	7.29
Awo_c17420	-	putative iron-sulfur flavoprotein	7.09
Awo_c24840	*neuB*	N-acetylneuraminate synthase NeuB	6.76
Awo_c09330	*gcvH1*	glycine cleavage system H protein GcvH1	6.28
Awo_c24850	-	acetyltransferase GNAT family	6.17
Awo_c09910	-	O-methyltransferase family 3	5.99
Awo_c24870	*rffA2*	lipopolysaccharide biosynthesis protein RffA2	5.99
Awo_c00090	*gatB1*	galactitol-specific phosphotransferase enzyme IIB component	5.88
Awo_c06380	*bcaT1*	branched-chain amino acid aminotransferase BcaT1	5.78
Awo_c09260	*fhs1*	formyl-THF synthetase Fhs1	5.72
Awo_c13070	-	methyltransferase	5.63
Awo_c10910	*folBK*	dihydroneopterin aldolase/ 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase FolBK	5.55
Awo_c10930	-	metal-dependent phosphohydrolase	5.36
Awo_c26530	-	putative HIT family protein	5.23
Awo_c10920	*folP*	dihydropteroate synthase FolP	5.16
Awo_c06540	-	endoribonuclease	5.03
Awo_c06020	-	hypothetical protein membrane protein	4.68
Awo_c03380	-	hypothetical protein	4.59
Awo_c14240	*moaC*	molybdenum cofactor biosynthesis protein C	4.28
Awo_c26540	-	putative oxidoreductase	4.02



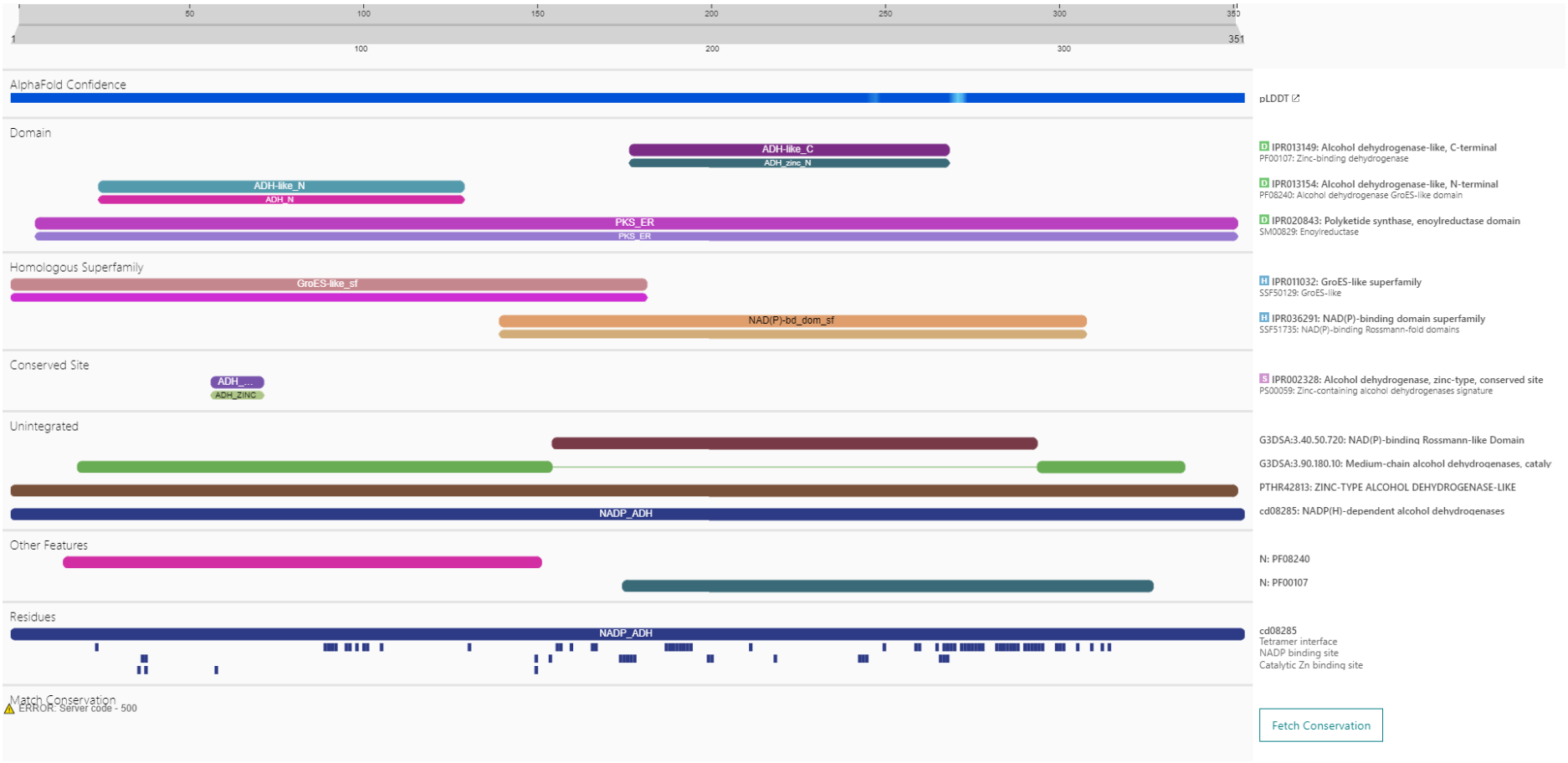
Supplementary Figure 3: InterPro data for *A. woodii* gene *mttC7*, protein H6LB49



