Metabolic engineering of *Cupriavidus necator* H16 for sustainable biomanufacturing

Simona Della Valle

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2023

‘*Well, I could have been wrong’, I said.*
‘And one way to know is to find out.’

- From *Cassandra at the Wedding,*
by Dorothy Baker

Abstract

Uncoupling fuel and chemical production from the consumption of fossil resources is a fundamental challenge for sustainable development. In industrial biomanufacturing, this challenge is most often addressed by using heterotrophic microbes to convert plant biomass into value-added compounds. To improve the sustainability of industrial bioprocesses, a shift to alternative feedstocks is required. This includes the direct utilisation of excess atmospheric carbon dioxide (CO_2), as well as other one carbon (C_1) or two-carbon (C_2) compounds which can be derived from abiotic CO_2 reduction or alternative waste streams. As these feedstocks are broadly inaccessible to common industrial microbes, their bioconversion requires engineering specialised production strains.

The facultative chemolithoautotroph *Cupriavidus necator* H16 (*C. necator*) holds demonstrable potential as a host strain for sustainable biomanufacturing, owing to its malleable metabolism and natural ability to utilise a wide range of feedstocks. To realise its potential, tailored molecular tools and extensive biological insights are needed. Addressing these requirements is the overarching aim of this thesis. A novel electroporation protocol was optimised, allowing for more rapid and efficient transformation of wildtype *C. necator* than previously attainable. This advancement, which accelerates the prototyping of synthetic DNA constructs, enabled the development of an efficient genome editing tool. The method uses homologous recombination and CRISPR-Cas counterselection to streamline the genomic manipulation of *C. necator*, facilitating the installation of new or enhanced cellular functions. During the development of this genome editing tool, the performance of several different genetic regulatory modules was characterised. These DNA parts are applicable in other biotechnological contexts, enabling the implementation of more complex genetic circuitry in *C. necator*. As an example, the implementation of a genetic logic gate was demonstrated.

Strategies to improve growth and production from C_1 and C_2 substrates were also investigated. An evolutionary engineering approach was implemented to improve the growth rate of *C. necator* on formate, leading to the identification of evolved isolates with enhanced phenotypic traits. Genotypic analysis of these isolates revealed likely causative mutations, which could be harnessed for further metabolic engineering. Finally, the bioconversion of acetate to ethanol was investigated. Ethanologenic *C. necator* strains were constructed, recording improvements in titer, rate and yield over previous reports. Taken together, the findings documented in this thesis provide valuable tools, methods and insights relevant to *C. necator* engineering, thereby contributing to its development as a platform strain for sustainable biomanufacturing.

Acknowledgments

At the time of submission, it is entirely evident to me that none of the work presented in this thesis would have been possible without the unwavering support of my colleagues, friends, and family. Firstly, I would like to thank my supervisors, Prof Wei Huang, Prof Harrison Steel and Prof Ian Thompson, for their guidance and mentorship over the course of the DPhil. Above all, thank you for always encouraging me to pursue my scientific interests with boundless curiosity, and for helping me draw the bounds when I could not see them clearly for myself.

To all my friends and colleagues in the SynBio CDT, at Begbroke and in the Engineering Biotechnology group, working alongside you for the past few years has been a great pleasure. I will always cherish our scientific discussions, but also our countless lunches, formal dinners, pub outings, and of course, our cathartic rants. Pursuing a DPhil is never an easy thing to do, but even less so against the unstable backdrop of a global pandemic. Without our community, camaraderie, and friendship, I am certain that I would not have made it to this stage. Thank you for sharing your journeys with me, too.

I would also like to thank all my friends and colleagues from my time at UCL. To my biochemist friends: I am so proud of everything we have achieved since we graduated. Even though we have since scattered to the winds, and it's not always easy to keep in touch, our friendships will always ground me and inspire me. To Dr Cameron Watson and Dr Sarah Jones, thank you for your continued mentorship and support throughout my academic career. You have never failed to provide perspective when I needed it. You were also the first people to show me how to make buffers, grow bacteria, take

care of “labmin”, and persevere through lab work setbacks back in 2016, which has come in very handy. To Prof Saul Purton, who supervised my Master’s work, thank you for supporting me early on in my development as a scientist. I remember my time in the Purton Lab most fondly; a bright starting point in my training.

Beyond my academic and professional environment, I would like to thank the strong women of the Oxford University Basketball Club and the Linacre Ladies that Lift Society. Training and competing with you all over the past few years has been immensely joyful. And of course, thank you to my original basketball teammate, my oldest friend Carlos. Thank you for always encouraging me to be the truest version of myself.

To my loving parents Rossana and Nicola, and my formidable sister Elena, thank you for always believing in me and supporting me as I strive to find my own path in the world, however erratically. A tree is only ever as strong as its roots.

To Lauren: I could have filled every single double-spaced line of this document with statements of gratitude, and it still would not have been enough. I have learnt a lot of things during my DPhil, but none of it even remotely comes close to the magnitude and importance of everything I have learnt from you.

Declaration of Authorship

I hereby declare that, unless otherwise stated, this thesis and the entirety of the research it contains is wholly my own work. I certify that, where applicable, the rightful authors of any collaborative work are duly acknowledged. I confirm that no part of this thesis has been submitted as part of any other degree at this or any other University.

Simona Della Valle

20th July 2023

Table of Contents

Part 2: Improving growth and production from C_1 and C_2 feedstocks

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 6

List of Abbreviations

ALE	Adaptive laboratory evolution
ATP	Adenosine triphosphate
C_1	One-carbon
C_2	Two-carbon
CBB	Calvin-Benson-Bassham
CDS	Coding DNA sequence
CFU	Colony forming units
Chl^R	Chloramphenicol resistance
cPCR	Colony PCR
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
DBTL	Design-Build-Test-Learn
dCas9	Nuclease-deficient Cas9
DIY	Do it yourself
DNA	Deoxyribonucleic acid
GC-FID	Gas-chromatography with flame ionisation detection
GFP	Green fluorescent protein
GOI	Gene of interest
gRNA	Guide RNA
HR	Homologous recombination
IS	Insertion sequence
JEx	Jungle Express
Kan^R	Kanamycin resistance
KO	Knock-out
LB	Luria-Bertani medium
mRFP	Monomeric red fluorescent protein
mRNA	Messenger RNA
NAD/H	Nicotinamide adenine dinucleotide (NAD) + hydrogen (H)
NalX	Nalidixic acid
OD	Optical density
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
PHA	Polyhydroxyalkanoate
PHB	Polyhydroxybutyrate
RBS	Ribosome binding site
rGly	Reductive glycine
RNA	Ribonucleic acid
SBOL	Synthetic Biology Open Language
SEVA	Standard European Vector Architecture

SIBR	Self-splicing intron-based riboswitch
SOB	Super Optimal Broth medium
SOF	Super Optimal Fructose medium
Tet^R	Tetracycline resistance
THF	Tetrahydrofolate
TIR	Translation initiation rate
UV	Ultraviolet
WT	Wildtype

1. General introduction

1.1. Thesis scope and objectives

The overarching aim of the research presented in this thesis is to advance the development of *Cupriavidus necator* H16 (hereafter referred to as *C. necator*) as a platform strain for environmentally- and economically-sustainable biomanufacturing. This facultatively chemolithoautotrophic bacterium holds demonstrable potential to produce value-added chemicals from renewable feedstocks at an industrial scale. However, there are several outstanding limitations for its application in bioproduction. The work presented in this thesis addresses these challenges at three different levels, with the following specific objectives:

i. To design and build tailored tools for advanced genetic manipulation, enabling more ambitious strain engineering.

ii. To engineer strains with improved growth on formate (C_1 substrate), which can be applied to increase the efficiency of C_1 bioprocesses.

iii. To demonstrate ethanol production from acetate (C_2 substrate) as a case study in the valorisation of renewably generated feedstocks.

1.2. Thesis layout

The current chapter provides an overview of the thesis and an introductory literature review (Section 1.3). Following this, the thesis' research contributions are presented, divided into two parts and five chapters (Figure 1.1), as specified below. Finally, Chapter 7 provides a general assessment of the thesis' main outcomes and limitations, as well as opportunities for future research.

Part 1: Molecular tools and methods for *C. necator* engineering

- **Chapter 2** describes the development of an optimised electroporation protocol, which both simplifies and expedites the delivery of recombinant plasmids into *C. necator*, and will enable the efficient transformation of large DNA libraries.

- **Chapter 3** outlines the tailoring of a novel CRISPR-based tool for genome editing in this bacterium, enabling highly efficient genomic manipulation on a reduced timeline.

- In **Chapter 4**, the regulatory module used to build the CRISPR-based tool was harnessed to implement a two-input AND gate, further validating its utility, and demonstrating how separate modules for the regulation of gene expression can be combined to deliver more complex genetic circuitry in *C. necator*.

Part 2: Improving growth and production from C_1 and C_2 feedstocks

- **Chapter 5** outlines the calibration, successes, and limitations of an adaptive laboratory evolution (ALE) experiment, where *C. necator* was evolved towards improved growth on formate.

- Finally, **Chapter 6** describes the implementation of a two-enzyme biosynthetic pathway, endowing *C. necator* with the ability to produce ethanol from acetate at the gram-per-litre scale.

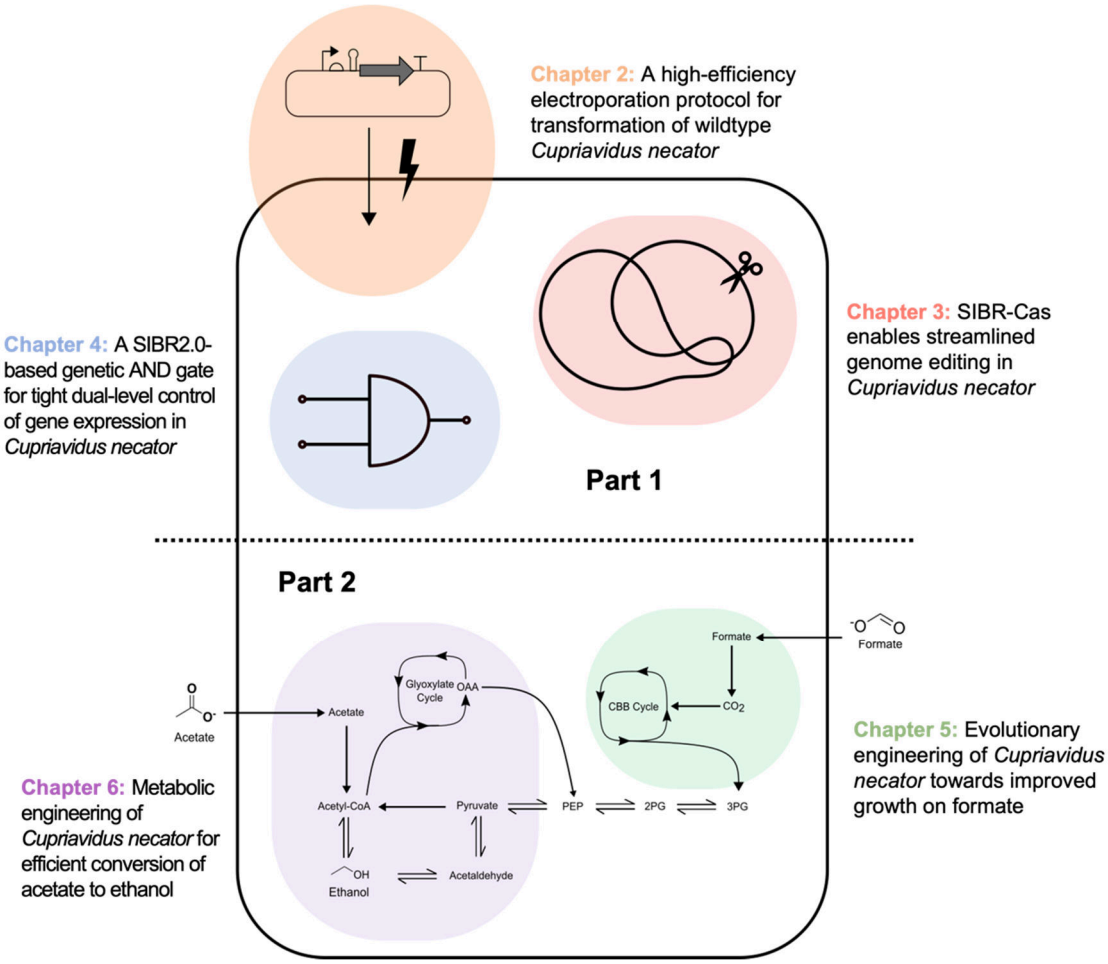


Figure 1. 1. Project Overview.

The focus areas of each Chapter are depicted, encompassing plasmid delivery, genome editing, advanced genetic circuitry, engineering formatotrophic metabolism, and harnessing acetotrophic metabolism for ethanol biosynthesis.

1.3. Background

1.3.1. The need for alternative substrates in industrial biotechnology

The accumulation of excess anthropogenic carbon in the environment is driving rapid change in planetary conditions, with adverse long-term consequences for both human society and biodiversity. Amongst other factors, this change is caused by the unsustainable use of fossil feedstocks for the manufacturing of chemicals and fuels, an entrenched process underlying much of modern society, with roots dating back to the Industrial Revolution[1].

Industrial biomanufacturing, which harnesses biological organisms as catalysts for the production of commodity and specialty chemicals, is a promising alternative to address these global challenges and diminish (or eliminate) our dependence on petrochemical-based manufacturing[2]. Traditionally, industrial bioprocesses have replaced fossil feedstocks with agricultural feedstocks. A prime example of this is the bio-based production of fuel ethanol from corn or sugarcane via microbial fermentation of ethanologenic yeasts[3] (predominantly *S. cerevisiae*) or bacteria[4] (most commonly *Zymomonas mobilis*, *Klebsiella oxytoca* and *Escherichia coli*). Bioethanol can be used as either an additive or replacement for petroleum-derived fuels. It is the most widely used biofuel, with a total global production of ~27 billion gallons a year as of 2021[5]. Bioethanol production exemplifies how a bioprocess can evolve from academic conception to industrial deployment when facilitated by a favourable economic, legislative and socio-political landscape[6].

The agricultural feedstocks for bio-based production of ethanol and products vary between first- and second-generation processes. First generation feedstocks include sugars, starch and vegetable oil derived from plant biomass. Second generation feedstocks include compounds derived from non-food lignocellulosic plant waste[7]. Though more sustainable than the petrochemical alternative, the green credentials of both first- and second-generation feedstocks are widely disputed. It has been argued that the large-scale cultivation of bioenergy crops strains the availability of limited resources such as arable land, freshwater, and fertilisers, so is a net detriment to the physical environment, biodiversity, and global food security[7]. Additionally, the efficiency of bioproduction processes based on using plant biomass as a feedstock is inherently limited by the low energetic efficiency of photosynthesis[8–11]. In response to this, metabolic engineering strategies have focused on implementing 'feedstock flexibility' within traditional biocatalytic strains, to enable the use of a wider and more sustainable range of substrates in industrial biomanufacturing, such as third generation feedstocks[12].

As an alternative, the possibility of engineering biology to construct microbial cell factories that can directly sequester and valorise waste carbon streams, particularly excess atmospheric carbon dioxide (CO_2) and other one-carbon (C_1) feedstocks, is steadily gaining prominence in both academic and industrial research arenas[13,14]. Recent studies have highlighted how autotrophic microbial hosts can deliver sustainable production of a variety of molecules from renewable C_1 feedstocks, thereby diminishing our reliance on both petrochemical processing and the use of plant

biomass (see Section 1.4). In this introductory review, a brief overview of the foundational technologies enabling the engineering of advanced microbial production strains is first provided. This is followed by a discussion of the general considerations and recent advances in the construction of microbial platform chassis for the valorisation of CO_2 and other C_1 feedstocks.

1.3.2. Synthetic biology as an enabling discipline for sustainable biomanufacturing

The prospect of engineering microbes as platforms for the production of value-added chemicals was first conceived following the development of recombinant DNA technology in the 1970s-1980s[15,16]. An early example of microbial biotechnology was the production of biosynthetic human insulin in *E. coli*, first demonstrated in 1978 by scientists at Genentech[17]. Since the early days of recombinant DNA technology, molecular tools for the manipulation of microorganisms have been in constant evolution. Building on these technical advances in the fields of molecular and systems biology, such as the development of high-throughput “omics” methods (genomics, transcriptomics, proteomics, etc.), over the past two decades synthetic biology has been cemented as a paradigm for biological engineering[18,19] (Figure 1.2).Though synthetic biology is now a broad church, encompassing different research areas, methodologies, and applications, its core notion is that engineering principles can be applied to both study and manipulate biological organisms. Below, we describe key developments catalysed by synthetic biology, which have advanced our ability to precisely engineer biology.

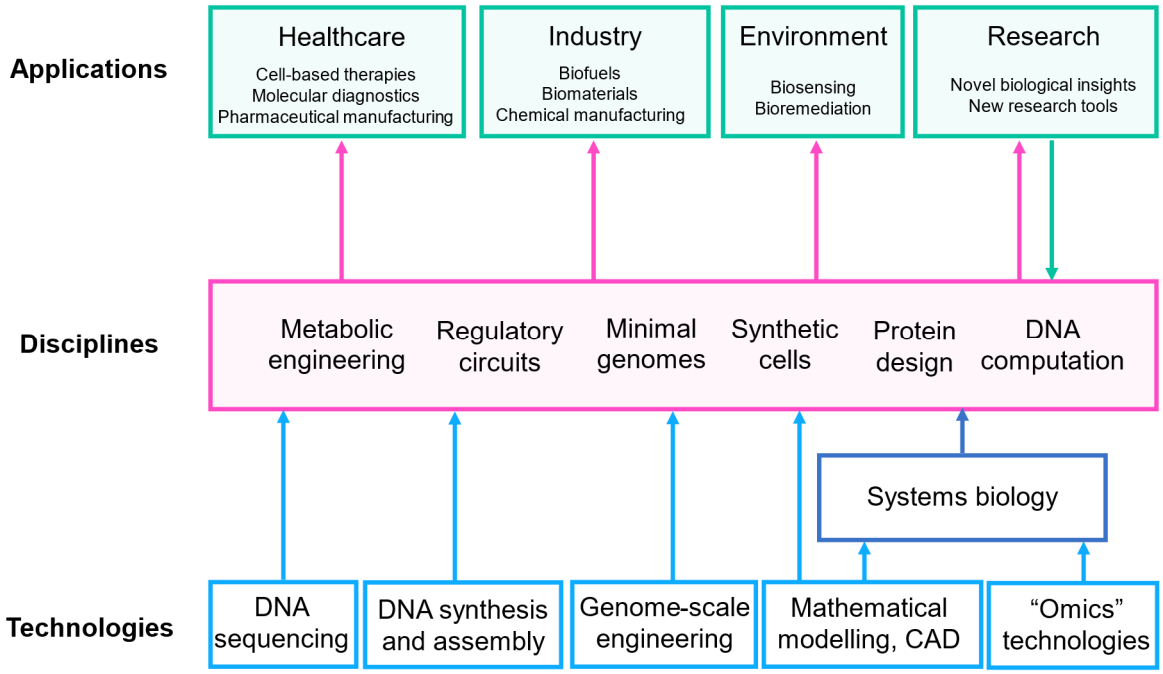


Figure 1. 2. Overview of synthetic biology.

The ongoing development of synthetic biology is underpinned by several technical advancements in molecular technologies, as well quantitative knowledge derived from systems biology. Synthetic biology approaches are incorporated into many disciplines, including metabolic engineering and the design of artificial regulatory circuits, which are discussed in this introductory review. Outputs from these disciplines can be applied in the healthcare sector, industrial biomanufacturing, or environmental monitoring and remediation. Additionally, they can provide tools and insights to inform further research. This figure has been adapted from ref. [20]. Abbreviations: CAD = computer aided design.

Modularity and standardisation of molecular parts enables the construction of regulatory networks

A core concept in synthetic biology is that regulatory networks operating within living cells can be viewed as hierarchical assemblies of modular units, similar to the interconnected architectures of electronic circuits. It should therefore be possible to engineer these networks by swapping or tuning their modular components. Novel molecular “parts”, such as synthetic DNA sequences, can also be designed and used to forward-engineer genetic networks. By implementing rewired or artificial regulatory architectures, host organisms can be endowed with optimised or new-to-nature capabilities, with countless applications in biotechnology and industrial biomanufacturing[21].

In an effort to establish the use of DNA parts as truly reusable modular components, synthetic biologists have adopted community standards for their design, construction, and characterisation[22]. For example, the Synthetic Biology Open Language (SBOL) was developed as a framework to unambiguously describe DNA designs and facilitate their exchange, as well as to improve the reproducibility of experimental results[23]. The SBOL standards include the use of a finite set of glyphs to represent the features of DNA sequences[24], including promoters, ribosome binding sites (RBSs), or terminators (Figure 1.3). To assemble DNA parts together and create new network architectures, standardised cloning methods are favoured, with Golden Gate assembly[25] as the preferred standard in the field. Numerous Golden Gate-compatible modular cloning toolkits have been developed, such as MoClo[26], CIDAR[27], or Start-Stop[28], all of which are readily available on Addgene[29]. This simultaneously facilitates the exchange of

DNA parts between researchers and enables their rapid assembly to form new genetic circuitry.

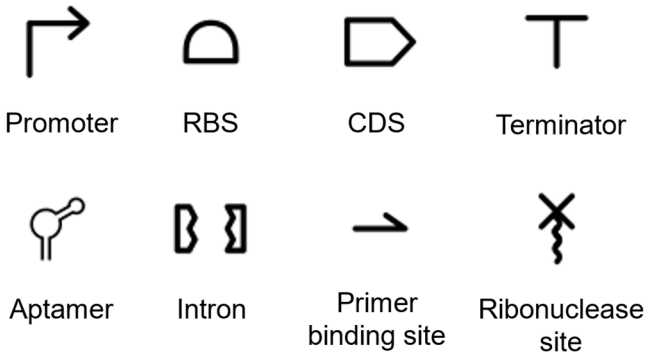


Figure 1. 3. SBOL graphical notation.

Selection of glyphs used to represent common sequence features when communicating DNA designs, as established by the SBOL visual standard, version 3.0[24]. SBOL symbols are used throughout this thesis. Abbreviations: RBS = ribosome binding site, CDS = coding DNA sequence.

Engineering resilience to biological complexity and diversity

By implementing assemblies of modular components, synthetic biologists have successfully built complex regulatory circuits. Examples include logic gates[30,31], oscillators[32], toggle switches (both as genetic networks[33] and protein networks[34]), and genetic clocks[35]. However, attaining these complex response functions almost always requires extensive calibration and optimisation of the engineered networks. This is because living cells are very different from computers, with many layers of biological complexity underlying their phenotypes. With this in mind, the reductionistic "plug-and-play" view of modular DNA parts and genetic circuits can be misleading[36,37]. For example, an inducible promoter part may yield very different gene expression outputs depending on its genetic surroundings (which gene is it regulating?) and genomic

context (which organism is it operating in?). This may be due to a wide range of factors, including but not limited to stimulatory or repressive interactions between sequence elements, long-range regulatory effects by transcription factors, the emergence of secondary structures in mRNA transcripts, or organism-specific biases in the recognition sequences for transcriptional and translational machinery. Further, when individual parts are combined into more complex assemblies, their behaviour is often unpredictable due to either unexpected interactions between molecular components, or between the engineered functions and the host cell's metabolism. Due to the challenges associated with implementing synthetic constructs into both model and non-model hosts, an iterative Design-Build-Test-Learn (DBTL) cycle is often used as a pipeline for optimisation (Figure 1.4).

Many design principles and practical methods can be implemented to tackle this complexity. For instance, molecular insulators can be used to reduce the influence of genetic context on the performance of synthetic parts or circuits[38,39]. Computational models can be harnessed to quantitatively predict the behaviour of regulatory networks, simplifying the forward-engineering of complex architectures[40]. Artificial genetic networks can be designed with the aim of being orthogonal to the host cell's metabolism, thereby minimising the impact of genomic context on circuit performance (though this effect is difficult to eliminate, as orthogonal circuits still largely rely on shared cellular resources for their operation)[41–44]. Additionally, feedback control systems can be implemented within living cells to limit the burden that engineered functions pose on host metabolism, thereby improving their stability[45].

In spite of the technical developments mentioned above, the challenges associated with ensuring the robust performance of synthetic functions are exacerbated when engineering non-model organisms, for which limited quantitative biological data is available. Advances in synthetic biology are often demonstrated and optimised in a narrow range of well-characterised, domesticated hosts such as *E. coli* or *S. cerevisiae*. Translating genetic designs to diverse hosts often involves extensive re-characterisation and optimisation (see ref. [46] for a recent example, where the CIDAR modular cloning toolkit[27] was modified to enable its use in a broader range of bacterial hosts). In some cases, the development of tailored molecular tools is also required[47].

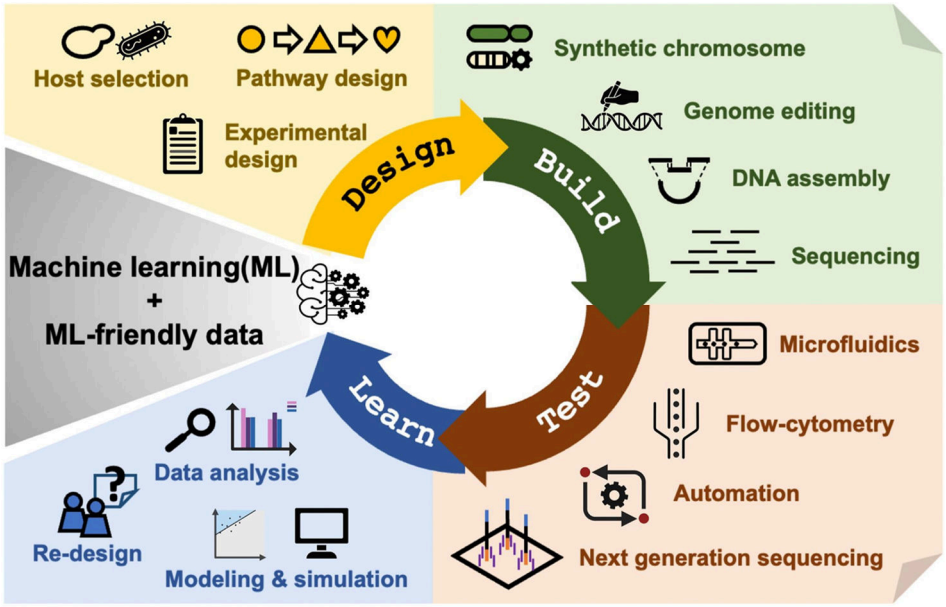


Figure 1. 4. Optimisation in synthetic biology.

The DBTL cycle is a productive pipeline to implement and optimise synthetic genetic designs. Key technical advances in molecular biology, automation, robotics, and computer science underpin each stage of the cycle. This figure was adapted from ref. [48].

Metabolic engineering

Synthetic biology principles can be applied in metabolic engineering to install or enhance biosynthetic pathways, as well as to optimise their performance within the

genomic and metabolic context of the host strain[49,50]. Rationally designed metabolic pathways can be constructed as synthetic “devices”, i.e. combinations of modular parts, and delivered into the host organism on episomal vectors. Broadly, devices used in metabolic engineering are designed with the goal of maximising metabolic flux towards a target product. Both endogenous and heterologous genes can be selectively expressed for this purpose, using synthetic or refactored regulatory modules.

Combinatorial approaches are often used to optimise the *in vivo* performance of such devices[51,52]. For instance, randomly assembling a set of biosynthetic genes with libraries of different promoters or ribosome binding sites, and screening for the best-performing constructs is an effective method to fine-tune the relative stoichiometries of enzymes along a metabolic pathway. Such combinatorial approaches amount to high-throughput transcriptional and translational engineering, allowing experimenters to define a pathway configuration for optimal production without the need for extensive *a priori* knowledge. This is facilitated by the availability of biophysical models that enable the design of synthetic parts such as promoters, RBSs, or CDSs with defined gene expression outputs (see refs. [53–56] for selected examples).

Other *in silico* methods, such as constraints-based modelling[57], are often used in metabolic engineering to facilitate the design and construction of platform strains with desirable phenotypes. In particular, there are several computational approaches that can harness (genome-scale) metabolic models to predict genetic interventions that lead to overproduction of a target chemical [58]. For instance, the OptKnock framework implements a bilevel optimisation algorithm that predicts gene deletion targets to couple chemical production with biomass formation, enabling the selection of

overproducing strains through genetic manipulation followed by growth adaptation[59]. Alternative tools such as OptReg or OptForce can be used to not only identify gene deletion targets, but also targets for other manipulations such as up- or down-regulations[60,61]. Other frameworks, such as OptStrain, can also predict the optimisation of overproducer strains through addition of heterologous reactions[62].

Genome engineering technologies

Implementing systems-level metabolic designs *in vivo* often requires genomic manipulation of the host strain. Tools for genome editing have become indispensable in metabolic engineering. In particular, the advent of CRISPR-Cas technologies over the past decade has greatly advanced our ability to precisely manipulate biological organisms[63].

CRISPR-Cas, which stands for clustered regularly interspaced short palindromic repeats and associated protein, are naturally-occurring components of prokaryotic adaptive immune systems. CRISPR arrays are genomic loci which contain short repetitive DNA sequences (known as CRISPR spacers)[64–66], some of which are of extrachromosomal origin, derived for instance from the genetic material carried by bacteriophages[67,68]. These sequences are kept as a record of past encounters, which protects the spacer-carrying strains from re-infection by enabling them to recognise foreign genetic material and inactivate it[69]. To do so, CRISPR spacers are paired with Cas proteins. Spacer sequences are transcribed and matured to form CRISPR RNA (crRNA) guides, the precise nature of which depends on the type of CRISPR-Cas system[70]. These RNA guides then form a complex with the Cas endonuclease,

directing it to a specific location on the target DNA (to which the crRNA is complementary) and thereby targeting it for cleavage. The target DNA location for the CRISPR-Cas system is not random; it is determined by a protospacer adjacent motif (PAM), the sequence of which varies between different types of CRISPR systems[71,72]. The biotechnological potential of CRISPR-Cas systems was first realised in 2012, when the Cas9 endonuclease and its associated RNA guide were successfully reprogrammed to cleave a plasmid *in vivo,* in a site-specific manner[73]. Since this ground-breaking report, Cas9 and other Cas nucleases (such as Cas12, formerly known as Cpf1) have been extensively developed as RNA-programmable tools for genome editing[74].

Gene knockouts can be performed by directing CRISPR-Cas cleavage to a specific target locus, generating blunt (Cas9) or staggered (Cas12) double-stranded DNA breaks. These can be subsequently repaired by error-prone non-homologous end joining[75], or by homologous recombination[76], for which a DNA template is provided. The latter repair mechanism can also mediate knock-ins[77]. Catalytically inactivate ("dead") variants of Cas nucleases (such as dCas9) can also be used to modulate gene expression in a process known as CRISPR interference (CRISPRi). The inactive Cas nuclease is directed to a target locus, interfering with the transcription machinery and resulting in gene silencing[78]. Conversely, in CRISPR activation (CRISPRa), inactive CRISPR-Cas systems are used to enhance transcription from a specific locus, for instance by fusing an RNA polymerase subunit to the Cas nuclease[79]. More recently, CRISPR-Cas systems have been re-engineered to mediate the precise editing of single DNA bases (base editing), as well as insertions and deletions (prime editing), without requiring double-stranded DNA breaks or donor DNA templates[80–82].

Though most advances in CRISPR-Cas genome editing methods have been reported in model bacterial species, toolkits tailored for use in non-model bacteria are becoming increasingly available[83]. By improving the genetic tractability of such microbes, these tools support their development towards advanced biotechnological applications.

Evolution presents challenges and opportunities

A fundamental property of all biological organisms is that, unlike machines, they evolve. When synthetic devices are deployed in a host cell, their sequence may be randomly altered due to errors arising during DNA replication. Often, cells in which these stochastic changes lead to the loss of the device's encoded function have a selective advantage, as they no longer need to allocate cellular resources to fulfil burdensome engineered functions[45,84]. Evolutionary stability is therefore a key consideration for synthetic biology and strain engineering, particularly when the long-term viability of genetic programs is required. Diverse strategies to improve the evolutionary stability of engineered functions have been demonstrated in the literature[85]. For example, targeted mutations can be introduced in the host strains' homologous recombination machinery, thereby limiting their ability to accumulate mutations via this particular cellular pathway. One such mutation, the deletion of the *recA* gene of *E. coli*[86], is a common genotypic feature of laboratory cloning strains such as *E. coli* DH5α. The removal of mobile genetic elements (such as transposons and insertion sequences) from a host strain's genome has also been shown to lower the potential for inactivating mutations to accumulate in engineered constructs[87]. Further to these strain engineering approaches, certain design principles can be implemented to yield synthetic devices with a reduced evolutionary potential. For

example, a study conducted using a GFP reporter system demonstrated that avoiding sequence repetition (i.e. the re-use of identical DNA parts) can reduce the accumulation of inactivating mutations[88].

Though the evolving nature of biological organisms often complicates their engineering, it may also be harnessed to optimise synthetic parts and devices, or to attain systems-level phenotypic changes that are inaccessible by rational design alone. For instance, adaptive laboratory evolution (ALE) is a widely-used tool that harnesses evolutionary dynamics to obtain strains with improved phenotypic traits. In ALE experiments, the processes of genetic diversification and natural selection are replicated at the bench scale. Microorganisms can be cultured in defined conditions for prolonged periods (in the range of hundreds to thousands of generations), creating the ideal conditions for the fixation of mutations that confer a fitness benefit in the specified environmental context[89]. Following this approach, strains can be evolved to have higher tolerance to a toxic substrate or product, increased nutrient uptake, and faster growth rates, all of which are key considerations for metabolic engineering and bioproduction[90,91]. Using genome engineering tools, mutations that are accrued during ALE experiments can be recapitulated in the ancestral wild-type strain, to determine causal relationships between individual (or combinations of) genotypic changes and the observed phenotypes. These insights can then be used to further improve rational strain designs (see ref. [92] for a recent example, which is discussed in more detail in Section 1.4.3 and Chapter 5).

Further to ALE experiments, evolution can be harnessed as a powerful tool to optimise imperfect metabolic designs. For example, growth-coupling designs which link a

bioproduction goal to biomass formation (such as those generated by the OptKnock tool[59]) can be optimised via growth adaptation. Engineered strains can be evolved towards improved growth, which simultaneously enables strains to accumulate nonintuitive mutations that further enhance their bioproduction capabilities. Evolutionary approaches, specifically directed evolution[93], can also be used to engineer or optimise the performance of individual parts, such as regulatory modules[94] or enzymes[95].

1.4. Engineering microbial platform hosts for CO_2 valorisation

In recent years, a range of synthetic and systems biology approaches have been applied to engineer microbes that can utilise CO_2 and other C_1 substrates for growth and bioproduction. Studies have focused on addressing the long list of limitations associated with autotrophic hosts. These include inefficient carbon assimilation, low biomass and product yields, a narrow spectrum of target molecules suitable for bioproduction, and the scarce availability of efficient tools for genetic manipulation of emerging autotrophic model organisms.

1.4.1. General considerations for autotrophic chassis engineering

Though a variety of engineering strategies have been demonstrated, three key sequential considerations for developing any platform chassis for CO_2 valorisation can be broadly defined (Figure 1.5.). Firstly, the optimal supply of both carbon and energy

sources must be tackled. Some microorganisms, such as acetogens in the genus *Clostridium*, are amenable to direct cultivation on carbon monoxide (CO)- and CO_2-rich waste gases, using hydrogen as an energy source in a process widely known as syngas fermentation[96–98]. Such gases can be cheaply derived from many sources such as steel mills or ethanol production plants, facilitating their implementation as cost-effective sustainable feedstocks[99].

Where direct cultivation on gaseous CO_2 is unfeasible or disadvantageous, a reduction-first approach has been proposed, in which CO_2 is initially reduced (for instance to formate) before being assimilated into the host cell's central carbon metabolism[100]. CO_2 reduction can be mediated intra- or extracellularly by native or engineered enzymes. Alternatively, efficient electrochemical (abiotic) methods of CO_2 reduction to C_1-C_3 compounds have been demonstrated[101,102]. The resulting compounds can be effectively used as electron mediators, and subsequently fed to microbes to support growth and synthesis of value-added compounds in a process known as electromicrobial production. In particular, the use of electrochemically-derived formate has gained considerable interest in recent years, as it has numerous advantages over direct cultivation on CO_2 or syngas mixtures[103–105]. Formate is miscible, non-flammable, easily stored and transported, and simultaneously provides a source of carbon and energy, thereby bypassing the need to use hydrogen gas as an energy carrier.

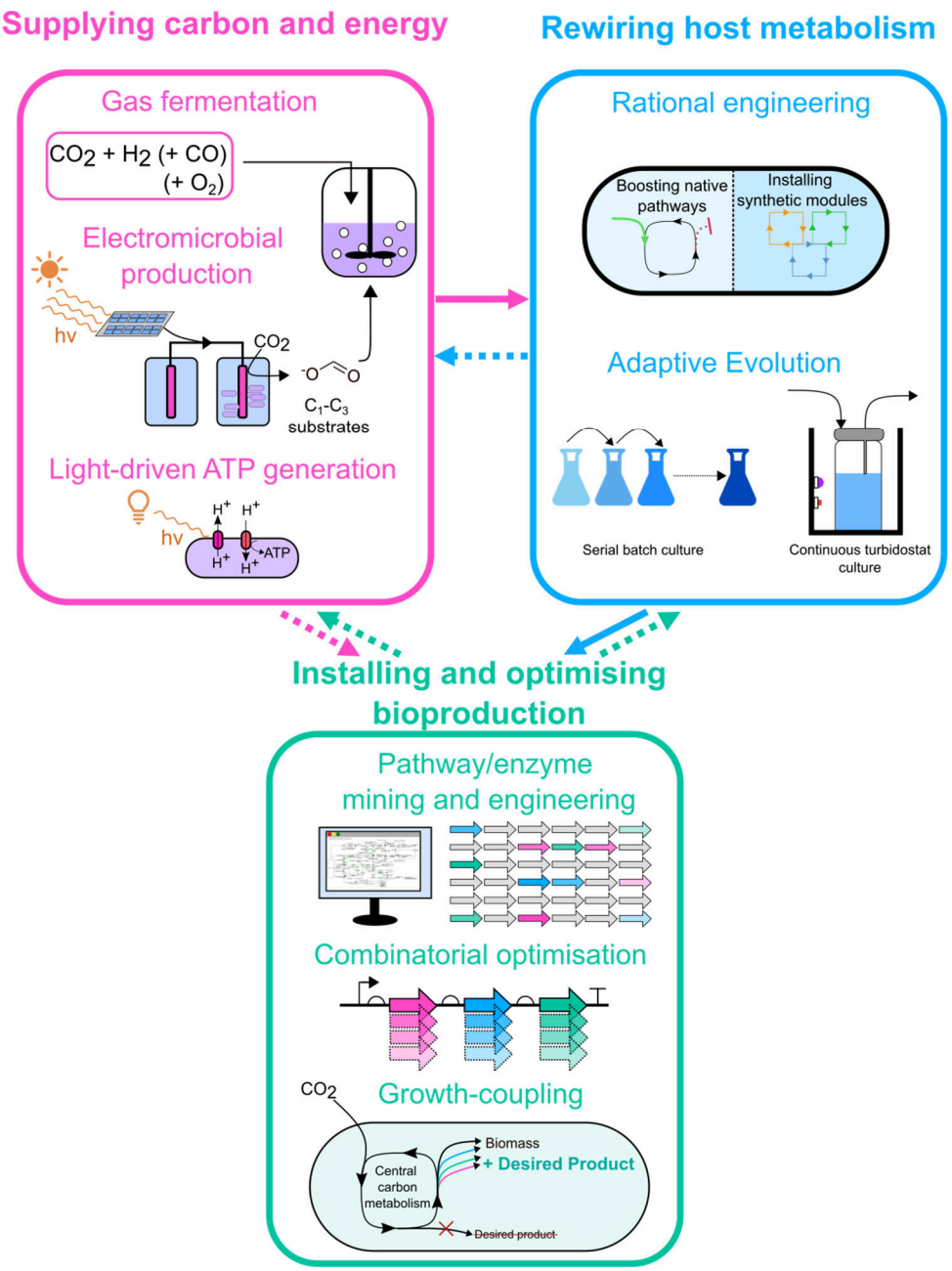


Figure 1. 5. Building microbial platform chassis for CO_2 valorisation.

Arrows indicate the interconnectivity between these three general considerations. Solid arrows correspond to the most common sequential flow of platform chassis engineering strategies; dashed arrows indicate that these optimisation strategies may be alternatively (and iteratively) connected. **Supplying carbon and energy.** Some chassis can be cultured directly on mixtures of gaseous substrates providing carbon (CO_2 and/or CO) and energy (H_2). In electromicrobial production, renewably-generated electricity can be used to power biohybrid autotrophic cultivation. Non-photosynthetic microbial hosts can also be endowed with light-harvesting capabilities to boost the supply of energy in autotrophic cultivation. **Rewiring host metabolism** can improve autotrophic growth

kinetics. This can be attained via rational engineering or adaptive evolution, with both strategies often being used in combination. **Installing and optimising bioproduction.** Biosynthetic pathways for both native and heterologous products can be established and optimised using a wide range of tools. Successful strategies have leveraged bioinformatic searches to find promising production pathways and enzymes, as well as experimental optimisation through combinatorial assemblies.

Secondly, the metabolism of the chassis is often rewired to enable autotrophic growth, or increase its efficiency where native autotrophic capabilities are present. A wide array of natural and synthetic pathways for the assimilation of C_1 substrates have been explored and optimised towards achieving sustainable bioproduction goals, including the prevalent Calvin-Benson-Bassham (CBB) cycle, the Wood-Ljungdahl pathway (also known as the reductive acetyl-CoA pathway, rAcP), and the synthetic-turned-natural reductive glycine (rGly) pathway. The natural and engineered architectures of these and other pathways for the assimilation of C_1 substrates have been extensively reviewed in the literature[105–109].

More traditional metabolic engineering strategies have largely been driven by overexpression of key native or heterologous enzymes along carbon assimilation pathways. Recently, modular engineering approaches have emerged[110] and been applied extensively to improve autotrophy, most prominently to implement the rGly pathway in a range of hosts[111–115], as well as to install synthetic capabilities such as light-driven energy supply in non-photosynthetic hosts[116–119]. These diverse metabolic engineering strategies, which are often implemented at the genome scale, can be paired with adaptive laboratory evolution (ALE). As a systems-level tool, ALE enables *in vivo* exploration of strain optimisation landscapes beyond what is accessible by rational design alone[90]. Recent reviews have detailed the methods available to

implement systems-level metabolic engineering in combination with diverse evolution and selection strategies to develop efficient cell factories[120,121].

Lastly, bioproduction capabilities may be installed or optimised in the chosen chassis organism to convert assimilated CO_2 into value-added products. Whilst some autotrophic organisms can naturally synthesise useful compounds such as bioplastics, acetate or ethanol, efforts have been made to expand the range of industrially relevant molecules that can be obtained through autotrophic microbial fermentation. Despite the diversity in target compounds and biosynthetic pipelines, some standardised metabolic engineering strategies can be identified for directing carbon into desired products. Notably, growth-coupled production has become an emerging standard[110,122,123]. Frequently supported by genome-scale metabolic models, which are increasingly available for non-model autotrophic organisms, growth-coupling principles intrinsically link the production of target molecules and biomass. This correlation ensures the evolutionary stability of engineered biosynthetic pathways and simplifies the DBTL cycle at the core of strain engineering[122].

Having discussed the fundamental considerations for autotrophic chassis engineering, the focus of this introductory review will now shift to the most recent progress in the development of microbial platform chassis for autotrophic growth and bioproduction. Focusing primarily on advances reported in the past five years, the following analysis is divided into two broad categories: improving existing chemolithoautotrophic organisms, and the more challenging endeavour of installing autotrophic metabolism into traditionally heterotrophic hosts. Special attention is given to studies on the model chemolithoautotroph *Cupriavidus necator* H16, as this bacterium is the study species

for the work presented in this thesis. Note that, in the interest of brevity, the development of photoautotrophic cell factories such as Cyanobacteria or microalgae will not be discussed. For comprehensive reviews on this topic, references [124–126] can be consulted.

1.4.2. Acetogens as versatile autotrophic platform chassis with industrial-scale potential

Naturally occurring chemolithoautotrophic organisms have evolved to fix CO_2, using inorganic compounds as electron donors. There are a limited number of biotechnologically relevant chemolithoautotrophic bacterial species. The most developed host organisms are acetogens such as *Clostridium ljungdahlii* and *Clostridium autoethanogenum*, alongside the bioplastic-producing bacterium *Cupriavidus necator* H16.

Acetogens assimilate CO_2 via the ATP-efficient Wood-Ljungdahl pathway, converting CO_2 to acetyl-CoA with hydrogen as an electron donor. Due to their uniquely efficient mechanisms for energy conservation, acetogens have been described as operating at the thermodynamic limit of life[127]. A range of native products can be produced by autotrophic fermentation of acetogens on C_1 gas mixtures, such as ethanol, butyrate, butanol, 2-oxobutyrate, 2-3-butanediol, and acetate (reviewed comprehensively in ref. [128]). Limited attempts have been made to produce non-native chemicals in acetogenic hosts. The first successful report of genetic engineering in an acetogen to produce butanol was reported in 2010[129]. More recently, Liew and colleagues demonstrated carbon-negative production of acetone and isopropanol at industrial scale using an

engineered strain of *C. autoethanogenum*[130]. Their novel approach relied on mining a diverse collection of biosynthetic enzymes and assembling the enzyme sequences into a combinatorial library for pathway optimisation via high-throughput screening. The chassis microorganism was also optimised. Knock-out targets to improve production parameters were successfully identified using a combination of metabolic modelling and a streamlined cell-free method[131] for screening of native enzymes that may interfere with biosynthesis. The optimised biosynthetic enzymes and host strains were combined to deliver a scalable fermentation process with production rates of ~3 g/L/h at ~90% selectivity. The report by Liew and colleagues provides a first-in-class demonstration of how synthetic biology tools can be applied to genome-scale optimisation of an autotrophic microbial cell factory to deliver truly sustainable and scalable biomanufacturing.

Acetogens have also been harnessed for electromicrobial production, a hybrid process which couples abiotic (electrochemical) CO_2 reduction with subsequent biological assimilation. In a demonstration of efficient “artificial photosynthesis”, a type of electromicrobial production which is driven by photovoltaics, Haas and colleagues[132] used a solar-powered CO_2 electrolyser module to continuously generate syngas (a mixture of CO, CO_2 and hydrogen). The electrolyser was directly coupled to a fermenter module containing a mixed population of the acetogens *Clostridium autoethanogenum* and *Clostridium kluyveri*. Combined, these species delivered the production of butanol and hexanol from the electrochemically-generated syngas. The ground-breaking efficiencies reported in this study (100% Faradaic efficiency and 78% energy conversion efficiency of butanol and hexanol formation from syngas) provide evidence of the technical and economic potential of bio-hybrid systems. Moreover, the

study demonstrates how the range of autotrophically synthesised products can be expanded by combining host organisms with distinct metabolic capabilities to form a bioproduction consortium.

1.4.3. *Cupriavidus necator* H16: a malleable chemolithoautotroph

The energetic limitations of the Wood-Ljungdahl pathway, as well as the requirement for anaerobic conditions, limit the range of products that can be synthesised in acetogenic hosts. In contrast, aerobic chemoautotrophic hosts can mediate the synthesis of value-added chemicals requiring higher ATP investments. The facultative chemolithoautotroph *C. necator* has emerged as a promising platform chassis due to its versatile metabolism, as well as the increasing availability of tailored molecular tools to facilitate synthetic biology strategies[133]. The latter will not be discussed in this introductory review. Instead, a more detailed account of the molecular toolkit available for *C. necator* (and its limitations) is provided in the introductions to Chapters 2, 3 and 4 (Sections 2.2, 3.2 and 4.2).

Genome and metabolism

Cupriavidus necator H16 is a Gram-negative, facultative chemolithoautotrophic β-proteobacterium. First isolated from a spring in Germany in the early 1960s[134], it is known alternatively as *Ralstonia eutropha* H16, formerly as *Wautersia eutropha*[135] or *Alcaligenes eutrophus*[136], and originally as *Hydrogenomonas eutropha*[134]. Its genome consists of three circular replicons: chromosome 1 (4,052,032bp), chromosome 2 (plasmid-like, 2,912,490bp) and the megaplasmid pHG1 (452,156bp), all of which

have been fully sequenced[137]. *C. necator* is a strictly respiratory bacterium, capable of both aerobic and anaerobic growth. Under anoxic conditions, the bacterium can use nitrate as an alternative electron acceptor[138]. Its growth proceeds via heterotrophic, lithoautotrophic or mixotrophic (including organoautotrophic) metabolism, depending on carbon and energy source availability (Figure 1.6)[137,139]. *C. necator* has been shown to reach high cell densities of ~200 g dry cell weight per litre whether cultivated under autotrophic or heterotrophic conditions[140,141].