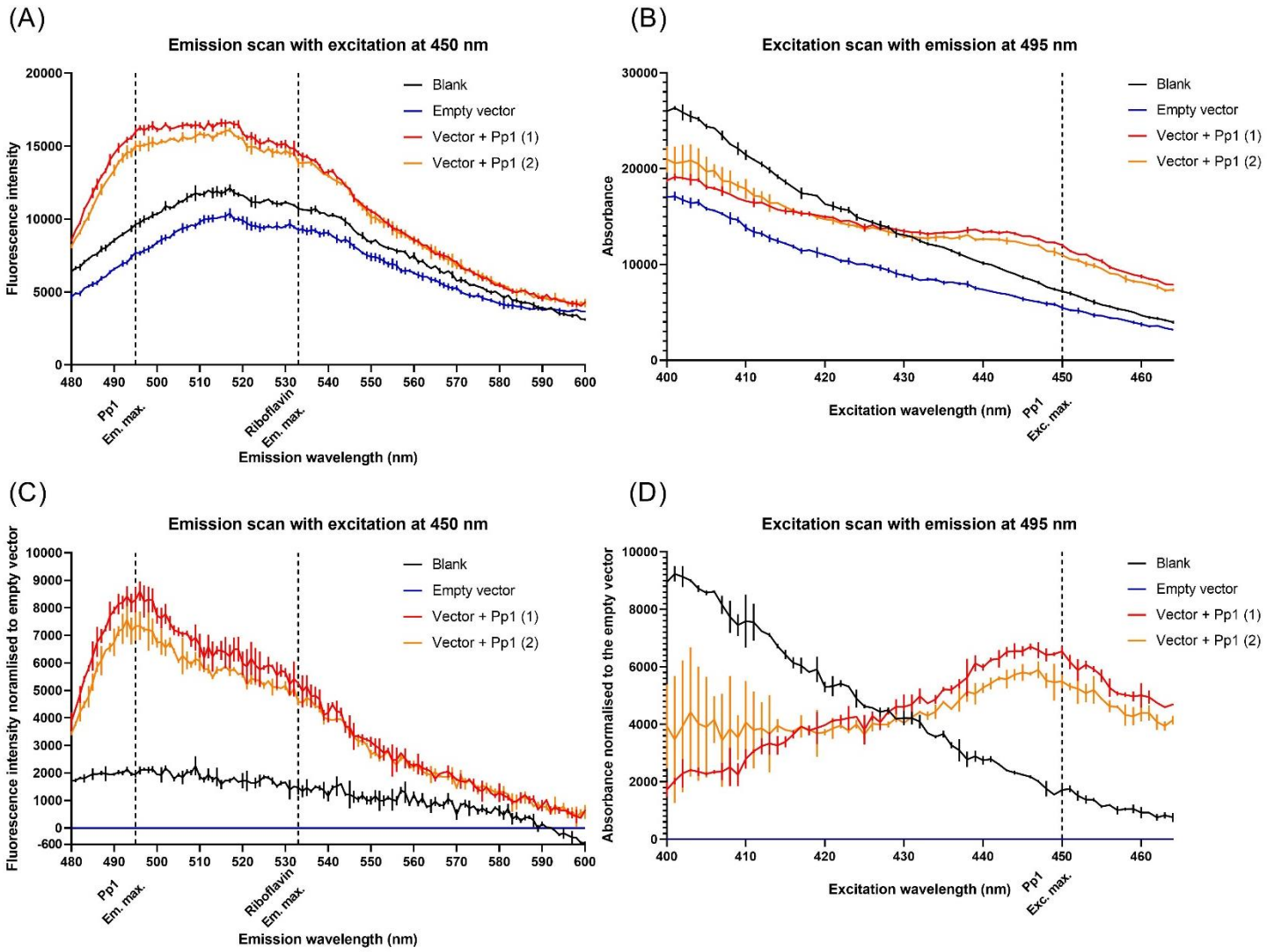
Supplementary Figure 4: InterPro data for *A. woodii* gene Awo_c03590, protein H6LH22