The bacterium can grow heterotrophically on a wide range of carbon sources. The preferred substrate for its heterotrophic cultivation is fructose, though it may utilise many other organic compounds, such as *N*-acetylglucosamine, gluconate, fatty acids, alcohols, glycerol, amino acids, lignin derivatives, TCA cycle intermediates and aromatic compounds[139,142,143]. In the absence of organic substrates, *C. necator* may grow lithoautotrophically using H_2 and CO_2 as sole energy and carbon sources. CO_2 fixation in this bacterium proceeds through a canonical CBB cycle[139]. The key enzymes of the CBB cycle are encoded in the cbb_c operon on chromosome 2, as well as a highly homologous second operon (cbb_p) located on the pHG1 megaplasmid. Both operons are regulated by CbbR, a LysR-type transcriptional regulator encoded on chromosome 2[144]. A second copy of the *cbbR* gene remains upstream of the cbb_p operon. The plasmid-borne copy is defective, having been inactivated by a series of short deletions[137,145]. To generate energy for chemolithoautotrophic growth, hydrogen (H_2) can be oxidised by two Ni-Fe hydrogenases encoded on pHG1[146]. Alternatively, formate may be used by the bacterium as an energy and/or carbon source for organoautotrophic growth[144,147]. At least four distinct formate dehydrogenase enzymes have been identified across the *C. necator* genome[137].

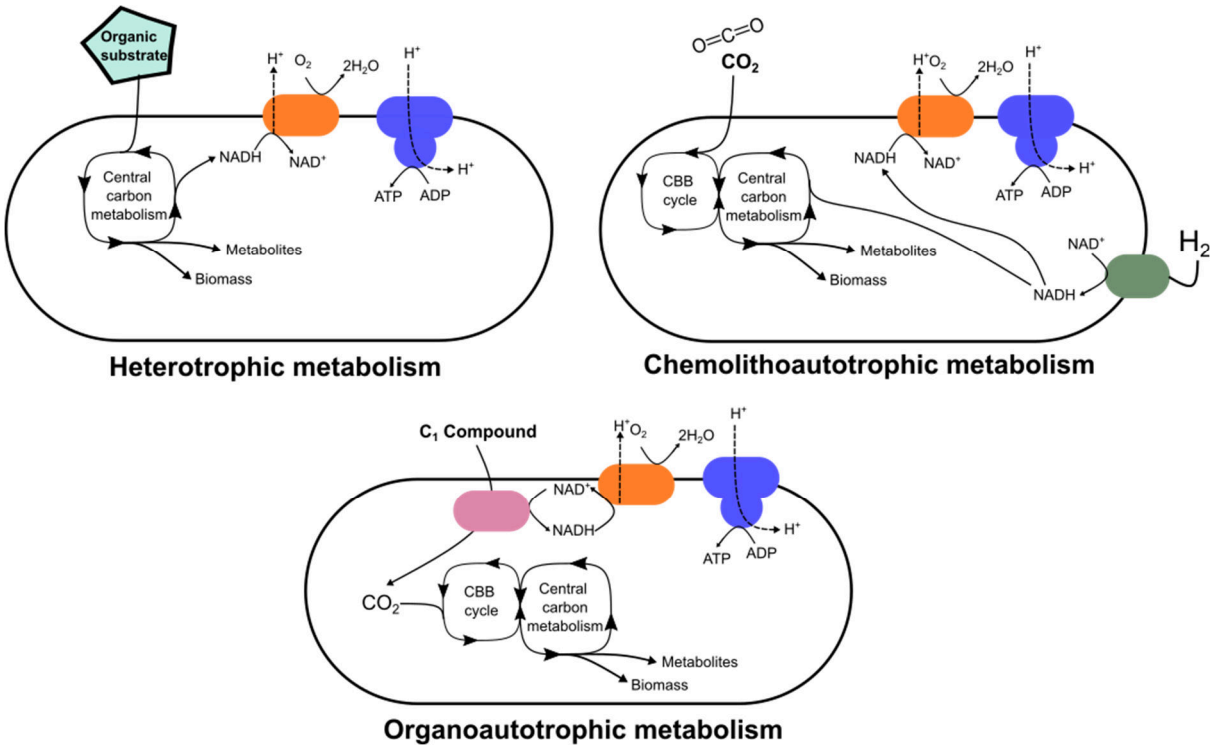


Figure 1. 6. Growth modes of *C. necator*.

During heterotrophic growth, a variety of organic substrates may be oxidised to generate reducing power (NADH) and energy (ATP). During chemolithoautotrophic growth, CO_2 fixation proceeds through the CBB cycle. Energy and reducing power are obtained from hydrogen via oxidation by hydrogenase enzymes. During organoautotrophic growth, a type of mixotrophy, a C_1 substrate is oxidised to CO_2 to generate energy and reducing power. The resulting CO_2 is then fixed through the CBB cycle, feeding into central carbon metabolism.

Harnessing *C. necator* to valorise waste carbon streams

There is a growing body of research on using sustainable feedstocks for *C. necator* fermentation and value-added chemical production. A valuable trend in this research is the valorisation of waste streams, for instance by using glycerol, coffee waste oil or plant oil as substrates[148–150]. Recent studies have also focused on CO_2 valorisation via metabolic engineering of this organism for chemolithoautotrophic growth and

production. In addition to naturally-accumulated polyhydroxyalkanoates (PHAs)[151], *C. necator* has been engineered to efficiently produce a wide array of molecules through autotrophic fermentation, such as sugars[151,152], carboxylic acids[153], alcohols[154,155], methyl ketones[156], terpenes[157], or alkanes[158].

Both wild-type and engineered *C. necator* strains have been implemented within bio-hybrid systems for electromicrobial production. The first demonstration of this was provided by Li *et al*. in 2012[154]. Here, *C. necator* was directly interfaced with a formate-producing electrochemical system to mediate the biotransformation of formate into isobutanol and 3-methyl-1-butanol. Interestingly, they observed that reactive oxygen and nitrogen species generated during electrochemical CO_2 reduction inhibit bacterial growth. Though this limitation could be alleviated by shielding the anode with an inexpensive ceramic cup, the resulting total biofuel titer of 140 mg/l was 10-fold lower than the titer recorded using a formate-feeding fermenter.

Recent studies have sought to push the limits of electrochemical CO_2 reduction to formate and subsequent biotransformation by *C. necator*. Chen and colleagues demonstrated an innovative approach where an FDH enzyme in the cathodic chamber catalyses CO_2 reduction to formate, rather than the cathode itself[159]. To minimise the toxicity of reactive oxygen and nitrogen species as observed by Li *et al*., they also implemented physical separation of the cathodic and anodic chambers using a proton exchange membrane. A recent study by Lim *et al*. also provides support for securing live *C. necator* in a separate fermenter module to minimise adverse effects on bacterial growth[160].

Using a different iteration of electromicrobial synthesis, *C. necator* has been cultivated directly on CO_2 using hydrogen generated by a water-splitting system[155,161]. These studies have demonstrated that, where the electrolyser module is powered by a solar-to-electricity device, solar-to-fuel efficiencies in the range of ~10% and ~7-8% can be attained for biomass and value-added products, respectively. These efficiencies are well in excess of those attained by natural photosynthetic systems (1-3%), which are limited by a variety of factors including photochemical inefficiency and photorespiration (see ref. [162] for a comprehensive review).

Metabolic engineering of *C. necator* towards improved growth on and production from renewable C_1 feedstocks

Successful attempts to improve autotrophic metabolism in *C. necator* via the native CBB cycle have recently been demonstrated. Two independent reports have shown that the carboxylation bottleneck in the CBB cycle can be partially overcome by the heterologous overexpression of a cyanobacterial Rubisco, boosting biomass and PHA yields[159,163]. Strains with improved autotrophic growth kinetics have also been obtained by ALE. Calvey *et al*. recently described a *C. necator* strain with a 24% improvement in maximum growth rate on formate relative to the wildtype[92]. The strain was isolated through an innovative pipeline, combining evolution, -omics data, and rational genome engineering, including genome reduction approaches. First, bacteria were cultured in a permissive formate concentration (50 mM) for a total of 400 generations. The best-performing individual isolates were identified and characterised by whole-genome sequencing, revealing a reduced set of unique and shared

mutations which were hypothesised to contribute to the desirable formatotrophic growth kinetics. To discern the effects of individual mutations, these were recapitulated in a wildtype background. Several engineered strains were created, including genome-reduced variants where the organism's pHG1 megaplasmid was fully removed. Some of these strains exceeded the formatotrophic performance of even the best-evolved isolates, providing further evidence for the synergistic combination of ALE and rational engineering approaches.

Further, *C. necator* has been used as a sandbox for engineered autotrophy, most notably for the implementation of the synthetic reductive glycine (rGly) pathway. First proposed by Arren Bar-Even as the most efficient synthetic formate assimilation pathway for engineered autotrophy[107], the pathway has since been identified in the naturally-occurring bacterium *Desulfovibrio desulfuricans*[164]. Work by Claassens *et al*. successfully rewired the central carbon metabolism of *C. necator* to incorporate formate via the rGly pathway, thereby replacing the native CBB cycle entirely (Figure 1.7)[111]. A modular engineering approach was implemented to accomplish this feat. First, a module allowing the conversion of formate to glycine via condensation with tetrahydrofolate (THF) was expressed. This required the expression of three heterologous enzymes to mediate flux from formate to 5,10-methylene-THF, as well as overexpression of native enzymes to mediate the conversion of 5,10-methylene-THF to glycine. The regulation and performance of this first metabolic module was optimised in a purpose-built *C. necator* strain exhibiting glycine auxotrophy. This approach represents an exemplary implementation of growth-coupled selection and modular optimisation of pathway designs. Next, glycine assimilation into biomass was investigated. Out of the two possible routes, proceeding either through glyoxylate or

serine as intermediates, the latter is more efficient and therefore preferable. Flux was successfully forced through the serine route by overexpression of native enzymes. Following further strain optimisation through short-term ALE, the final "CRG4" strain obtained in the study exhibited a formatotrophic biomass yield similar to the wildtype. Since this ground-breaking report, work by Dronsella, Claassens and colleagues has further demonstrated that genomic integration of the modules encoding the rGly pathway results in *C. necator* strains with formatotrophic biomass yields exceeding the wildtype (14% yield increase)[112]. This work constitutes the first report of an engineered C_1-assimilation pathway surpassing the yield of an organism's native pathway, providing a framework for the development of more proficient autotrophic host organisms through synthetic and systems biology.

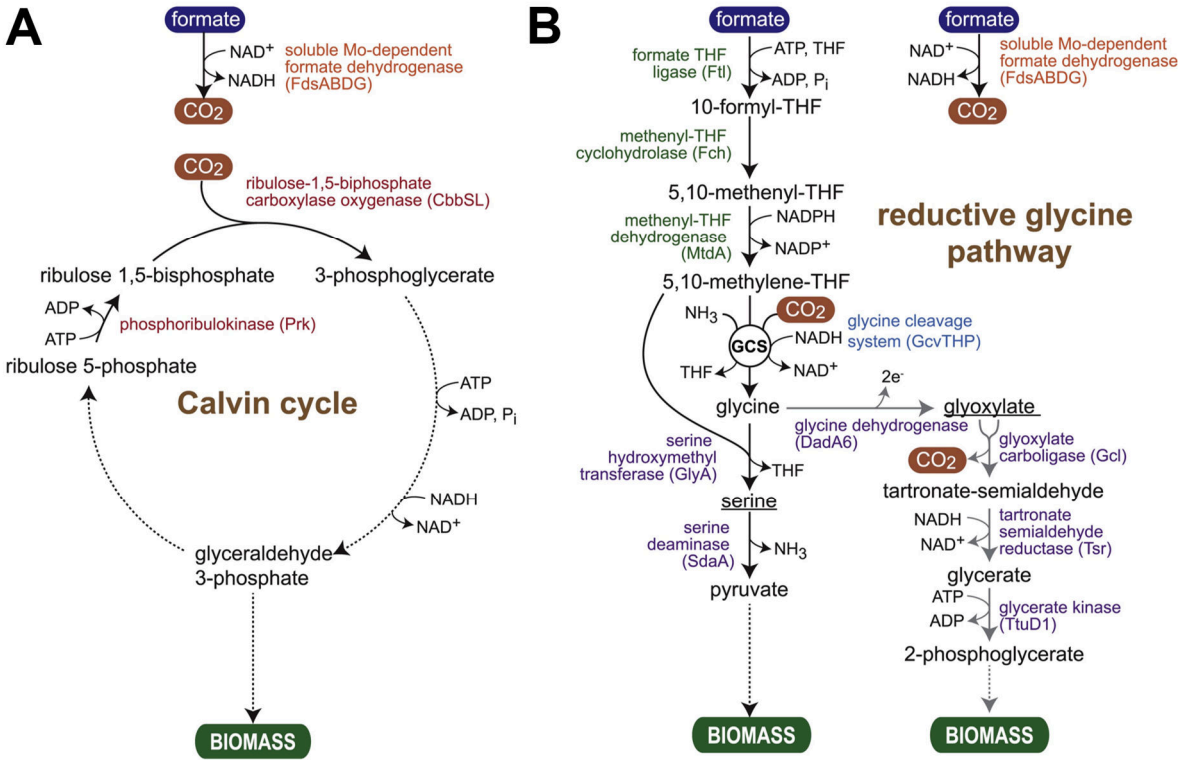


Figure 1. 7. Schematic overview of the CBB and rGly pathways in *C. necator*.

(A) During formatotrophic growth, formate is oxidised to CO_2 by a molybdenum-dependent soluble formate dehydrogenase enzyme. CO_2 is then assimilated into biomass via the native CBB cycle. **(B)**

Two variants of the rGly pathway can be implemented. In the original design (left), glycine is assimilated via serine and pyruvate. Alternatively, glycine can be converted to glyoxylate, which is subsequently assimilated via the glycerate pathway (right). This figure was taken from ref. [111]. For further interpretation of the figure abbreviations and colour schemes, the original publication should be consulted.

1.4.4. Engineering synthetic autotrophy

The studies discussed thus far have unequivocally evidenced that engineered autotrophic platform chassis such as *C. necator* can mediate sustainable bioproduction. Yet these organisms are still relatively unknown to industry compared to their heterotrophic counterparts, such as *E. coli* or *S. cerevisiae*. In line with this, recent work in platform chassis development has focused on endowing well-established heterotrophic hosts with autotrophic capabilities, to expedite the use of (at least partially) autotrophic hosts for more sustainable industrial applications.

Heterologous expression of Rubisco in *E. coli* has been explored as early as the 1980s[165], setting the stage for more ambitious metabolic engineering of this heterotrophic host strain towards carbon fixation, as has been demonstrated in recent years. The first example of carbon assimilation through an engineered CO_2 fixation module in *E. coli* was reported by the laboratory of Ron Milo in 2016. Here, hemi-autotrophic growth was successfully achieved by decoupling the metabolic modules needed for energy production and carbon assimilation[166]. Since then, work by the same group provided the first example of an *E. coli* strain capable of generating all biomass carbon from CO_2, using formate as an electron source[167]. Efficient assimilation of C_1 substrates in *E. coli* has also been reported by engineering a tetrahydrofolate cycle, similarly delivering strains that could grow on CO_2 and formate alone[168–171]. The rGly pathway has also been successfully implemented in *E. coli* to enable growth on formate and methanol[113]. In a paradigm-shifting study, Satanowski

et al. demonstrated that autotrophic metabolism can be attained in *E. coli* without the need to engineer new pathways, or even express heterologous enzymes. Instead, its endogenous metabolism can be rewired towards thermodynamically-feasible carbon assimilation pathways[172]. Interestingly, work by the Savage group has proposed harnessing carbon-concentrating mechanisms to boost carbon assimilation in non-native autotrophic hosts. They demonstrated that heterologous expression of a 20-gene cluster from *Halothiobacillus neapolitanus*, encoding an α-carboxysome, was sufficient to allow CO_2-dependent growth of engineered *E. coli* in ambient air[173].

In addition to *E. coli*, two recent reports have demonstrated the possibility of engineering C_1 metabolism in the industrially-relevant bacterium *Pseudomonas putida*[114,115]. Both studies implement the tried-and-tested modular engineering approach to install the rGly pathway in this host, further highlighting the portability of this pathway and its potential to deliver autotrophy to novel platform chassis.

Eukaryotic microbes have also been subjects for engineered autotrophy. An early study showed that heterologous Rubisco expression in *S. cerevisiae* could be harnessed to re-oxidise NADH using CO_2 as an electron acceptor, leading to an 8% increase in ethanol production yields[174]. More ambitious metabolic rewiring towards autotrophy have also been explored in yeast. Notably, Gassler and colleagues successfully converted *Pichia pastoris* into an autotroph able to use CO_2 as sole carbon source[175]. Their novel approach harnessed the host's native peroxisomal xylulose monophosphate pathway, which enables methanol assimilation, as a background against which a heterologous CBB cycle could be installed. In an effort to expand C_1 substrate utilisation in industrial yeast strains, Zhan *et al*. implemented

combinatorial pathway construction, modular engineering and ALE to successfully enable growth of *S. cerevisiae* on methanol as sole carbon source[176].

1.5. Conclusions and Outlook

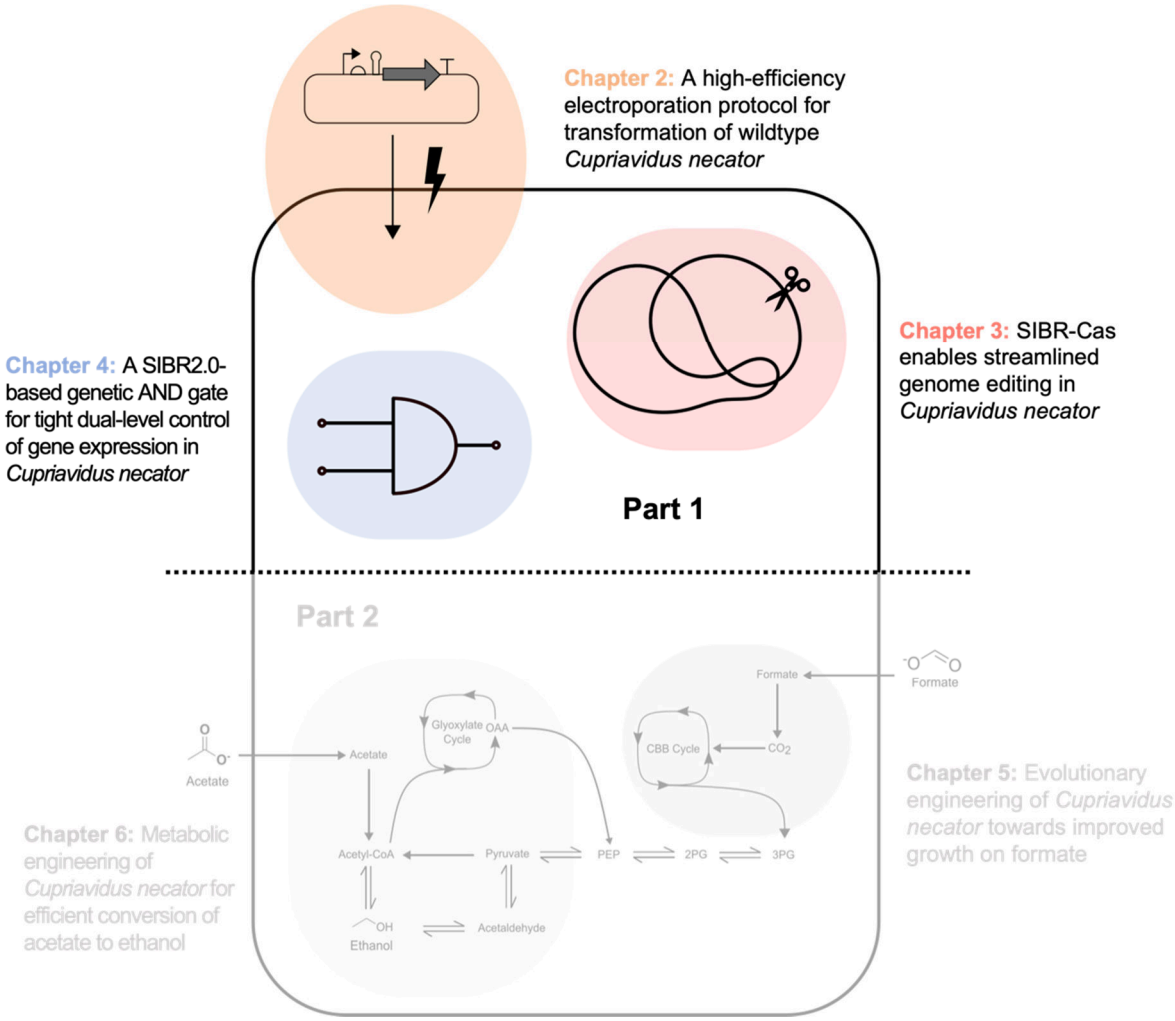
Autotrophic platform chassis, whether native or synthetic, hold the potential to catalyse the shift towards using renewably generated feedstocks, such as C_1 substrates, to produce value-added compounds. As highlighted in this introductory literature review, a wide range of strategies have been implemented to construct microbial cell factories with attractive capabilities for sustainable bioproduction. These engineered strains have been made possible by significant advances in the fields of synthetic and systems biology, such as genome engineering technologies, combinatorial pathway construction and optimisation, cell-free prototyping systems, metabolic modelling, and evolutionary engineering.

High conversion rates and efficiencies are needed in order for biocatalytic strains to deliver economically- and environmentally-sustainable C_1 bioprocesses at the industrial scale. Yet a large fraction of the autotrophic platform chassis that have been developed to date (particularly fully synthetic autotrophs) exhibit poor performance parameters, such as slow growth rates, low biomass and product yields, and metabolite or energetic imbalances. Thus, further optimisation is required before these chassis can be efficiently applied to industrial processes. Additionally, some promising frameworks for CO_2 valorisation, such as electromicrobial production, face significant technical and economic challenges related to scale-up[105].

Nevertheless, considering the development of platform chassis with performance parameters that not only match but surpass native autotrophy, truly sustainable biomanufacturing from CO_2 and other C_1 substrates is becoming an increasingly viable alternative to petrochemical-based production. The coming years are expected to yield substantial advances in the field of CO_2 valorisation, with important implications for bioindustry and the broader shift to a circular carbon bioeconomy.

The research presented in this thesis delivers substantial contributions towards this shift, specifically by advancing the development of *C. necator* as a platform chassis for sustainable production from renewable substrates. Chapters 2-4 describe the development of advanced methods for DNA delivery, genome editing, and precise regulation of gene expression. These tools have been tailored and optimised for use in *C. necator*, and are therefore primed to enable the implementation of more ambitious strain engineering designs in this non-model bacterium. In Chapter 5, an evolutionary engineering strategy is implemented to obtain *C. necator* strains with improved growth on formate, which could be applied to deliver more efficient C_1 bioprocesses. Finally, in Chapter 6, bioproduction from the C_2 substrate acetate is explored, demonstrating how the expression of key endogenous and heterologous genes in *C. necator* can improve production titer, rate and yield.

Part 1:
Molecular tools and methods for *C. necator* engineering



2. A high-efficiency electroporation protocol for transformation of wildtype *Cupriavidus necator*

2.1. Aims and Objectives

The aim of the work presented in this Chapter is to overcome the inefficient transformation of *C. necator*, which limits the scope of genetic engineering efforts in this organism. In contrast with other studies, which sought to address low transformation efficiencies through strain and plasmid engineering, our objective is to identify novel protocol options for efficient electroporation of the wildtype strain. To do so, we aim to undertake a systematic investigation of protocol parameters, optimising key variables which underlie electroporation efficiency.

2.2. Introduction

C. necator is a genetically tractable organism, with a sizeable repertoire of molecular biology tools available for strain engineering and synthetic biology. A wealth of plasmid vectors for recombinant DNA construct expression in *C. necator* have been developed. These include a wide range of replication elements, antibiotic resistance markers, controllable gene expression systems and strategies for plasmid stability and

maintenance, as reviewed recently[133]. Of these, the broad-host range vector pBBR1 and its derivatives are most widely used, with several versions readily available through the Standard European Vector Architecture (SEVA) repository[177,178].

A basic technique used for plasmid delivery into *C. necator* is conjugative transfer. Whilst efficient, this method is cumbersome and lengthy. The conjugation process requires co-culture of specialised *E. coli* donor strains (typically *E. coli* S17) and *C. necator* recipient strains to enable plasmid transfer, commonly referred to as mating. Subsequently, *C. necator* transconjugants are isolated by multiple rounds of culture on selective medium containing gentamycin, exploiting *C. necator*'s natural resistance to this antibiotic, which the donor strains lack. As a result, conjugation requires at least five days from transformation of donor strains to isolation of transconjugants, with the process becoming even lengthier when multiple rounds of selection are required.

When considering the number of recombinant DNA constructs typically required for metabolic engineering and synthetic biology, plasmid transfer by conjugation may become prohibitively time-consuming. Thus, alternative methods for plasmid transfer based on electroporation have been developed. Such methods are substantially less laborious than conjugation, requiring only a single, direct transformation of plasmid DNA into the target strain. The highest transformation efficiency reported in the literature is in the range of 10^5 CFU/µg DNA (when transforming a standard broad-host range plasmid vector in the wildtype strain)[179]. This lags far behind the efficiencies attained in *E. coli,* routinely reported in the range of 10^9-10^{10} CFU/µg DNA**.**

The ability to transform recombinant DNA constructs with high efficiency is critical to advanced metabolic engineering and synthetic biology strategies. In general, these methods operate on an engineering-inspired Design-Build-Test-Learn (DBTL) cycle (see Section 1.3.2), with rapid prototyping as a core capability. Combinatorial optimisation is often required. Large DNA sequence libraries are constructed, for instance, to test different architectures for biosynthetic pathways, scrambling the relative positions of protein-coding genes and genetic control elements. Large libraries of enzyme sequence variants may also be built for *in vivo* screening. In this context, the transformation efficiency of the host strain becomes an important bottleneck in the DBTL cycle, as it restricts the size of the combinatorial space which can be explored in the “Test” step.

Previous studies have sought to overcome the limited transformation efficiency of *C. necator*. The first successful electroporation protocol with a broad host range plasmid was reported by Park and colleagues in 1995, with a modest efficiency of ~10^1-10^2 CFU/µg DNA[180]. They proposed using 10% glycerol as the electroporation buffer, which allows the freezing of competent cell aliquots for long-term storage. More recently, a report by Tee and colleagues provided the most extensive characterisation of factors affecting electroporation efficiency in the organism and reported an efficiency of ~4x10^5 CFU/µg DNA[179]. Key innovations from this study were the pre-treatment of cells with 50 mM $CaCl_2$, the use of 0.2 M sucrose in the transformation buffer, and the addition of fructose to growth and outgrowth media.

Several parameter ranges and effects identified by both Park and Tee are now widely accepted and fixed in literature reports. Electrocompetent cells are invariably prepared from mid-exponential phase cultures, with an optimum optical density (OD) in the range of 0.4-0.8. Cells are concentrated during treatment; for a given amount of DNA, transformation efficiency has been shown to increase with an increasing number of cells in the aliquot used for electroporation. Efficiency is thought to improve linearly with increasing amounts of DNA, up to 1 µg (as reported by Tee *et al*.[181]). Other parameters, such as buffer composition, field strength, or the choice of plasmid vector and recipient strain, vary more widely in the literature (Table S2.1).

Broadly, studies aiming to improve transformation (almost exclusively electroporation) efficiency in *C. necator* have predominantly focused their optimisation on the narrow range of parameter values which are commonly used for the model organism *E. coli*, or fixed in early studies. The search of the parameter space for the optimal electroporation protocol for *C. necator* has not been exhaustive. Having recognised this, we investigate whether modifying a reduced set of key variables could lead to a significant increase in transformation efficiency. We propose an updated protocol for electroporation of wildtype *C. necator* which delivers the highest transformation efficiency reported to date (routinely in the range of 10^7-10^8 CFU/µg DNA), representing (at least) a ~10-fold improvement over previously reported values. In addition, we demonstrate that the optimised protocol can be used with a low-cost, do-it-yourself (DIY) electroporation kit, enabling an easy electroporation operation without the requirement for sophisticated commercial instruments.

2.3. Materials and Methods

Bacterial strains, plasmids, and culture conditions

All plasmids and bacterial strains used in this study are listed in Table S2.2 and Table S2.3, respectively. Plasmids pSEVA231, pSEVA331, and pSEVA531 were obtained from the SEVA repository[177,178]. Plasmids pCAT201, pCAT204, and pCAT206 were a gift from the laboratory of Thomas Howard (Addgene plasmids #134878, #134881 and #134883, respectively)[182]. All plasmids were isolated from *E.* coli DH5α, using the NEB Monarch® Miniprep Kit according to the manufacturer's specifications. Plates for bacterial growth were Luria-Bertani broth (LB) supplemented with agar. Liquid media was LB or Super Optimal Broth (SOB), supplemented with 20 mM fructose, as indicated. Where appropriate, antibiotics were used at 100 µg/ml or 200 µg/ml (kanamycin), 10 µg/ml (tetracycline) or 50 µg/ml (chloramphenicol). All *E. coli* strains were incubated at 37 °C with vigorous shaking (200 rpm). *C. necator* strains were incubated at 30°C with moderate shaking (150 rpm). OD of bacterial cultures was measured using a UV-1800 UV/Vis spectrophotometer (Shimadzu) or the built-in spectrophotometer of a Chi.Bio device[183].

Electrocompetent cell preparation and transformation

C. necator competent cells were prepared by adapting an existing protocol developed for *Pseudomonas aeruginosa*[184]. Bacterial strains were streaked out from glycerol stocks onto LB agar plates and incubated for 48 h at 30°C. Bacterial cultures in 5 ml

SOB medium were inoculated from single colonies and incubated overnight (16-18 h). A small volume (~200 µl) of the saturated overnight cultures was used to inoculate large 50 ml cultures in SOB medium, in 250 ml conical flasks, which were grown to the desired OD (0.2-5, as indicated). Following incubation, large liquid cultures were split into two 50 ml tubes and pelleted by centrifugation at 4000 rpm for 10 min. Cell pellets were washed twice in room-temperature 1 mM $MgSO_4$, and each pellet resuspended in 1 ml 1 mM $MgSO_4$. Cells were pooled, and sterile 50% (v/v) glycerol was added to a final concentration of ~25% (v/v). Cells were divided into 50 µl aliquots and frozen at -80°C. For transformation, competent cell aliquots were thawed at room temperature and mixed with plasmid DNA (0-500 ng, as indicated). Electroporation was performed using room-temperature 0.1 or 0.2 cm gap cuvettes, at 1.8 or 2.5 kV, using default settings Ec1 or Ec2 in a Bio-Rad MicroPulser electroporator (Bio-Rad Laboratories). Immediately after electroporation, 0.95 ml of recovery medium (SOB, unless otherwise stated) was added. The resulting 1 ml of culture was transferred to a 1.5 ml microcentrifuge tube and incubated at 30ºC with 150 rpm shaking for 2-4 h (outgrowth). Following outgrowth, cells were serially diluted and plated onto selective LB plates to enable quantification of transformation efficiency (always quantified as CFU/µg DNA), or onto non-selective LB plates to quantify the total number of cells present in the transformation culture. The procedure for serial spot microdilution is illustrated in Figure S2.1.

Building and testing the DIY electroporator

A DIY electroporator was built following the guidelines outlined for ElectroPen[185]. Briefly, a piezoelectric crystal was extracted from a refillable gas kitchen lighter (L&H,

Amazon ASIN B08DMTCK2R). Copper wires were soldered onto the piezo igniter as indicated. The igniter was insulated using heat-shrink tubing and electrical tape, and could be operated using a hammer-action striking mechanism, preserved from the original kitchen lighter. For electroporation, the copper wires were attached to the metal plates of commercially available electroporation cuvettes (0.1 or 0.2 cm gaps, as indicated) with electrical tape. The cuvettes were filled with the same mixture of competent cells and DNA used with the conventional electroporator. This cell suspension was shocked 1-10 times with the DIY electroporator, after which recovery medium (0.95 ml SOB) was immediately added. Outgrowth and selection were performed as described for the conventional electroporator.

2.4. Results

2.4.1. SOB medium supports rapid, high-density cultivation of C. necator

Several rich growth media are routinely used to culture bacteria for electrocompetent cell preparation. Amongst these, Luria-Bertani (LB) and Super Optimal Broth (SOB) media are most common, often supplemented with glucose to induce catabolite repression and provide additional carbon for growth. To determine which rich growth medium can support high-density cultures of *C. necator* for competent cell preparation, we measured the specific growth rate (calculated as described in ref. [186]) and carrying capacity (final cell density of the bacterial population, quantified as the OD reached after 25 hours of cultivation) of this bacterium when cultured in LB and SOB alone, or

supplemented with 20 mM fructose. Fructose was used in place of glucose as wildtype *C. necator* cannot metabolise glucose.

The highest specific growth rate (~0.55 h^{-1}) and carrying capacity (OD ~6.3) were observed for SOB cultures (Figure 2.1). Fructose supplementation did not improve either parameter. SOB was therefore selected as the culture medium for electrocompetent cell preparation, as it enables culturing at a wider range of densities than LB. Moreover, due to the faster doubling time of *C. necator* in SOB, cultures of any density can be harvested more rapidly, thereby expediting the electrocompetent cell preparation protocol.

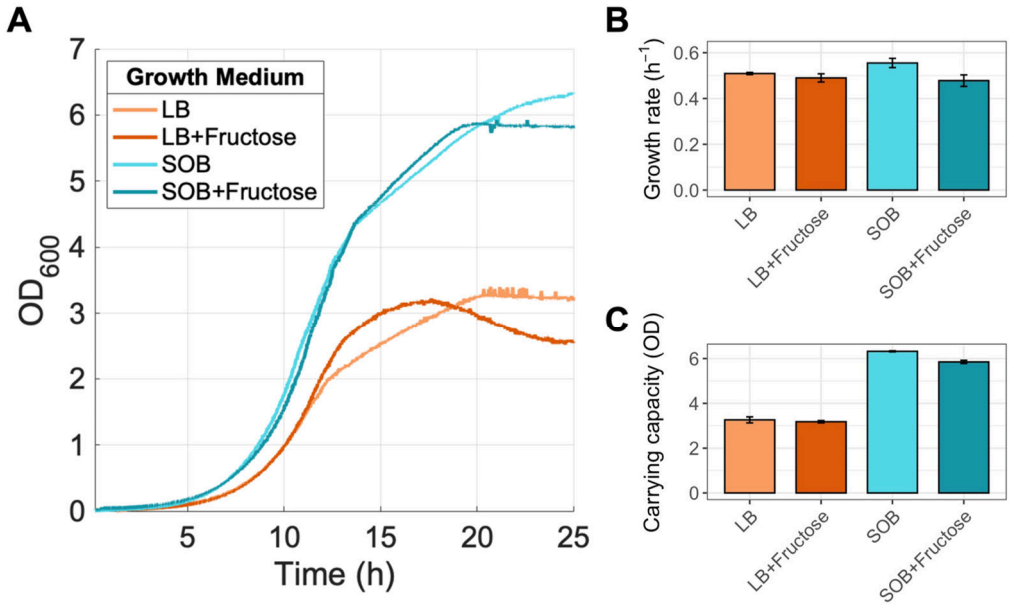


Figure 2. 1. Growth kinetics of *C. necator* in rich culture media.

(A) Representative growth curves measured in Chi.Bio devices operated in batch mode. Cells were grown at 30ºC with vigorous magnetic stirring in LB, LB supplemented with 20mM fructose, SOB or SOB supplemented with 20mM fructose. **(B)** Specific growth rate and **(C)** carrying capacity of *C. necator* cultured in each growth medium. For (B) and (C), bars indicate the mean ± one standard deviation of *n* = 3 estimations.

2.4.2. Using high-density cultures for competent cell preparation increases electroporation efficiency

Previous studies have proposed that maximal transformation efficiency (defined as the number of colony forming units recovered on selective plates per microgram of plasmid DNA supplied) can be achieved by preparing electrocompetent cells from *C. necator* cultures harvested in early to mid-exponential phase, corresponding to an OD in the range of 0.4-0.8[179,180,187]. Cell harvest at higher optical densities has not been explored.

To determine whether using high-density cultures can improve transformation efficiency, we prepared electrocompetent cells from *C. necator* cultures harvested at ODs in the range of 0.2-5 using a novel procedure (see Materials and Methods for details). Briefly, 1 mM $MgSO_4$ supplemented with ~25% glycerol was used as the final electroporation buffer. This buffer has been previously proposed for efficient electroporation of *Pseudomonas aeruginosa*[184] and, to the best of our knowledge, has not been previously tested for *C. necator*.

While varying the OD of cultures used for the preparation of competent cells, all other protocol parameters were kept constant. Competent cells were prepared as detailed in Section 2.3. For electroporation, 0.2 cm gap cuvettes were used, and a field strength of 12.5 kV/cm was applied. During recovery (outgrowth), cells were grown in SOB medium for two hours. All protocol steps were performed at room temperature. The standard plasmid vector pSEVA231, which includes the pBBR1 origin of replication and a kanamycin resistance cassette (*kanR*), was used for electroporation (250 ng).

In this first experiment, solid medium (LB agar) supplemented with 100 or 200 µg/ml kanamycin was used to select transformants. Both working antibiotic concentrations are commonly reported in the *C. necator* literature and were tested in parallel to eliminate the possibility of high transformation efficiency being observed due to sub-optimal stringency of selection.

No significant differences in transformation efficiency or frequency were observed between the two kanamycin concentrations, indicating that either concentration is suitable for stringent selection of transformants, as shown in Figure 2.2. Considering this, 100 µg/ml was used as the working kanamycin concentration in all subsequent experiments. For competent cells harvested at ODs of 0.4-0.8, the recorded transformation efficiency was comparable to that obtained in other publications using the pBHR1 plasmid[179]. Interestingly, we found that transformation efficiency increases linearly with increasing OD (Figure 2.2 B). Maximal transformation efficiency was obtained when cultures were harvested at an OD of 5 (i.e. near-saturated cultures), with a value of $(5.8\pm0.4)\times10^{7}$ CFU/µg DNA. This represents a ~10^{2} improvement in this value over previous reports[179], though direct comparisons are not possible (*see* Section 2.5 for details). Transformation frequency (defined as the number of colony forming units recovered on selective plates per survivor cell in the competent cell aliquot) was in the range of 10^{-3}-10^{-2} CFU/cell, with no significant differences in frequency between competent cells prepared from cultures at different densities (Figure 2.2 C).

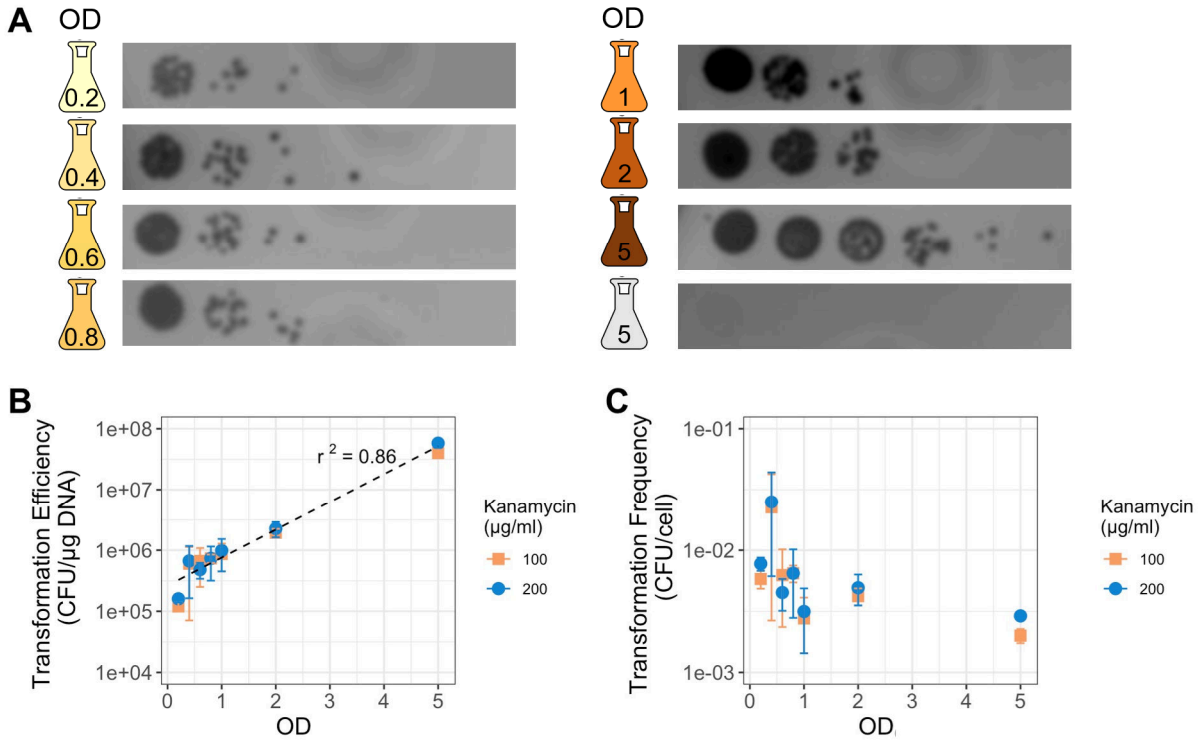


Figure 2. 2. Optimisation of culture OD at the time of cell harvest.

(A) Spot microdilutions of transformation cultures on LB agar solid medium, supplemented with 200 µg/ml kanamycin, which were used to quantify transformation efficiency (CFU recovered on selective plates per µg of plasmid DNA) and transformation frequency (CFU recovered on selective plates per surviving cell in the competent cell aliquot). For each microdilution, the OD of the cell culture used to prepare the competent cell aliquots is indicated within the flask icon. In all cases, plasmid pSEVA231 was used for transformation. The last dilution (grey icon) corresponds to a cell aliquot prepared from an OD = 5 culture, where no plasmid DNA was added prior to electroporation (i.e., negative control). **(B)** Transformation efficiency and **(C)** frequency of cell aliquots prepared from cultures at each OD. Individual points indicate the mean ± one standard deviation of *n* = 3 replicates. Least-means linear regression is shown as a dashed black line, with the indicated r^2 coefficient.

2.4.3. Effect of outgrowth time and medium on electroporation efficiency

Transformation protocols include an outgrowth (recovery) step immediately following DNA delivery via heat shock or electroporation. Cells are briefly grown in a non-selective rich medium, typically SOC (SOB supplemented with 20 mM glucose for

catabolite repression), to allow transformants to express antibiotic resistance genes prior to selection on solid medium. Previous studies have shown that the duration of outgrowth, as well as the composition of the outgrowth medium, can affect transformation efficiency in *C. necator*[179]. In line with this, we investigated whether outgrowth duration and medium composition are important parameters for our protocol. All other parameter values were kept constant, and competent cells were prepared from saturated (OD 5) SOB cultures, as indicated in Section 2.4.2.

We observed that transformation efficiency improves as a function of outgrowth time for up to two hours, in line with previous reports (Figure 2.3 A). No significant improvements in transformation efficiency were observed upon increasing outgrowth to three or four hours. Notably, high transformation efficiency (in the order of 10^6 CFU/µg DNA) was achieved even when the outgrowth step is omitted, highlighting the streamlined nature of the protocol. In subsequent experiments, an outgrowth time of two hours was used.

A range of recovery media have been used for transformation of *C. necator*, including SOC. This is surprising since *C. necator* cannot metabolise glucose, meaning the addition of this sugar does not induce catabolite repression nor does it provide an additional carbon source to boost growth. Fructose, which can be preferentially metabolised in the wildtype strain, has also previously been added to recovery media. We tested whether conducting the outgrowth step in SOB or SOB with added fructose could lead to higher efficiency than using a standard rich growth medium with no added sugar, namely LB (Figure 2.3 B). No significant differences in transformation efficiency were observed, indicating that the exact composition of the recovery medium is not a

limiting factor in the proposed protocol. Hereafter, the recovery step was performed by growing cells in SOB medium for two hours.

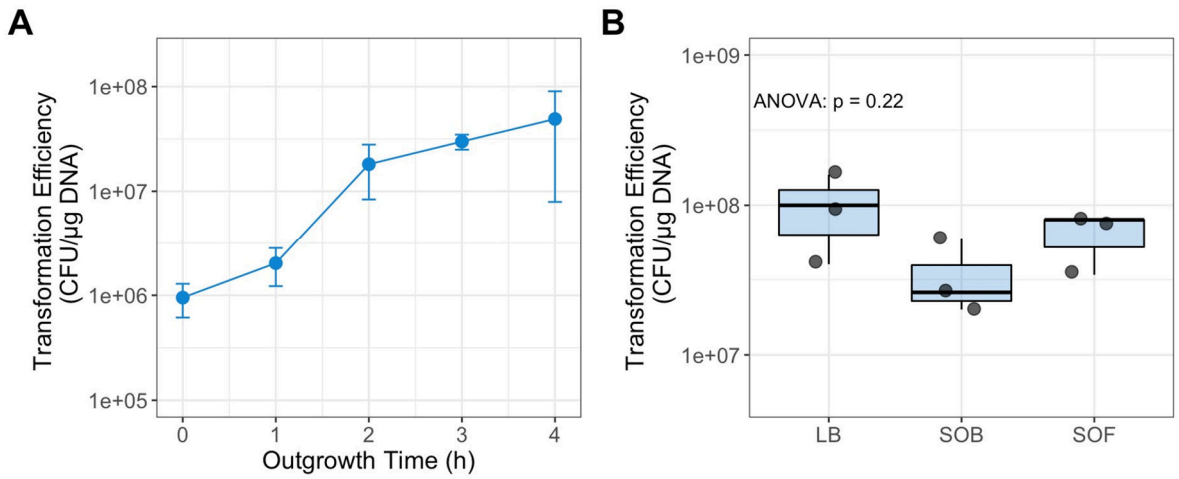


Figure 2. 3. Outgrowth optimisation.

(A) Transformation efficiency as a function of outgrowth duration. The outgrowth step was performed in 1 ml SOB cultures, as described in Section 2.3, and all other electroporation protocol steps were kept constant, as indicated in Section 2.4.2. Individual points represent the mean ± one standard deviation of $n = 3$. **(B)** Transformation efficiency as a function of recovery medium composition (LB, SOB, or SOB supplemented with 20mM fructose) for transformation cultures grown for two hours following electroporation. Individual observations ($n = 3$) are shown as grey dots. The p-value of a one-way ANOVA test is shown.

2.4.4. The electroporation method is robust to changes in field strength

An important parameter in any electroporation method is the strength of the electric field that is applied for DNA delivery, typically reported in kV/cm. Optimal field strength is known to be method-dependent, as the exact composition of the electroporation buffer, as well as the cell type and density, modulate the conductivity of the cell suspension. Thus far in our optimisation, we consistently applied a field strength of

12.5 kV/cm at room temperature. We investigated whether varying the field strength (while keeping all other parameters constant, as described above) could further increase the transformation efficiency for our protocol. All tested field strengths delivered transformation efficiencies higher than previously reported (>10^5 CFU/ µg DNA), with 12.5 kV/cm as the optimum (Figure 2.4).

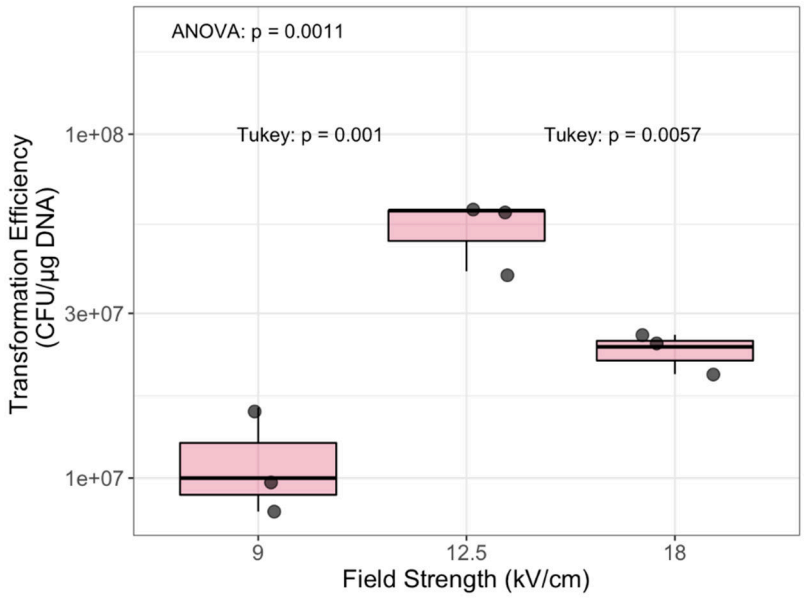


Figure 2. 4. Field strength optimisation.