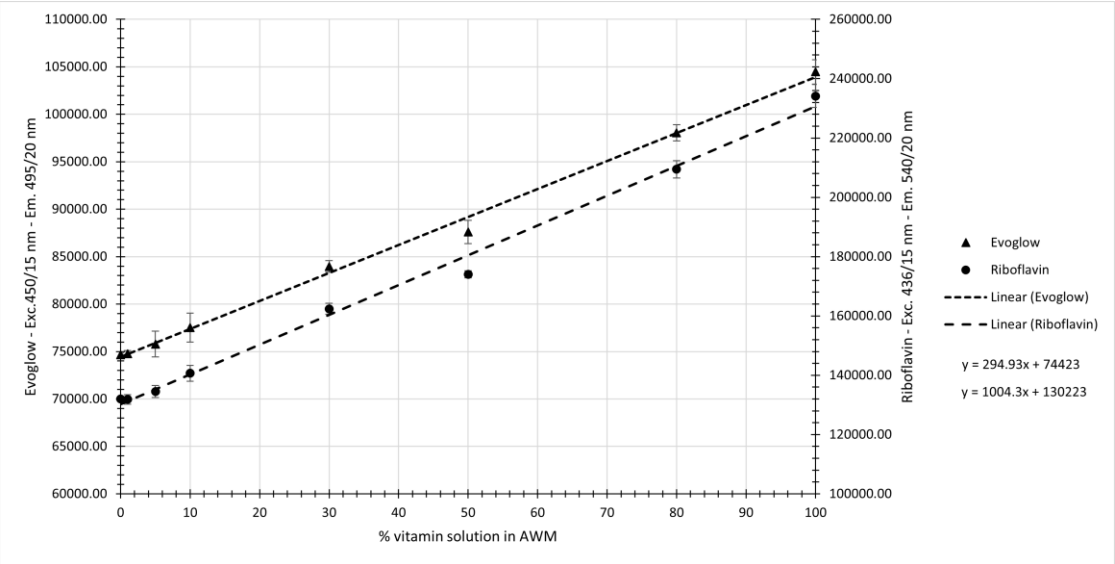
Supplementary Figure 5: InterPro data for *C. autoethanogenum* SADH protein, U5RQB1



Supplementary Figure 6: InterPro data for *C. beijerinckii* SADH protein, P25984



Supplementary Figure 7: Emission and excitation scan of *A. woodii* strains containing the Pp1 protein expressed with the P_{thl} promoter. Represented are two strains containing the fluorescent gene (plasmid pMTL84152_Pp1) (in red and orange) and controls with the medium only (in black) or *A. woodii* with the empty vector (plasmid pMTL84152) (in blue). Emission scan after excitation at 450 nm, intact (A) or after subtraction of the plasmid control (C). Excitation scan with emission at 495 nm, intact (B) or after subtraction of the plasmid control (D)



Supplementary Figure 8: Standard curve of fluorescence in relation to the percentage of vitamin solution present in the AWM medium. Measurement of the fluorescence at the wavelengths corresponding to the Evoglow genes (Exc.450/15 nm - Em. 495/20 nm) with a gain of 2198 and at the wavelengths corresponding to riboflavin (Exc. 436/15 nm - Em. 540/20 nm) with a gain of 2375. Both set of fluorescence measurements follow a linear trend showing increasing fluorescence with increasing vitamin concentration.