Transformation efficiency as a function of electroporation field strength. Individual observations (*n* = 3) are shown as grey dots. ANOVA and post-hoc Tukey test p-values are shown.

2.4.5. Transformation efficiency is saturated at low DNA concentrations

Another important electroporation parameter is the amount of DNA that is added to competent cell aliquots for transformation. Previous reports have proposed that transformation efficiency in *C. necator* increases with increasing amounts of DNA, up to 1 µg[179]. This indicates that DNA concentration might be a limiting factor for efficient

electroporation. Thus far, 250 ng of plasmid DNA was used for all electroporations. When adding different amounts of DNA (1-500 ng) to the competent cell aliquots, we observed that the electroporation efficiency attained with our protocol is fully saturated even for the lowest tested amount (Figure 2.5). Notably, no significant differences in transformation efficiency were observed between 1 and 100 ng, and a drop was recorded when 200 or 500 ng were used. This is in direct contrast with other proposed methods and indicates that the amount of DNA is not a limiting factor for high electroporation efficiency with the proposed protocol (when electroporating plasmid pSEVA231).

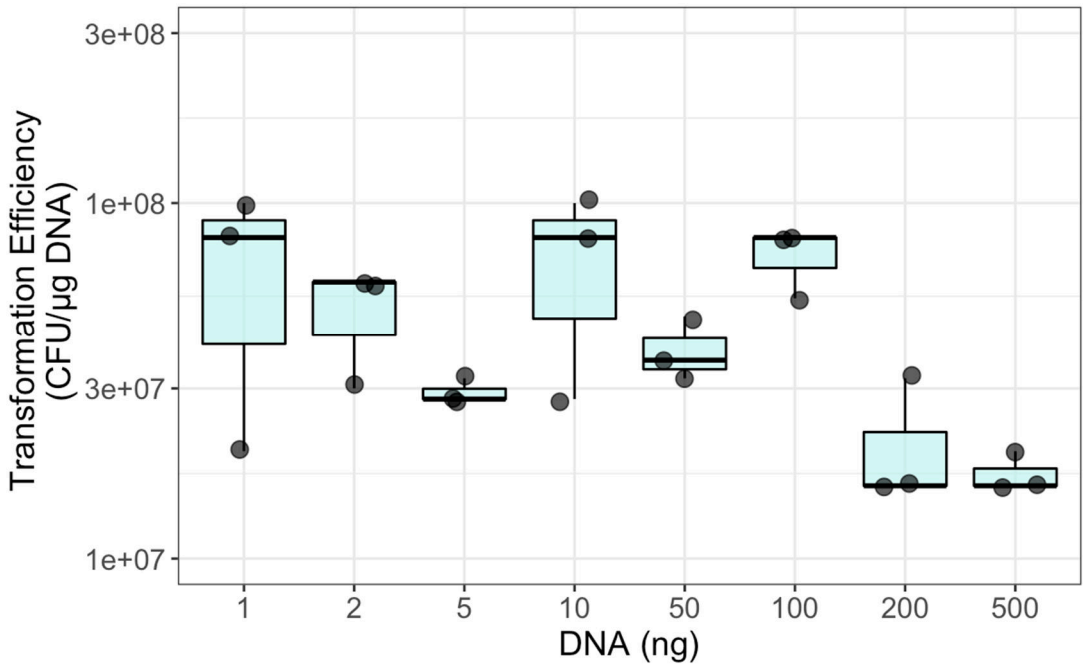


Figure 2. 5. Small amounts of DNA fully saturate transformation efficiency.

Transformation efficiency as a function of plasmid DNA mass (pSEVA231) added to competent cell aliquots. All protocol steps were performed as standard (see Sections 2.3 and 2.4.2). Individual observations (*n* = 3) are shown as grey dots.

2.4.6. Previously characterised deletion mutants do not improve transformation efficiency

Bacterial defence (immunity) mechanisms, such as restriction-modification systems, may limit the efficiency with which recombinant plasmids can be transformed. For model organisms such as *E. coli*, some common laboratory strains have been disarmed to increase their genetic tractability[188]. For *C. necator*, strains carrying deletions in restriction-modification systems have been reported to increase electroporation efficiency[189]. We tested whether one of these strains, namely *C. necator*Δ0006, could be transformed with higher efficiency than the wildtype using the optimised protocol. Specifically, the deleted H16_A0006 locus encodes the endonuclease subunit of a Type I restriction enzyme (KEGG Orthology ID K00153, similar in function to the product of the *hsdR* gene from *E. coli*[190,191]). Using the protocol optimised as described in Sections 2.4.1 to 2.4.5 (see Table S2.4 for a summary), we did not observe a significant difference in transformation efficiency between the two strains (Figure 2.6), indicating that the intervention of restriction-modification systems may not be as significant a limitation as previously reported in the context of this optimised protocol.

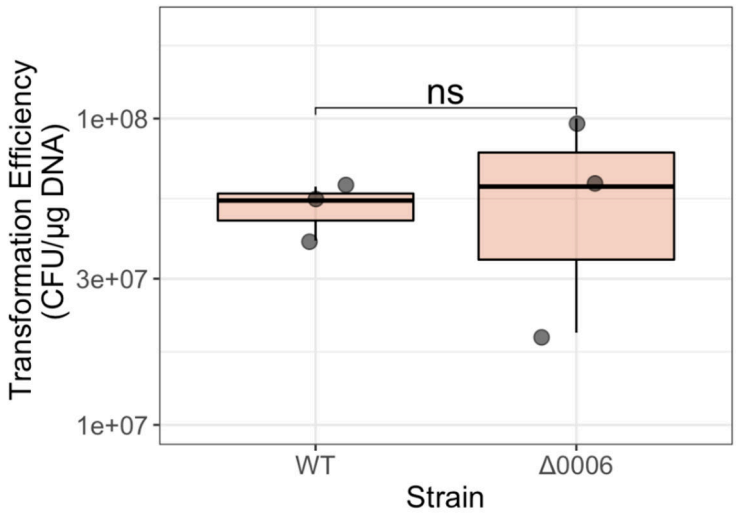


Figure 2. 6. Electroporation efficiency in wildtype and Δ0006 strains of *C. necator*.

Individual observations (*n* = 3) are shown as grey dots. ns = p>0.05, two-sample t-test.

2.4.7. Performance of the electroporation method with SEVA and pCAT standardised plasmid vectors

The plasmid used for protocol optimisation, pSEVA231, shares a common sequence syntax with other plasmids developed within the SEVA framework[177,178]. These plasmids are commonly used by laboratories worldwide to facilitate DNA assembly, collaboration, and standardised characterisation of DNA parts. We investigated whether other SEVA plasmids could also be transformed with high efficiency using our method (as summarised in Table S2.4). Plasmids pSEVA331 and pSEVA531 were used for this purpose. These share the same origin of replication (pBBR1), but encode chloramphenicol and tetracycline resistance cassettes, respectively. We observed variable transformation efficiencies between the three plasmids (Figure 2.7 A). High efficiency (>5x10^6 CFU/µg DNA) was observed for both pSEVA231 and pSEVA331. The tetracycline cassette-containing plasmid pSEVA531 was transformed with lower efficiency (~2x10^4 CFU/µg DNA).

For vectors carrying the pBBR1 origin of replication, Azubuike *et al*. proposed that the inefficient electroporation of *C. necator* could be due to the sub-optimal nature of the antibiotic resistance cassette[182]. Namely, they proposed that observed differences in transformation efficiency between pBHR1 and pBBR1-MCS2 plasmids were due to the kanamycin resistance gene on these vectors encoding different classes of aminoglycoside *O*-phosphotransferase enzymes (class I and class II, respectively), though both enzymes belong to subtype (a). The pCAT suite of standardised plasmids was designed to overcome this limitation. pCAT vectors were modelled on the SEVA architecture, and shown to transform with higher efficiency than previously reported with pBHR1 or pBBR1-MCS2, and to be compatible with both electroporation and heat-shock protocols[182].

We tested whether the transformation efficiency of these optimised vectors could be further improved by coupling the pCAT sequence design with our novel electroporation protocol. We tested the best-performing pCAT backbone plasmids conferring kanamycin (pCAT201), chloramphenicol (pCAT204) and tetracycline (pCAT206) resistance. For all three vectors, we attained a significantly higher electroporation efficiency than reported by Azubuike *et al*. (Figure 2.7 B). Notably, only plasmid pCAT204 was transformed with higher efficiency than its SEVA counterpart (Figure S2.2), indicating that standardised vectors from either suite could in practice be used interchangeably for highly efficient electroporation of *C. necator*. Additionally, sequence alignments showed that, though the DNA sequences of the kanamycin resistance cassettes on pSEVA231 and pCAT201 are not identical (Figure S2.3), they encode the same protein sequence (Figure S2.4).

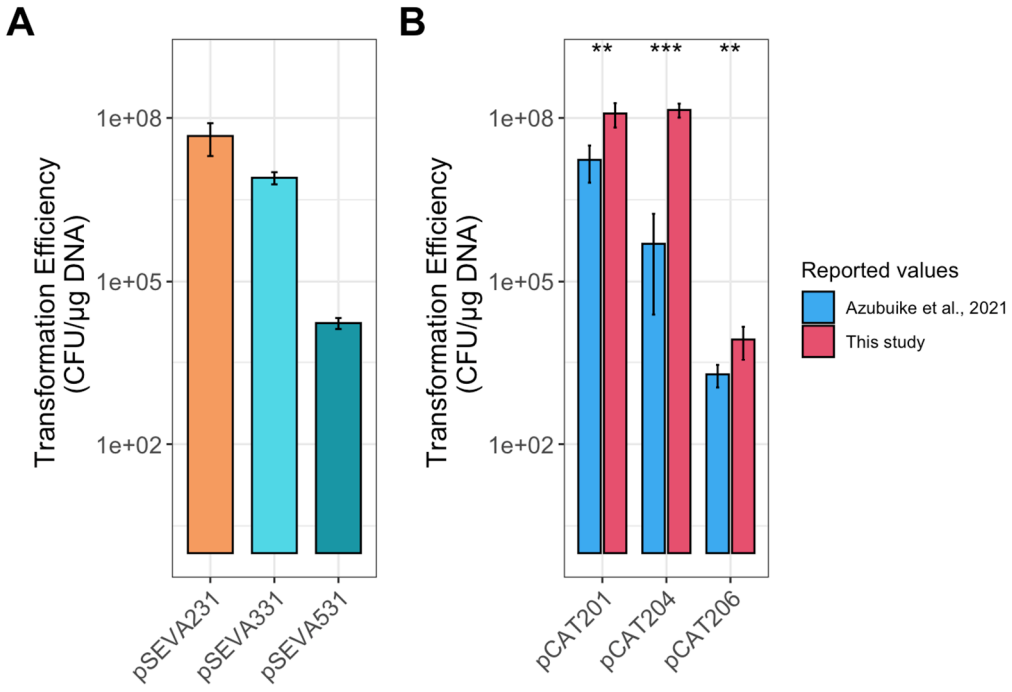


Figure 2. 7. Electroporation of standardised plasmids.

(A) Transformation efficiency for representative SEVA suite vectors pSEVA231 (Kan^R), pSEVA331 (Chl^R) and pSEVA531 (Tet^R), as obtained in this study using optimised protocol parameters. **(B)** Transformation efficiency values for pCAT vectors pCAT201 (Kan^R), pCAT204 (Chl^R) and pCAT206 (Tet^R) obtained in this study and reported in the literature. Bars indicate the mean ± one standard deviation of *n* = 3. *** = p<0.01; ** = p=<0.05; two-sample t-test.

2.4.8. Building and testing a DIY electroporator

A considerable drawback of electroporation is that it requires a specialised instrument which, though relatively inexpensive, may not be available in all research laboratories. A widely used alternative is heat-shock transformation, which has only been reported once for C. *necator*[182]. Though viable, the transformation efficiency obtained with the heat-shock method is substantially lower than with electroporation, meaning it may not be suitable for all applications. Moreover, a large amount of DNA (1 µg) is required for efficient heat-shock transformation, which may also limit the application of this method.

Where a commercial electroporator is not available and heat-shock transformation is not viable, another alternative is to construct a budget-friendly version of the electroporation instrument. This was first proposed as part of the "ElectroPen" project[185]. Briefly, the piezoelectric crystal contained in any widely available kitchen lighter can be used to generate a short electric pulse with an output voltage and exponential decay constant which fall within the permissible range for electroporation of bacterial cells.

We tested whether competent cells prepared using the optimised protocol could be electroporated with a DIY electroporator, constructed following the ElectroPen guidelines (Figure 2.8 A-D). The exact voltage output of the DIY instrument could not be measured due to the lack of a voltmeter or oscilloscope that could tolerate high voltages. Considering this, we tested the performance of the DIY instrument with two different sizes of commercially available cuvettes (0.1cm and 0.2cm gaps) to maximise the likelihood of a suitable field strength being generated. In both cases, transformants were recovered, with 0.1cm gap cuvettes leading to the highest transformation efficiency of (1.67±0.42) $x10^3$ CFU/µg DNA (Figure 2.8 E).

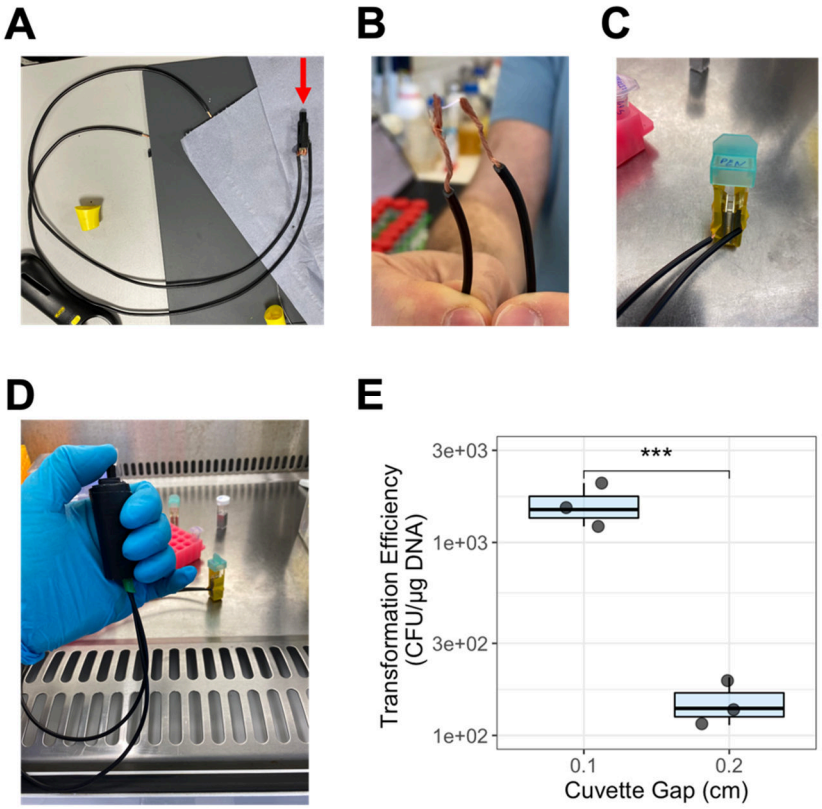


Figure 2. 8. Building and testing a DIY electroporator.

(A) Following the ElectroPen design, copper wires were attached to the piezo igniter obtained from a gas kitchen lighter (red arrow). **(B)** High-voltage output was confirmed by holding the ends of the copper wires close together and observing the appearance of a spark between them when operating the hammer-action striking mechanism. **(C)** The ends of the copper wires were attached to a commercially-available electroporation cuvette with electrical tape. **(D)** Electroporation was performed by striking the hammer on the piezo igniter, which was insulated using heat shrink tubing and electrical tape, which also enabled ease of hand-held operation. **(E)** Transformation efficiencies obtained by electroporation with the DIY instrument using different size cuvettes. Individual observations (n = 3) are shown as grey dots. *** = $p<0.01$, two-sample t-test.

2.5. Discussion and Conclusions

The work outlined in this Chapter delivers a streamlined protocol for high-efficiency electroporation of *C. necator*. We identified the starting protocol parameters, chiefly buffer composition, by taking inspiration from a literature reporting of optimal electroporation conditions for the Gram-negative bacterium *Pseudomonas aeruginosa*[184]. By investigating key factors, namely cell density, outgrowth time and medium composition, field strength, and DNA concentration, we delineated the range of acceptable parameters for optimal electroporation. We report what is, to the best of our knowledge, the highest electroporation efficiencies ever attained for *C. necator*. However, since a range of different plasmid vectors have been used for development and benchmarking of *C. necator* electroporation protocols (Table S2.1), direct comparisons are not possible and indeed could be misleading for protocol users. We have attempted to address this limitation by demonstrating the superior performance of our protocol with the SEVA and pCAT suite of standardised vectors, which can be universally applied for advanced genetic and metabolic engineering strategies in this bacterium. Specifically, we have demonstrated that plasmid vectors from either suite can be used interchangeably and that they direct the expression of the same gene product for kanamycin resistance, though the DNA sequence of this marker gene is not identical.

There are three key points where our protocol diverges from previous reports in the literature. The first is the composition of the electroporation buffer. To our knowledge, 1mM $MgSO_4$ has not been previously used for transformation of *C. necator*. As mentioned, we chose this buffer composition noting that it had been used for

electroporation of another Gram-negative bacterium[184]. We specifically took inspiration from this report because the authors tested a range of electroporation buffers, including water, 10% glycerol, 0.2-0.3 M sucrose, and HEPES buffer, all of which have been proposed as optimal for *C. necator* in various studies. They found that these buffers led to significant cell lysis during competent cell preparation. In contrast, when using $MgSO_4$ in the transformation buffer the structural integrity of cells could be maintained, and no adverse effects on electroporation were observed. Though our study did not directly investigate the effects of buffer composition on cell lysis, the superior performance of the $MgSO_4$ buffer over previous reports suggests that the loss of cell structural integrity during competent cell preparation may also be a limiting factor for efficient electroporation of *C. necator* in previously reported protocols. This hypothesis could be directly assessed in future experiments. Additionally, it is possible that varying the concentration of $MgSO_4$ in the transformation buffer, or altering its chemical composition altogether (neither of which were explored in this study), may yield different optimal parameters for electroporation and further improve transformation efficiency.

Secondly, we propose the use of near-saturated SOB cultures (OD ~ 5) for competent cell preparation. Previous reports have looked no further than an OD of ~1, and have identified an optimum in the range of 0.4-0.8, in line with what is known to be optimal for *E. coli*. By contrast, our results show that harvesting high-density cultures can lead to a significant increase in transformation efficiency. We report that electroporation efficiency increases linearly as a function of the optical density of cultures at the time of cell harvest. Though not previously described, this observation has precedent in the literature. In their early report, Park *et al*. showed that high-density competent cell

aliquots led to higher efficiency. Tee *et al*. similarly showed that increasing the number of cells used for electroporation by concentrating them 80-fold drastically improves transformation efficiency. However, they argued that harvesting cultures at OD ≥ 0.8 was disadvantageous, due to an observed drop-off in transformation frequency. Here, we did not observe any trade-offs between cell density, transformation efficiency and transformation frequency, evidencing the fact that the absolute number of cells used for electroporation (and their structural integrity) is a key parameter for high-efficiency transformation of *C. necator*. We did not investigate whether we could improve efficiency by concentrating competent cells even further; a standard ~20-fold concentration factor was consistently applied. This may be a promising avenue for future improvements. However, it is possible that having hyper-dense competent cell aliquots (by concentrating competent cells further) may lead to false positives appearing on transformation plates – if too many cells are plated on a single selective agar plate, some cells may experience a local concentration of kanamycin which is below the threshold concentration required for selection. As a result, colonies may appear on the agar plate that do not carry the plasmid of interest.

Lastly, in contrast to other studies, we do not observe that transformation efficiency increases when increasing amounts of DNA are used for electroporation, i.e. the DNA mass used for electroporation is not a limiting factor for high efficiency. Instead, we report that efficiency is saturated even at the lowest tested DNA concentration of 20 pg/µl, i.e. when 1 ng of plasmid DNA is added to 50 µl of competent cells. We hypothesise that this may be due to the increased number of structurally intact competent cells within the aliquot relative to other protocols. This is consistent with previous literature reports, where increasing the concentration of cells within the

aliquot whilst keeping the DNA mass constant led to higher electroporation efficiency[180,179]. The ability to use small amounts of DNA for electroporation is advantageous. For instance, it may be difficult to obtain large amounts of plasmid DNA when isolating low- or medium-copy plasmids from *E. coli* strains.

Interestingly, when ≥200 ng of plasmid DNA was used, a drop-off in transformation efficiency was recorded. This may be because adding large amounts of plasmid DNA, which is resuspended in water, to the competent cell aliquot significantly dilutes the transformation buffer, thereby altering the properties of the medium. This indicates that maintaining the relative concentrations of cells, $MgSO_4$, and glycerol within the competent cell aliquot is important. The total volume of liquid in the electroporation cuvette may also be an important factor to consider.

There are other minor points of divergence between our protocol and previous reports. For instance, we did not reproduce the observation that adding fructose to the outgrowth medium improves electroporation efficiency[179]. Mutant strains with deletions in restriction modification systems did not display significantly higher transformation efficiency than the wildtype, though this may depend on the exact sequence (and methylation pattern) of the plasmid which is being transformed, as proposed by Xiong *et al*. (this concept is explored further in Chapter 3)[189]. Most notably, though field strength is temperature dependent and most other literature reports mention cooling cuvettes and reagents, we chose to perform all experiments at room temperature (22 ºC). This was found to work best in the *P. aeruginosa* study[184], and delivered higher electroporation efficiencies than previously reported for

C. necator in our early experiments. Removing the requirement for ice-cold reagents and cuvettes further streamlines the protocol.

Though successful, our investigation has several limitations. Firstly, the granularity of optimisation for certain parameters (such as field strength) was restricted by instrument availability and technical specifications, such as the limited range of settings available on the Bio-Rad MicroPulser electroporator. Similarly, we were not able to test the performance of our protocol with plasmids pBHR1 and pBBR1-MCS2, as these could not be sourced. This prevents a direct comparison between our method and other protocols other than in the context of the pCAT or SEVA vectors, which carry the optimal kanamycin resistance cassette also present on pBHR1. It would be interesting to test whether our improved protocol could compensate for the sub-optimal nature of the kanamycin resistance gene encoded on the pBBR1-MCS2 backbone. We were also unable to test protocol performance with plasmids of different sizes. It is difficult to test the relationship between vector length and transformation efficiency without introducing confounding effects from sequence composition and burden. Therefore, the suitability of our protocol for the transformation of large vectors would have to be assessed on a case-by-case basis. In our laboratory, we have been able to transform vectors of up to 18.5 kb, with strains carrying these plasmids having been used in published research[117].

A final consideration is that the transformation buffer used in our protocol includes glycerol as a cryoprotectant, meaning that competent cells can be frozen for long-term storage. This is an advantage over other optimised methods where “fresh” competent cells had to be prepared immediately before electroporation. We have gathered

evidence that competent cells can be used with no drop in efficiency for at least one year, though this was not investigated systematically and is therefore not reported conclusively in this chapter.

In summary, we propose an electroporation protocol with remarkable simplicity, robustness, and efficiency which we expect will be readily established as the method of choice in the field. We envision further optimisation and benchmarking, for instance investigating protocol performance with plasmids of different sizes and sequences, will be undertaken in the wider community of researchers working with *C. necator*. Implementing multi-factorial design of experiments (DOE) strategies may yield improvements beyond what we report in this study, where a one factor at a time optimisation method was implemented. Importantly, this protocol was instrumental to the work presented in all subsequent chapters of this thesis and has therefore already facilitated significant advances in the development of *C. necator* towards sustainable biomanufacturing.

2.6. Chapter 2 Appendix – Supplementary Information

Table S2. 1. Transformation efficiency of *C. necator*.

*Where possible, the standard error of the mean has been recorded. [a)] The competent cells were frozen and thawed before transformation. [b)] A modified *C. necator* strain was constructed to enhance transformation efficiency. c) Engineered plasmids were used to enhance electroporation efficiency. NA = not reported.

Plasmid		Optical Density (OD)	Electroporation Conditions		Transformation Efficiency	Ref.
Name	Size (kb)		Transformation Buffer	Field Strength (kV cm^{-1})	(cfu µg DNA^{-1})*	
pBS29-P2-gfp	8.7	NA	0.3M sucrose	25	(4.04 ± 0.07) x 10^3	187
pBHR1	5.3	NA	0.3M sucrose	25	(4.70 ± 0.02) x 10^2	187
pKK-II 89-6	7.8	NA	0.3M sucrose	25	(1.20 ± 0.04) x 10^2	187
pCUP3	7.6	0.5	ddH_2O	25	(1.11 ± 0.08) x 10^3	192
pJRD215	10.2	0.5	ddH_2O	7.5	3.4 ± 0.6	192
pKT230	11.9	0.8	10% glycerol	7.5	8.0 x 10^{1} [a)]	180
pBHR1	5.3	0.6	0.2M sucrose	11.5	(3.86 ± 0.29) x 10^5	179
pBBR1-rfp	NA	0.6-08	10% glycerol	11.5	4.00 x 10^{4} [b)]	189
pCAT201	3.2	0.4-0.8	10% glycerol	12.5	(1.70 ± 0.24) x 10^7	182
pMTL71101	5.3	0.2-0.3	1mM HEPES, 10% glycerol	12.5	4.70 x 10^5	193

Table S2. 2. Bacterial strains used in this study.

Strain	Purpose	Genotype
E. coli DH5α	Plasmid DNA storage and isolation	F– φ80*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *rec*A1 *end*A1 *hsd*R17(rK^-, mK^+) *pho*A *sup*E44 λ–*thi*-1 *gyr*A96 *rel*A1
C. necator H16	Electroporation testing	Wild-type *C. necator* strain
*C. necator*Δ0006	Electroporation testing	Type I restriction modification system mutant of *C. necator* (KEGG locus ID H16_A0006)[189]

Table S2. 3. Plasmids used in this study.

*Plasmid maps and sequences can be accessed via hyperlink, by clicking on the individual plasmid names.

Plasmid*	Description	Ref.
pSEVA231	pBBR1 OriV-Rep; MCS; Kan^R	177
pSEVA331	pBBR1 OriV-Rep; MCS; Cm^R	177
pSEVA531	pBBR1 OriV-Rep; MCS; Tc^R	177
pCAT201	pBBR1 OriV-Rep; MCS; Kan^R	182
pCAT204	pBBR1 OriV-Rep; MCS; Cm^R	182
pCAT206	pBBR1 OriV-Rep; MCS; TcR	182

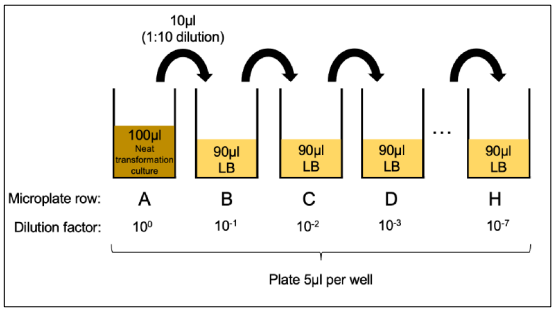
Table S2. 4. Optimal parameter values identified in this study.

This table is provided for ease of reference. For a step-by-step description of the electroporation protocol, see Section 2.3. [a] When electroporation is performed at room temperature. [b] For a 3.1kb plasmid, corresponding to 0.5-50 fmol of double-stranded DNA. [c]Recovery times longer than 4 h were not tested.

Protocol parameter	Value
Growth medium for electrocompetent cell preparation	SOB
OD at which cells are harvested	5
Electroporation cuvette gap size	0.2 cm
Electroporation field strength[a]	12. 5 kV/cm
Amount of DNA added[b]	1-100 ng
Outgrowth medium	LB or SOB
Outgrowth time[c]	> 2 h

Spot microdilutions are performed in 96-well plate format:

1. Cultures are diluted 1:10 along each of 8 plate columns.
2. A multi-channel pipette is then used to spot 5uL from each column on an agar plate. The spots are left to dry completely before inverting the plates and placing them in the incubator.
3. Transformation efficiency can then be accurately quantified by counting the cells of the first countable spot on the agar plate.



Example plate:

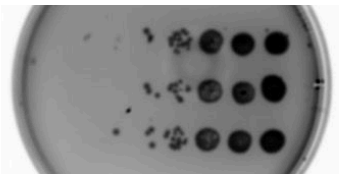


Figure S2. 1. Spot microdilution procedure for quantifying transformation efficiency.

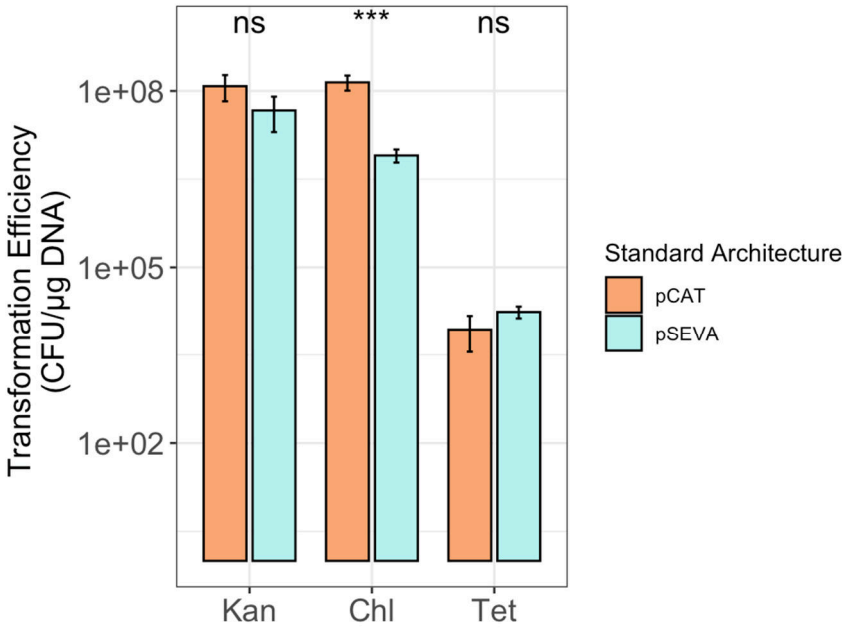


Figure S2. 2. Comparison of protocol performance with pSEVA and pCAT plasmid vectors.

Bars indicate the mean ± one standard deviation of *n* = 3. Individual observations (*n* = 3) are shown as grey dots. *** = p<0.01, ns = not significant, two-sample t-test.

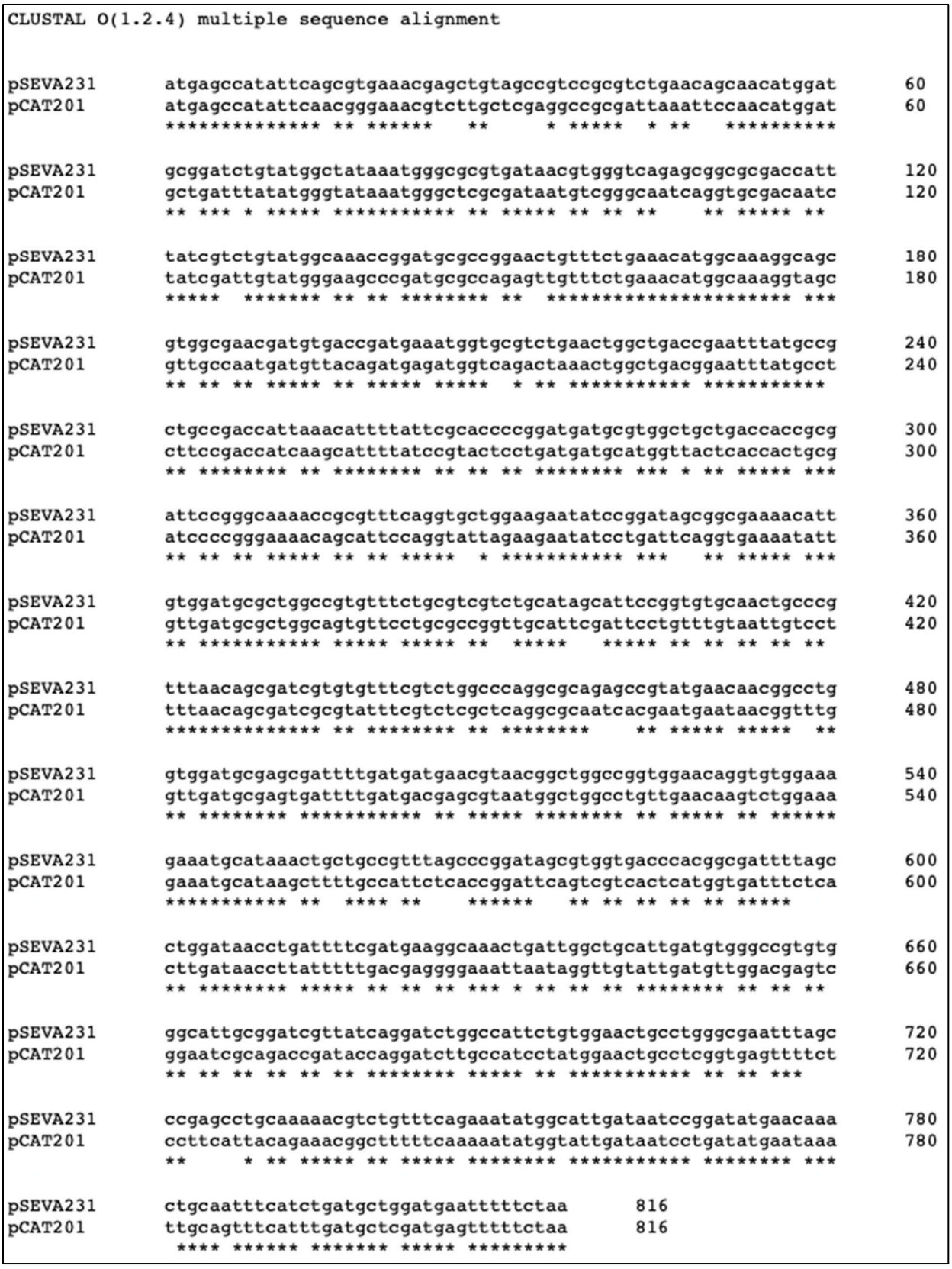


Figure S2. 3. DNA sequence alignment of *neo* (*kan*^{R}) genes from pSEVA231 and pCAT201 vectors.

Sequence alignment performed with Clustal W[194]. Matching bases are annotated as '*'.

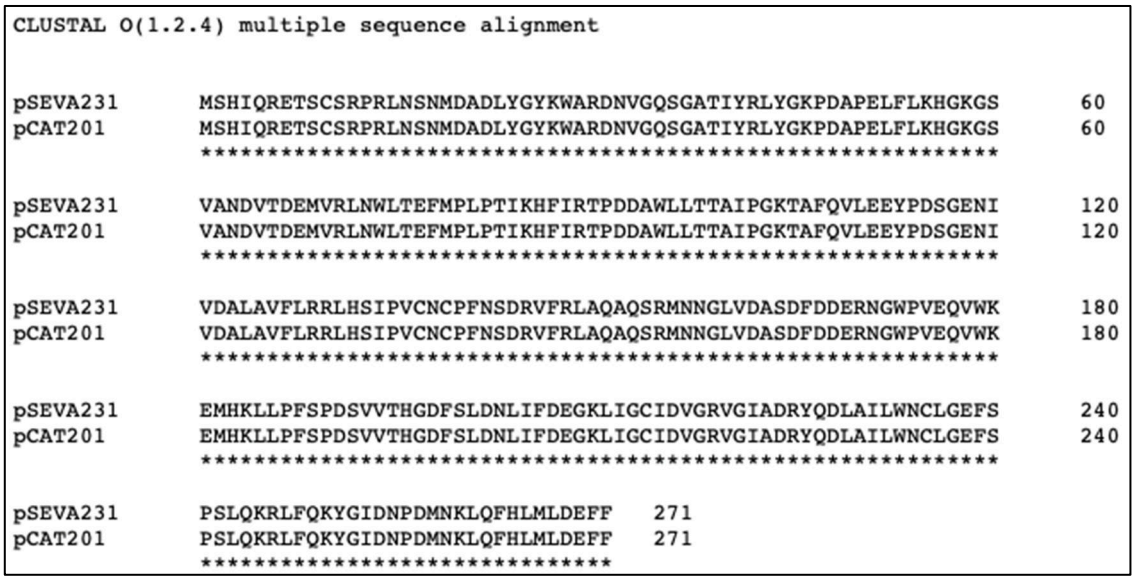


Figure S2. 4. Protein sequence alignment of *neo* (*kan*^{R}) gene products from pSEVA231 and pCAT201 vectors.

Sequence alignment performed with Clustal W[194]. Matching bases are annotated as '*'.

3. SIBR-Cas enables streamlined genome editing in *Cupriavidus necator*

3.1. Aims and Objectives

The inability to perform gene knockouts rapidly and reliably in *C. necator* hinders its development as a platform chassis for synthetic biology and metabolic engineering. To address this limitation, we aim to develop a CRISPR-based system to provide a simplified, streamlined, and efficient genome editing tool for this bacterium. In doing so, we also aim to characterise and demonstrate the use of a novel inducible gene expression system, which can be used as a regulatory module in diverse biotechnological applications.

3.2. Statement of collaborative contributions

This work was conducted in collaboration with Dr Enrico Orsi (Technical University of Denmark) and Dr Constantinos Patinios (Helmholtz Institute for RNA-based Infection Research). Dr Orsi led the characterisation of SIBR-Cas9 in *C. necator*, which is included in this chapter for completion though it is not its main focus. As co-inventor of SIBR-Cas, Dr Patinios contributed to project ideation and management, most prominently to the design of the SIBR2.0 system. I personally developed the SIBR-Cas12 and SIBR2.0-Cas12 systems as outlined below, including the characterisation of constitutive and inducible CRISPR-Cas12a targeting efficiency, implementation of the SIBR2.0 system, and its application to mediate efficient genome editing.

3.3. Introduction

Synthetic biology and metabolic engineering broadly aim to install novel and improved capabilities in biological organisms. The genetic circuitry required to execute these programs is most often encoded on recombinant plasmids, which are delivered into host cells and maintained through cultivation in selective media. In some cases, genomic manipulation of host organisms is required to optimise their performance. For instance, it may be necessary to delete certain genes to eliminate competing metabolic pathways and drive carbon flux towards a product of interest. Chromosomal integration (knock-in) of heterologous genes may also be necessary to improve functional parameters. A wide range of genome engineering tools have been developed in recent years, enabling increasingly complex applications in microbial biotechnology, sustainable bioproduction and beyond (see refs. [63] and [195] for comprehensive up-to-date reviews). However, the application of these advanced genome editing methods has broadly been limited to well-characterised model bacteria such as *E. coli*. Though universal in their scope, these tools are often not portable, i.e. their functionality in bacterial hosts other than the one for which they were developed is limited.

Metabolic engineering of non-model organisms such as *C. necator* requires the development of tailored molecular tools. Despite recent advances in synthetic biology methods for *C. necator*[133], its genetic manipulation remains challenging, partially due to the dearth of effective genome editing protocols. To date, three methods have been developed for genome editing in this bacterium. The first is based on the use of suicide plasmids to mediate knock-outs and knock-ins at specific loci[196–200]. These vectors

carry origins of replication which prevent their propagation in *C. necator*. When transferred into the bacterium by electroporation or conjugation, the plasmids integrate into a target locus via homologous recombination (HR). Chromosomal integration via a single crossover event can be selected for, most commonly by using an antibiotic selection marker. An additional marker is included in the vector to counter-select for a second crossover event, where the plasmid is excised from the chromosome, leaving behind either an edited or a wild-type locus (Figure 3.1 A). Two counterselectable markers, *sacB*[197–201] and the Cre/*loxP* system[202], have been successfully used in *C. necator*. Mobile genetic elements have also been applied to genome editing in *C. necator*. The Tn5 transposon, which randomly integrates into the bacterium's chromosome, has been used to mediate both gene deletions[203,204] and insertions[205]. More recently, the RalsTron system was developed, which enables gene deletions via site-specific chromosomal insertion of a group II intron[206]. Finally, there is a single literature report of a CRISPR-Cas9 editing method for *C. necator*[207]. Here, an editing plasmid containing homology arms and an inducible CRISPR-Cas9 cassette is transformed into *C. necator*. The CRISPR-Cas9 complex can mediate efficient counterselection of WT genomes following a double crossover event (Figure 3.1 B). Specifically, when the expression of the CRISPR-Cas9 complex is induced, any cells which have not undergone HR at the target locus and still retain a copy of the WT allele will be targeted for genome cleavage. As *C. necator* is not able to undergo non-homologous end joining[207], double-strand breaks cannot be alternatively repaired.

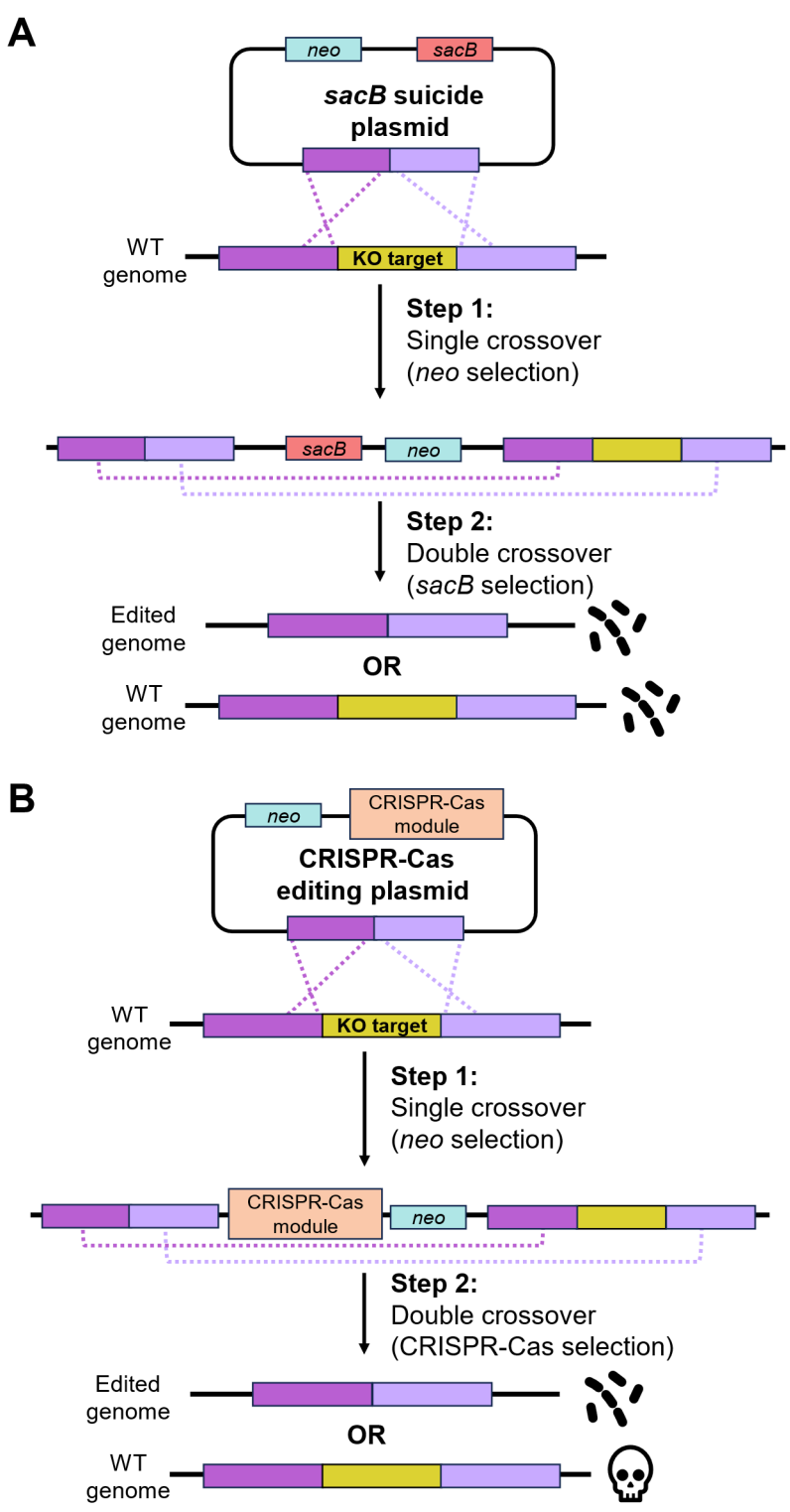


Figure 3. 1. HR-based methods for genome editing in *C. necator*.

(A) Knock-out (KO) of a target locus using suicide plasmids with *neo* and *sacB* selectable markers. After transformation, cells are grown in kanamycin-containing medium to select for the integration of the suicide plasmid onto the *C. necator* chromosome via a single crossover event (for simplicity, only upstream integration is shown, though downstream integration is also possible). To select for double crossover events, transformants are cultured in sucrose-containing medium without kanamycin. Any cells retaining the chromosomally-integrated copy of the plasmid (including the *sacB* marker) will not survive due to the accumulation of the cytotoxic compound levans[208]. After the double crossover event, (theoretically) a 50-50 mixture of edited and WT genomes will be recovered. **(B)** KO of a target locus mediated by CRISPR-Cas counter-selection. Following transformation, cells are grown in antibiotic-(kanamycin) containing medium to retain the editing plasmid. Sufficient cultivation time is allowed for the cells to undergo HR at the target locus (sequential single and double crossover events). Both crossover events are selected for in a single step by inducing the expression of the CRISPR-Cas module on the editing plasmid, such that any cells retaining the WT locus sequence cannot survive. Theoretically, only cells with edited genomes can proliferate following CRISPR-Cas counterselection.

Though useful, these genome editing protocols have significant limitations, which in turn limit the scope of synthetic biology and metabolic engineering efforts in this organism. Methods relying on suicide vectors and mobile genetic elements operate with low editing efficiency and long timelines, requiring two to three weeks for isolation of edited strains. The CRISPR-Cas9 method described by Xiong *et al*. also requires at least two weeks, though it delivers far superior editing efficiencies in the range of 78-100%[207]. However, the plasmid architecture and editing protocol are not standardised, and the characterisation of the arabinose-inducible CRISPR-Cas9 system is very limited. No details on inducible targeting efficiency are provided. Moreover, the editing plasmids developed in this study are not readily available. These factors limit reproducibility, widespread acquisition, and further development of the technology.

Here, we aim to overcome these limitations by developing a novel tool to perform gene knockouts in *C. necator* using Self-splicing Intron-Based Riboswitch-Cas (SIBR-Cas). The method uses an inducible self-splicing intron to tightly control the expression of a CRISPR-Cas complex, which can be used for counterselection of the WT strain following homologous recombination[209] (for ease of reference, a schematic illustration of the SIBR-Cas technology is provided in Figure S3.1.1). This genetic control framework is designed to be organism-independent, as it does not require the expression of any additional endogenous or heterologous transcription factors or enzymes. It is therefore primed for use and further development in non-model wildtype bacterial species such as *C. necator*, where the expression of complex heterologous molecular machinery can be challenging. A key feature of SIBR-Cas is that it enables distinct temporal separation of HR and CRISPR-Cas counterselection, which is an

important consideration for successful editing, particularly in bacterial species with slow or inefficient HR systems.

We have developed and extended the standardised SIBR-Cas architecture for use in *C. necator*, demonstrating the inducible expression of functional CRISPR-Cas9 and CRISPR-Cas12 complexes that can target the bacterial chromosome for cleavage at specific loci. We successfully applied SIBR-Cas to genome editing, providing an efficient, streamlined, and standardised protocol that can be executed in a single working day. In doing so, we provide a step-by-step protocol that can be followed to implement the SIBR-Cas toolbox to other wildtype bacterial species where genomic manipulation remains challenging.

3.4. Materials and Methods

Bacterial strains, plasmids, and culture conditions

All bacterial strains and plasmids used in this study are listed in Table S3.1.1 and Table S3.1.2, respectively. Unless otherwise stated, plasmids were isolated from *E.* coli DH5α, using the NEB Monarch® Miniprep Kit according to the manufacturer's specifications. Plates for bacterial growth were LB agar. Liquid media was LB or SOB, as indicated. Where appropriate, antibiotics were used at 100 µg/ml (kanamycin) and 50 µg/ml (rifampicin). All *E. coli* strains were incubated at 37 °C with vigorous shaking (200 rpm). *C. necator* strains were incubated at 30 °C with moderate shaking (150 rpm), unless otherwise stated. Optical density at 600 nm (OD) of bacterial cultures was measured using a UV-1800 UV/Vis spectrophotometer (Shimadzu).

Flow cytometry for single-timepoint fluorescence measurements

Fluorescence measurements were performed to measure the gene expression output of the lacUV5 promoter (P_{lacUV5}) in *C. necator*. Strains carrying test and control plasmids were cultured overnight (16-18 h) in selective LB medium (100 µg/ml kanamycin). All cultures used in these experiments had a total volume of 5 ml, in 50 ml conical centrifuge tubes. Overnight precultures were used to inoculate fresh cultures in selective mineral medium (M9), with 20 mM fructose as sole carbon source, at an OD of 0.05-0.1. Cells were grown to mid-exponential phase (OD 0.3-0.6), at which point 1 ml of each culture was transferred to 1.5 ml microcentrifuge tubes. Cells were pelleted by centrifugation and washed twice in phosphate buffered saline (PBS). After the final wash, cell pellets were resuspended in 1 ml PBS, then diluted in PBS to an OD of 0.01. The cells were analysed using a BD FACSCalibur flow cytometer (BD Biosciences). mRFP fluorescence was measured with a 488 nm laser and a 585/42 nm emission band-pass filter (corresponding to the instrument's FL2 channel). The voltage of the FL2 detector was set to 705 V and the amplitude gain was adjusted to 1.0. At least 100,000 events were collected for each sample. Flow cytometry data was analysed using the proprietary FlowJo software (BD Biosciences).

Theophylline toxicity assay

Theophylline toxicity in *C. necator* was quantified via a minimum inhibitory concentration assay at the microplate scale. Cells were cultured overnight (16-18 h) in LB medium (5 ml cultures in 50 ml conical centrifuge tubes). Cells from saturated

overnight cultures were pelleted by centrifugation, spent medium was discarded, and the cell pellet was resuspended in fresh LB medium, adjusting the cell density to an OD of 1. Cells were used to inoculate fresh cultures in a 96-well microplate by diluting them 1:10 into the plate wells, giving a starting OD of 0.1. The total volume of each well was 200 µl. Theophylline was added to microplate wells by diluting a stock solution to give working concentrations in the range of 0-20 mM, as indicated. The microplate was incubated at 30 °C with double orbital shaking in a Spark microplate reader (Tecan), with OD measurements taken automatically at 15 min intervals over a period of 24 h.

Assembly of targeting and editing plasmids

A custom Python software was used to design the crRNA spacer sequences for each target locus (see Section 3.8 for details). The final sequences were synthesised as oligonucleotides (Tables S3.2.1 and S3.2.2). Forward and reverse oligonucleotides were mixed in equimolar amounts and annealed by incubating the mixture at 95 ºC for 5 min, followed by cooling at room temperature (22 ºC) for 2 h. The annealed double-stranded oligonucleotides were assembled into SIBR plasmids via a Golden-Gate protocol, as described previously[209]. SIBR2.0 constructs were assembled by PCR-amplifying the SIBR intron sequence and the universal SIBR plasmid backbone from pSIBR001 and pSIBR005, respectively. All PCRs were conducted using Q5® High-Fidelity DNA polymerase (New England BioLabs, NEB), following the manufacturer's instructions. PCR products were checked by agarose gel electrophoresis and purified using the Monarch® PCR & DNA Cleanup Kit (NEB). Gibson assembly of double-stranded DNA fragments was carried out using the NEBuilder® HiFi DNA Assembly

(NEB). To assemble the editing plasmids, homology arms covering ~1 kb upstream and downstream of the target locus were amplified from *C. necator* genomic DNA. PCR products were checked and purified as described above. Gibson assembly of homology arms into the SIBR plasmids was performed as described previously[209]. For all plasmids constructed in this study, correct assembly was confirmed via Sanger sequencing. Editing plasmids were additionally checked by full-plasmid Nanopore sequencing.

CRISPR-Cas targeting and editing assays

To measure the targeting efficiency of CRISPR-Cas9 and CRISPR-Cas12a complexes in *C. necator*, the transformation efficiency of non-targeting and targeting versions of SIBR plasmids was quantified. Following confirmation of successful assembly, plasmids were electroporated into *C. necator*. Electrocompetent cells were prepared and transformed following the protocol outlined in Chapter 2. Recovery was performed in 1 ml cultures within 1.5 ml microcentrifuge tubes for 2-4 h. Transformation efficiency was quantified as CFU/µg DNA via spot microdilution on selective LB agar plates. For inducible targeting assays, dilutions were also spotted on selective plates with 5 mM theophylline. For editing assays, the recovery step in the electroporation protocol was extended to 8 h, whilst the volume was kept constant at 1 ml. Transformation efficiency was quantified by spot microdilution. For test conditions where efficiency was low, the entire 1 ml transformation culture was pelleted by centrifugation and spread on selective LB agar plates (± 5 mM theophylline) to accurately quantify transformation efficiency. For each plasmid and condition, editing efficiency was quantified by colony PCR (cPCR) using DreamTaq®

DNA polymerase (ThermoFisher Scientific), following the standard protocol. PCR primers used to amplify the *acoC* locus were 5'-ATCTGCCACGCGCTAACC-3' (forward primer) and 5'-GAATTGCCTGTTGATCTCGGTC-3' (reverse primer). A maximum of 16 colonies were analysed for each replicate, plasmid, and condition. For all assays, transformation plates were incubated at 30 ºC for 48 h before single colonies could be counted and genotyped by cPCR, as required.

Plasmid curing assays

To investigate plasmid loss following the genome editing, single colonies corresponding to deletion mutants (as confirmed by cPCR) were picked from editing assay plates and precultured in 5 ml selective LB medium (100 µg/ml kanamycin) in a 50 ml conical centrifuge tube. Following preculture, test cultures for each curing condition were inoculated in 5 mL of the appropriate medium, as indicated, in 50 mL conical centrifuge tubes. Cells were grown to saturation in each condition and serially passaged every ~16 h by diluting the cultures 1:100. At each passage, the total number of cells in the culture was quantified by spotting serial dilutions on non-selective solid medium (LB agar). Plasmid-bearing (kanamycin-resistant) cells were analogously quantified on selective plates (LB agar with 100 µg/ml kanamycin).

3.5. Results

3.5.1. Active CRISPR-Cas12a and CRISPR-Cas9 complexes are readily expressed in *C. necator*

First, we investigated whether functional CRISPR-Cas complexes could be expressed in *C. necator* from plasmids following the canonical SIBR architecture. The simplest demonstration of CRISPR-Cas activity is provided by targeting assays, where both the Cas nuclease and its cognate CRISPR-derived RNA (crRNA) guide are constitutively expressed. In the SIBR framework, the lacUV5 promoter is used to drive constitutive gene expression of *Cas12a*[209]. We confirmed that this promoter is suitable for use in *C. necator* using a fluorescent reporter gene (Figure S3.1.2).

Next, we built versions of the canonical pSIBR005 plasmid, where constitutive *Cas12a* expression is paired with either a non-targeting or a targeting spacer within the crRNA guide (see Table S.3.1.2 for detailed plasmid descriptions and sequences). We modified the non-targeting spacer sequence from the default sequence present on the universal SIBR plasmids, as its seed region was found at multiple locations on the *C. necator* genome which could lead to off-target effects. We chose to target the *phaC* locus on the *C. necator* genome (KEGG ID H16_A1437), a non-essential gene within the *phaCAB* operon, which encodes a poly(3-hydroxybutyrate) polymerase necessary for PHA synthesis[137]. A custom Python script was used to design suitable targeting crRNA spacers for this locus. Details on how spacer sequences were designed and selected are provided in Section 3.8. Spacer sequences for the *phaC* locus are provided in Table S3.2.1. For the targeting assay, we built a total of four versions of

the pSIBR005 plasmid, that differed only in the sequence of the crRNA spacer (Figure 3.2 A). We tested three distinct targeting spacer sequences for the *phaC* locus (T1, T2 and T3, as shown in Figure 3.2. B), as different spacers may lead to different targeting efficiencies, an important consideration for CRISPR-mediated genome editing[210].

Targeting efficiency was quantified as the difference in transformation efficiency between non-targeting and *phaC*-targeting versions of the pSIBR005 plasmid, using a *phaC* deletion mutant strain (Δ*phaC*) as a negative control. Transformation efficiency was quantified as the number of colony forming units recovered on selective plates following electroporation of SIBR plasmids. For the best-performing targeting spacers (T1 and T3), we observed a significant difference in transformation efficiency ($\geq 10^3$ CFU/µg DNA) between non-targeting and targeting plasmids, consistent with a functional CRISPR-Cas12a complex being constitutively expressed. This difference was not observed for the Δ*phaC* strain, further confirming the specificity of targeting (Figure 3.2 C-D). Additionally, using an analogous experimental procedure we demonstrated that functional CRISPR-Cas9 complexes can also mediate efficient targeting in *C. necator* (Figure S3.1.3).

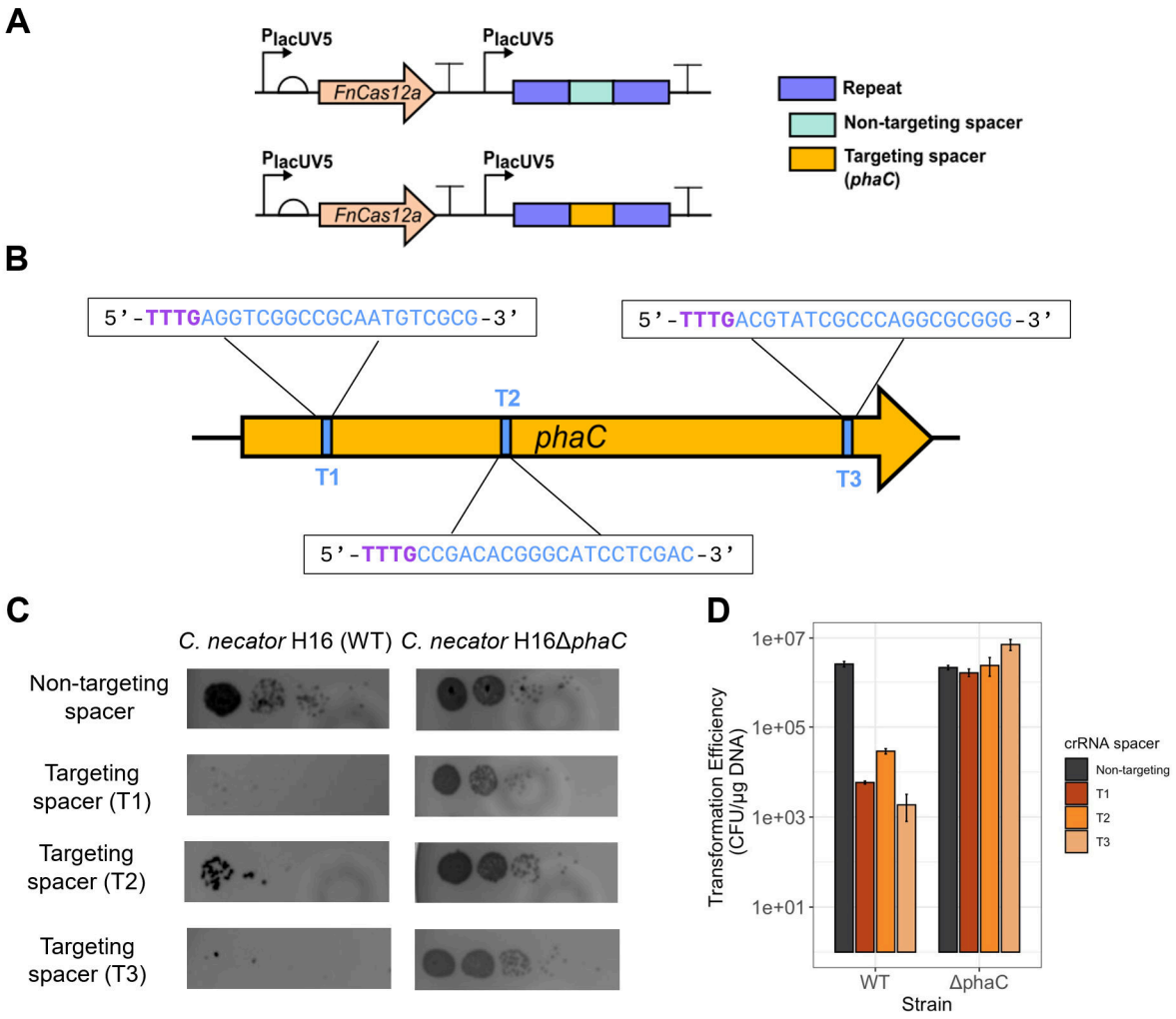


Figure 3. 2. Constitutive targeting of the *phaC* locus by an active CRISPR-Cas12a complex.

(A) Variants of the pSIBR005 plasmid used for the targeting assay. The constitutive P_{lacUV5} promoter drives expression of both *Cas12a* and the crRNA cassette, which contains either a non-targeting spacer or a *phaC*-targeting spacer. Three versions of the latter were tested (T1, T2 and T3). In the assay, targeting efficiency can be quantified as the difference in transformation efficiency between non-targeting and *phaC*-targeting versions of the pSIBR005 plasmid. **(B)** Location of the targeting crRNA spacers (T1, T2, and T3) along the *phaC* locus. The sequence of the Protospacer Adjacent Motif (PAM) is shown in bold purple font (consensus is 5'-(T)TTV-3', where V = G, C, A^{211}). Each 20 bp spacer sequence is shown in blue. **(C)** Spot microdilutions of pSIBR005 transformation cultures. Each plasmid was transformed into both WT and Δ*phaC C. necator* strains, with the latter being used as a negative control. Each spot represents a 10-fold dilution over the previous column (left to right). Spacers T1 and T3 resulted in the highest efficiency of targeting, only in the WT strain ($\geq 10^3$ reduction in transformation efficiency relative to the non-targeting plasmid). **(D)** Transformation efficiency, quantified as CFU/µg DNA from the spot microdilution assays shown in (C), for the non-targeting and *phaC*-targeting versions of pSIBR005. Bars indicate the mean of *n* = 3, ± one standard deviation.

3.5.2. SIBR-mediated regulation of *Cas12a* expression is defective in *C. necator*

SIBR-Cas12a and SIBR-Cas9 constructs cannot be efficiently transformed into WT *C. necator*

Having demonstrated constitutive targeting, we investigated whether the expression of CRISPR-Cas12a and CRISPR-Cas9 complexes could be controlled by the SIBR system in *C. necator*, using theophylline as an inducer. To the best of our knowledge, theophylline has not been previously used as an inducer in *C. necator*. We conducted a theophylline toxicity assay to simultaneously determine (a) whether the molecule is taken up by the bacterium and (b) which concentrations are tolerated and could therefore be used for induction. Theophylline was found to be toxic only when added to the growth medium at concentrations of ≥ 5 mM (Figure S3.1.4), confirming its suitability to mediate inducible gene expression in *C. necator* at lower concentrations (typically 2-7 mM[209]).

Non-targeting and *phaC*-targeting versions of plasmids pSIBR001-pSIBR004 were built. These plasmids each contain one of four variants of the SIBR intron, with different variants having been shown to be optimal in different genomic contexts (see ref. [209] for details). We conducted an inducible targeting assay to determine which intron variant performs best in *C. necator*, but found that very few colonies (or none) could be recovered when the SIBR-Cas12a constructs were transformed into *C. necator*, regardless of the nature of the crRNA spacer sequence (Figure 3.3).

We hypothesised that this could be due to an active bacterial defence system recognising the SIBR intron sequence and targeting the SIBR-Cas12a plasmids for degradation. To circumvent this, we tested whether SIBR-Cas12a constructs could be transformed into *C. necator*Δ0006. This strain carries a deletion that inactivates one of the bacterium's restriction-modification systems and has been previously reported to increase electroporation efficiency[207] (the Δ0006 mutant was also used in Chapter 2, see Section 2.4.1). We found that non-targeting SIBR-Cas12a plasmids could be readily electroporated into the Δ0006 strain (Figure 3.3 B). Interestingly, plasmids re-isolated from *C. necator* instead of *E. coli* DH5α were also efficiently transformed into both the WT and Δ0006 strains. This suggests that *E. coli*-derived methylation patterns within the SIBR intron sequence may be recognised as foreign and targeted by the host restriction-modification machinery. We also confirmed that the Δ0006 strain does not yield higher electroporation efficiency for the control plasmids pSEVA231 and non-targeting pSIBR005 (Figure S3.1.5D), consistent with the results reported in Chapter 2. Low electroporation efficiency in the WT strain was also encountered with SIBR-Cas9 plasmids, and could analogously be resolved by using the Δ0006 mutant as the recipient strain (Enrico Orsi, personal communication, data not shown).

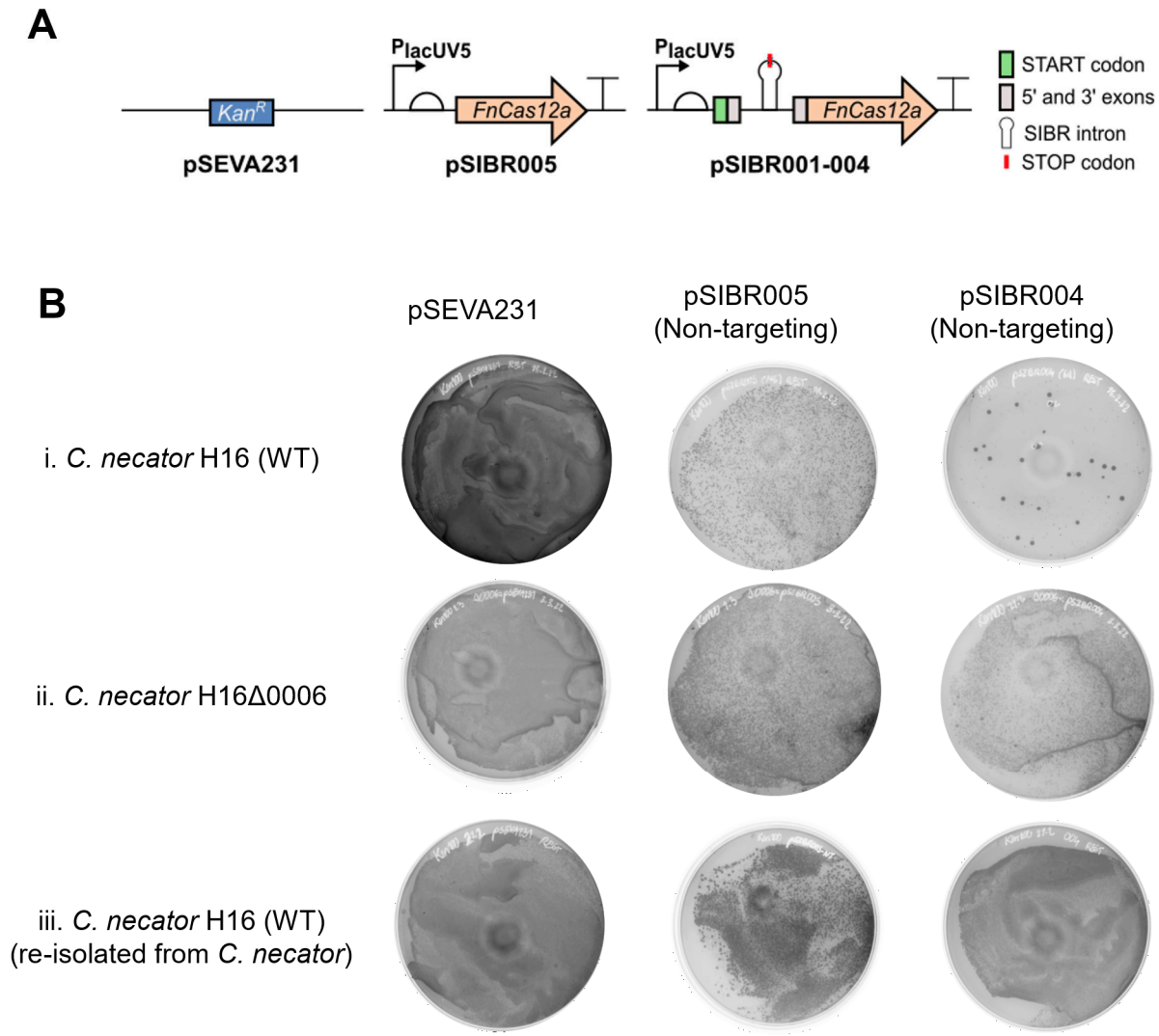


Figure 3. 3. SIBR-Cas plasmids cannot be readily transformed in wildtype *C. necator*.

(A) Plasmids used for electroporation testing. A positive control was provided by the empty backbone plasmid pSEVA231, which is identical to the backbone of SIBR plasmids. Non-targeting versions of plasmids pSIBR005 and pSIBR001-004 were used for testing, which mediate constitutive and theophylline-inducible *Cas12a* expression, respectively. **(B)** Representative transformation plates for all pairwise combination of plasmids and host strains. Note that plasmid pSIBR004 is shown as a representative example for plasmids pSIBR001-pSIBR004. Each plate contains 900 µL of the appropriate transformation culture, as indicated. (i) When plasmids isolated from *E. coli* (as done routinely) were electroporated into a WT background, low transformation efficiency was observed only for SIBR-Cas12a plasmids (pSIBR004 plates are shown as a representative example). Low electroporation efficiency could be improved both by (ii) transforming the plasmids into the *C.necator*Δ0006 host strain, or (iii) by re-isolating plasmids from *C. necator* transformants and re-electroporating them into the WT.

SIBR-Cas12a constructs fail to mediate inducible *Cas12a* expression

To conduct inducible targeting assays, alternative target loci had to be selected, as our laboratory's stock of the *C. necator*Δ0006 strain also has a Δ*phaC* background. We designed spacer sequences to target the *acoC* locus (Table S3.2.2), which encodes the E2 subunit of a branched-chain alpha-keto acid dehydrogenase (KEGG ID H16_B0146). *acoC* and its enclosing *acoXABC* operon are involved in the catabolism of acetoin in *C. necator*[212,213]. Genes within this locus are not essential and have been previously deleted as part of metabolic engineering efforts[214,215]. Prior to building *acoC*-targeting versions of SIBR-Cas12a plasmids, we confirmed targeting efficiency for two different spacer sequences (T2 and T8) in both the WT and Δ0006 strains (Figure S3.1.5). Spacer sequences for the *acoC* locus are listed in Table S3.2.2.

Inducible targeting assays were conducted by transforming all versions of the SIBR-Cas12a constructs (pSIBR001-pSIBR004 with non-targeting, *phaC*-targeting, or *acoC*-targeting crRNA spacers) into C*. necator*Δ0006*.* Constructs containing non-targeting and *phaC*-targeting crRNA spacers provided negative controls, as the Δ0006 mutant also has a Δ*phaC* background. Following transformation, cells were selected on solid medium with or without theophylline. The difference in transformation efficiency between induced an uninduced conditions was quantified to provide a measure of inducible targeting efficiency. For all variants of the SIBR-Cas12a constructs, we observed a significant difference in transformation efficiency between non-targeting and *acoC*-targeting versions of the plasmids regardless of the presence or absence of theophylline (Figure 3.4). This is consistent with active CRISPR-Cas12a

genome targeting even in uninduced conditions, indicating that SIBR regulation of *Cas12a* expression was defective. Interestingly, this regulatory defect was not observed for SIBR-Cas9 constructs, where theophylline-inducible *Cas9* expression could be mediated by the SIBR intron (Figure S3.1.6).

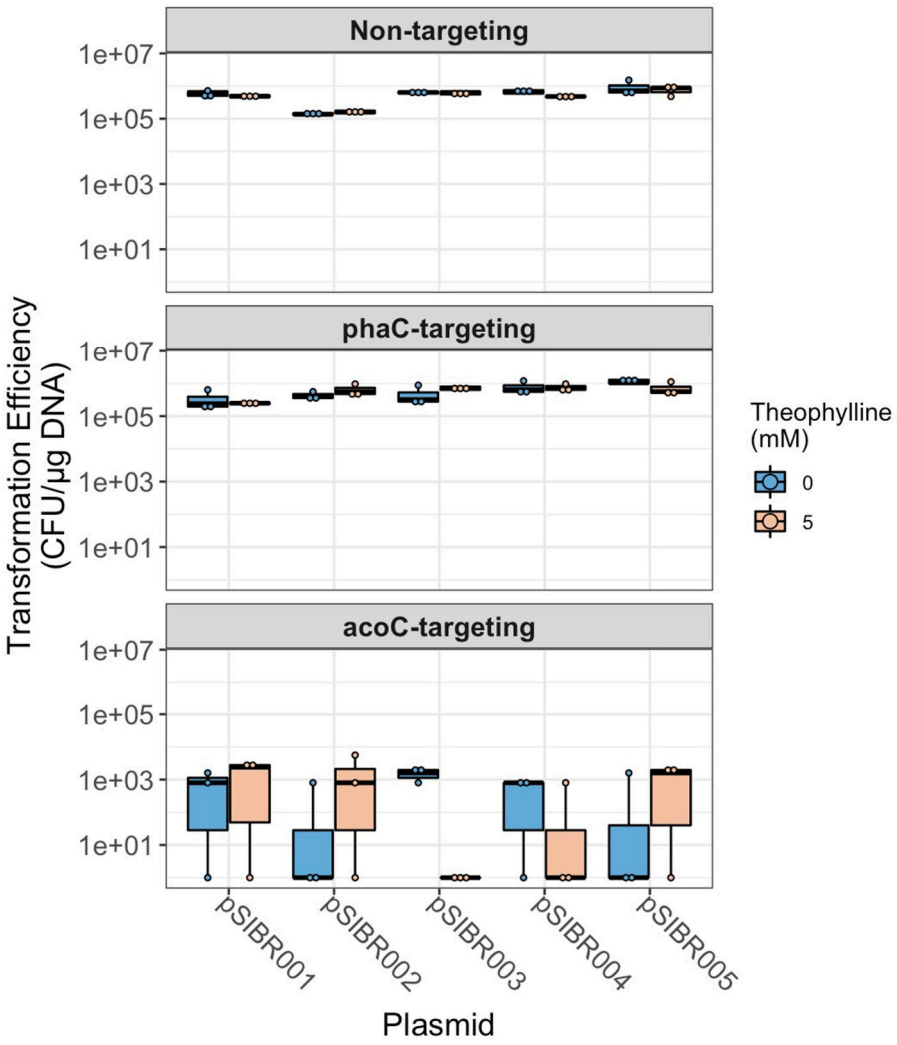


Figure 3. 4. SIBR-mediated regulation of Cas12a expression is defective in *C. necator*.

Plots show the transformation efficiency for pSIBR001-005 plasmids, carrying either non-targeting, *acoC*-targeting or *phaC*-targeting crRNA spacers, into the *C. necator*Δ0006Δ*phaC* strain. Plasmid pSIBR005 directs constitutive *Cas12a* expression, whilst plasmids pSIBR001-004 direct inducible *Cas12a* expression, mediated by four distinct variants of the SIBR intron (each plasmid encodes a unique variant). Following transformation, cells harbouring the test plasmids were plated on solid medium (LB plates) supplemented with kanamycin and either 0 or 5mM of the inducer theophylline. All variants of the non-targeting and *phaC*-targeting plasmids, which were used as negative controls, were

transformed with high efficiency across all possible induction conditions, as expected (i.e., no constitutive or inducible genome targeting was detected for these plasmids). All *acoC*-targeting constructs were transformed with a ≥10^3-fold lower efficiency than their non-targeting counterparts, regardless of the theophylline concentration. Inducible constructs (pSIBR001-004) displayed a targeting pattern which was indistinguishable from that of the constitutive pSIBR005 plasmid, indicating the defective induction of *Cas12a* expression in these plasmids.

3.5.3. Defective inducible expression in SIBR-Cas12a constructs is likely due to translation from an alternative RBS within the intron sequence.

Although previous data from *E. coli*, *P. putida* and *Flavobacterium* IR1 demonstrate strict SIBR-mediated inducible control of *Cas12a* expression, it is evident that the same regulatory kinetics may not be observed when SIBR is used in a different host organism. Additionally, favourable data from inducible targeting assays with SIBR-Cas9 plasmids in *C. necator* highlight that the performance of the gene regulatory system may be dependent on both the host strain (genomic context) and the immediate genetic context of the SIBR intron, suggesting an interplay between these two factors. Taking this into account, we hypothesised that there are two potential causes for the defective regulation of *Cas12a* expression in SIBR-Cas12a constructs. It is possible that the SIBR intron is self-splicing from *Cas12a* pre-mRNA transcripts in the absence of theophylline (leaky self-splicing). Alternatively, functional Cas12a could be translated from pre-mRNA transcripts from a secondary RBS either within the intron sequence or elsewhere near the 5’ end of the CDS. These hypotheses are not mutually exclusive.

To eliminate the possibility of leakiness due to the splicing of SIBR in the absence of theophylline, we introduced a STOP codon within the 5' exon sequence of SIBR-Cas12a, yielding plasmid pSIBR00X (Figure 3.5 A). This design ensures that even if SIBR splices out in the absence of theophylline, a premature STOP codon will preclude the translation of functional Cas12a, preventing in turn counterselection even in the presence of a targeting spacer. As done previously, non-targeting and *acoC*-targeting versions of pSIBR00X were transformed into *C. necator*Δ0006, and cells were plated on solid medium with and without theophylline. As shown in Figure 3.5 B, the presence of a premature STOP codon at the 5' exon did not eliminate the translation of Cas12a, as genome targeting was observed both in the absence and presence of theophylline. This result indicates that factors other than leaky self-splicing result in the expression of the Cas12a protein. Though it may seem counterintuitive that targeting is observed even in the presence of the inducer from the pSIBR00X constructs, it is important to note that splicing of the SIBR intron occurs with low efficiency when theophylline is present, a key feature contributing to tight control of *Cas12a* expression in the SIBR-Cas framework. Thus, many un-spliced pre-mRNA transcripts remain even in induced conditions, from which Cas12a could be translated if an alternative RBS were present.

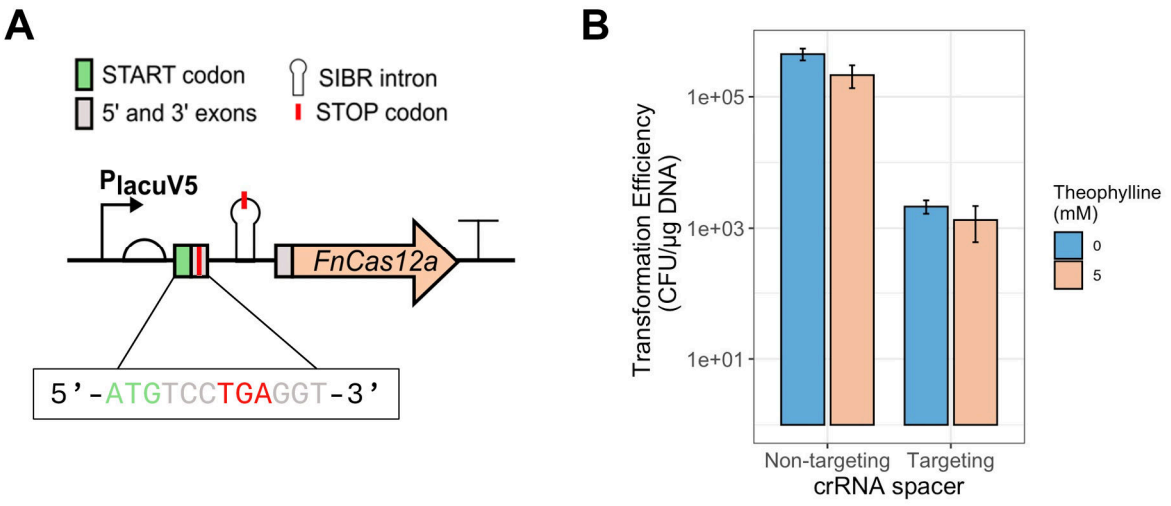


Figure 3. 5. The SIBR-Cas12a splicing defect is likely due to alternative translation of Cas12a pre-mRNA transcripts.

(A) Schematic diagram of the *Cas12a* expression cassette on plasmid pSIBR00X. This is identical to that of the pSIBR001-004 plasmids, except for an additional STOP codon within the 5' exon flanking the SIBR intron. Provided that no internal RBSs are encoded within the SIBR intron sequence, the presence of an additional premature stop codon at this location should preclude expression of *Cas12a* regardless of the presence or absence of theophylline, as self-splicing of the SIBR intron would not remove the additional stop codon. **(B)** Inducible targeting assay for pSIBR00X. Transformation efficiency, quantified as CFU/µg DNA, for the non-targeting and *acoC*-targeting (T3) versions of plasmid pSIBR00X on LB medium supplemented with kanamycin and 0 or 5mM theophylline. A $>10^2$ reduction was observed for the *acoC*-targeting version of pSIBR00X, regardless of the theophylline concentration. Bars indicate the mean of $n = 3$, ± one standard deviation.

We conducted a bioinformatic analysis of the SIBR-Cas pre-mRNA sequences to identify any potential RBSs from which near-full length, functional Cas12a could be alternatively translated. Using the RBS Calculator biophysical model[53,216,217], we compared the predicted translation initiation rates (TIR) in *C. necator* over the SIBR intron sequence for both SIBR-Cas9 and SIBR-Cas12a constructs. Interestingly, we identified a site near the 3' end of the intron sequence where the TIR was predicted to spike for SIBR-Cas12a constructs, but not for SIBR-Cas9 constructs (Figure S3.1.7 A). The site corresponds to a methionine codon which is downstream from the final STOP codon in the SIBR intron sequence.

3.5.4. SIBR2.0: transferring the SIBR intron along the *Cas12a* CDS leads to tight control of gene expression

One of the main proposed advantages of the SIBR system is its portability. It functions as a modular unit for controlling gene expression, and it can theoretically be applied to any gene of interest (GOI). In this work, we have demonstrated this by applying the intron to control *Cas9* expression in *C. necator*. However, we have also shown that the functionality of the SIBR intron is dependent on its genomic and genetic context, meaning that its portability may be limited and ought to be assessed on an *ad hoc* basis.

Recognising this limitation, we developed a novel iteration of the SIBR system (SIBR2.0) which allows the intron to be transferred from its canonical location immediately following the start codon to the inner CDS of the GOI. We hypothesised that this strategy may alleviate the regulatory defect observed for SIBR-Cas12a constructs in *C. necator*. The reasoning underlying this hypothesis is two-fold. First, altering the immediate genetic context of the intron may reduce alternative translation from within the SIBR sequence. Second, the SIBR intron can be placed at strategic locations along the *Cas12a* CDS such that essential functional domains are excluded from any protein translated from within the intron sequence.

We designed alternative versions of the SIBR-Cas12a constructs, where the SIBR intron was moved from its original location (at amino acid position two) to suitable alternative CDS locations (Figure 3.6 A and B). These locations were identified with

the help of a computational model, integrating knowledge on splicing and binding efficiency of the theophylline-dependent T4 *td* intron obtained during the original characterisation of the SIBR system[209]. A custom Python script was used to execute the model and predict the best alternative locations for the intron, which could be applied beyond *Cas12a* to place the SIBR intron within any GOI (Constantinos Patinios, personal communication, data not shown). Following bioinformatic analysis, plasmids pSIBR414 and pSIBR818 were constructed, containing the SIBR2.0 intron sequence at residue positions 414 and 818, respectively.

Non-targeting and *acoC*-targeting versions of pSIBR414 and pSIBR818 were built and used to conduct an inducible targeting assay (Figure 3.6 C). For both sets of constructs, a reduction in transformation efficiency was only observed for the targeting plasmids in the presence of theophylline. These results confirm that the SIBR2.0 intron can mediate tightly inducible control of *Cas12a* expression when placed at alternative locations within the CDS.

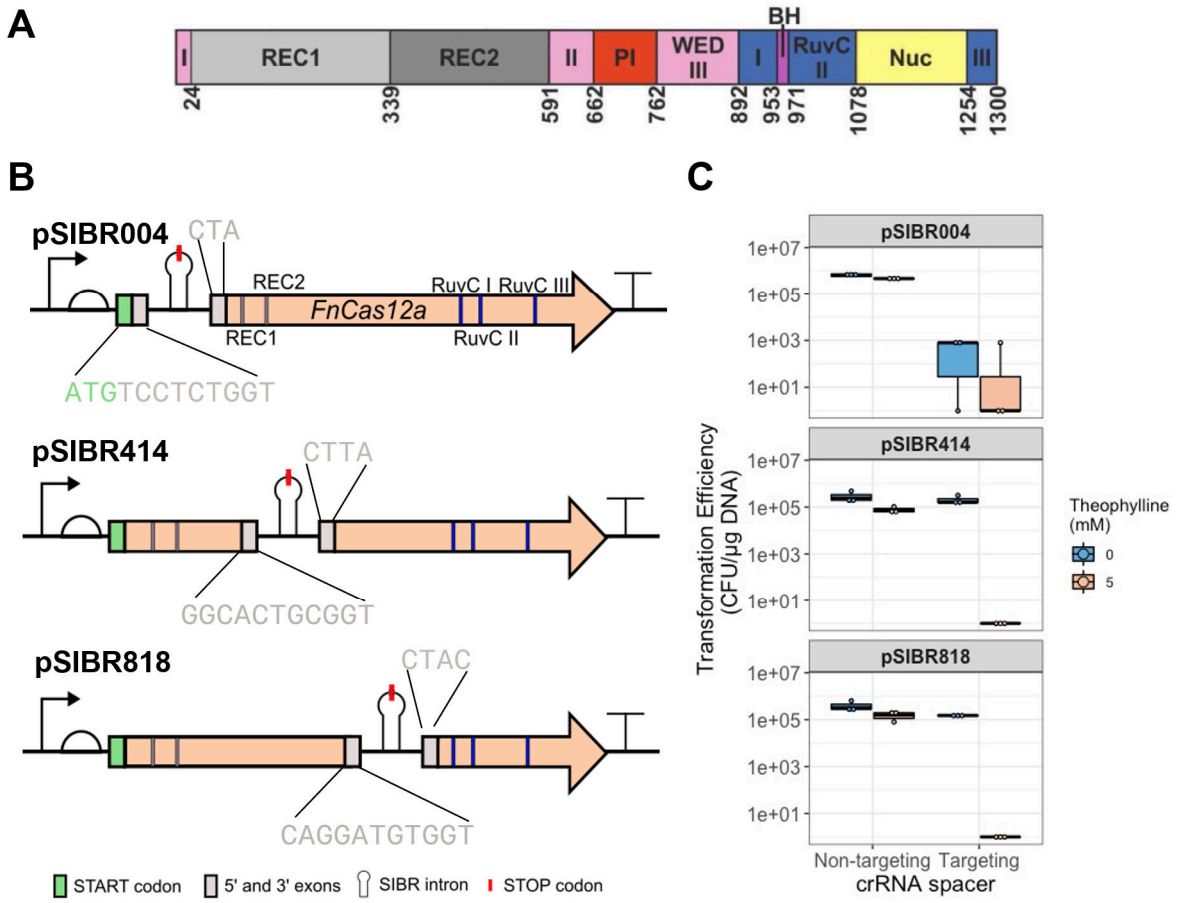


Figure 3. 6. Alternative SIBR placement along the Cas12a CDS enables inducible regulation of gene expression.

(A) Functional domains in the Cas12a protein (taken from ref. [218]). REC = recognition, PI = PAM interacting, WED = wedge, BH = bridge helix, Nuc = nuclease. **(B)** In SIBR2.0, the intron sequence and its flanking 5' and 3' exons are relocated from their canonical location at the 5' end (as in pSIBR004, top) to the inner CDS of *Cas12a*, with the goal of eliminating translation of functional Cas12a from within the SIBR sequence. Plasmids pSIBR414 and pSIBR818 were constructed, placing the intron at amino acid positions 414 and 818, respectively. **(C)** Inducible targeting efficiency. Transformation efficiencies for plasmids pSIBR004, pSIBR414, and pSIBR818 are compared.

Further, we analysed the SIBR2.0 sequence within the pre-mRNA context to determine whether relocating the intron lowers the predicted TIR from any internal RBSs. Interestingly, we found that there is a site within the SIBR2.0 sequence which is predicted to drive translation at a high rate in both pSIBR414 and pSIBR818, comparable to the TIR of the original SIBR intron in plasmid pSIBR004 (Figure S3.1.7

B). However, the predicted translation initiation site occurs upstream of the final stop codon within the SIBR2.0 sequence, meaning that translation of functional protein from this site is not possible, consistent with the results of inducible targeting assays.

3.5.5. SIBR2.0-Cas12a enables efficient genome editing in *C. necator*

Having shown that the SIBR2.0 system can be used to tightly control the expression of CRISPR-Cas12a complexes in *C. necator*, we sought to apply the tool to genome engineering. Namely, we tested whether SIBR2.0-Cas12a could be used to perform a knockout of the *acoC* locus via homologous recombination (HR) with CRISPR-mediated counterselection (Figure 3.7 A). First, homology arms were added to pSIBR818 (non-targeting and *acoC*-targeting) to facilitate HR at the target locus. These were designed to be ~1 kb each, as described previously[197,198]. Editing versions of plasmid pSIBR005 were also constructed and used as controls.

Editing plasmids were electroporated into *C. necator*Δ0006. Transformants were grown for six to eight hours in non-selective rich medium before cells were plated on selective solid medium (supplemented with kanamycin), with or without the inducer theophylline. This recovery time was defined as an optimum through preliminary optimisation experiments, where we observed a trade-off between transformation efficiency and editing efficiency as a function of outgrowth duration (Figure S3.1.8)

Editing efficiency was quantified by colony PCR (cPCR) of transformants. Specifically, we define editing efficiency as the fraction (percentage) of recovered colony forming

units where the target genomic locus had been deleted. Primers for cPCR were validated before use, and shown to yield three possible band patterns: a single high molecular weight DNA band corresponding to the WT locus, a shorter band corresponding to the edited locus, and a mixed pattern of bands corresponding to merodiploids, i.e. strains where the second HR step had not occurred (data not shown). For reference, binding sites for cPCR primers in WT, merodiploid, and edited genomes are indicated in Figure 3.7 A. Primer sequences are detailed in Section 3.4. Only transformant colonies yielding a single short DNA band were counted as fully edited. WT and merodiploid colonies were counted as unedited.

For all editing constructs, final editing efficiencies are provided in Figure 3.7C. Agarose gel images for representative cPCRs conducted as part of the editing assays are provided in Figure S3.1.9. For the control plasmids (pSIBR005), low editing efficiency (≤50%) was recorded in all cases, and no significant differences were observed between induced and uninduced conditions. For pSIBR818 editing plasmids, a final editing efficiency of (70±29)% was recorded only for targeting constructs under induced conditions. These data demonstrate that counterselection of the WT by SIBR2.0-Cas12a is necessary and sufficient to mediate highly efficient genome editing in *C. necator*. We note that targeting versions of pSIBR818 editing plasmids are transformed with low efficiency relative to their non-targeting counterparts in both induced and uninduced conditions (Figure 3.7 B). Thus, even though editing efficiency is high, only a small number of colonies were obtained. As expected, targeting pSIBR818 plasmids were the only constructs for which we observed a significant difference in transformation efficiency between induced and uninduced conditions.

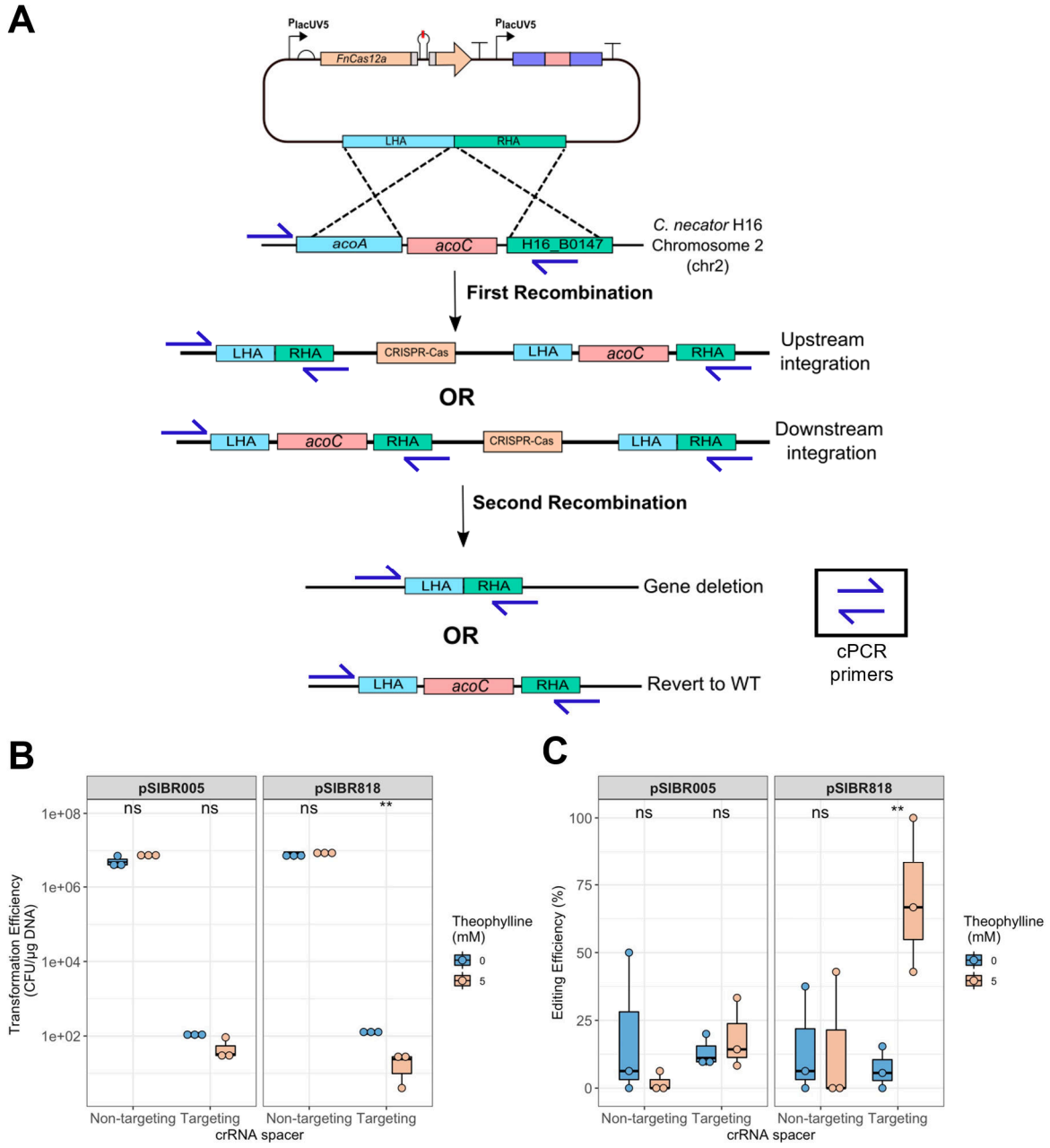


Figure 3. 7. SIBR2.0-Cas12a mediates efficient genome editing.

(A) Overview of the genome editing strategy via HR and CRISPR-Cas12a counterselection. The editing version of the pSIBR818 plasmid (top) contains the SIBR2.0-Cas12a module, as well as homology arms for upstream (left hand arm, LHA) and downstream (right hand arm, RHA) chromosomal integration at the *acoC* locus. Two sequential HR events are required for full editing. First, the editing plasmid is integrated into the chromosome either upstream or downstream of the target locus via a single crossover event, resulting in merodiploid strains. Following a second HR event (double crossover), strains can either revert to the WT locus or delete the target sequence, with a theoretical editing efficiency of 50%. The SIBR-Cas12a module on the editing plasmid can be induced at this stage to counterselect the WT strains from the population and increase the editing efficiency. Red arrows indicate all available binding sites for forward and reverse cPCR primers. **(B)** Transformation efficiency

for each editing construct (pSIBR005 or pSIBR818, non-targeting or *acoC*-targeting) in *C. necator*Δ0006. **(C)** Editing efficiency (percentage of cells were the target locus had been deleted, as quantified by cPCR) for each construct. In (B) and (C), individual observations are shown as dots. ns = not significant, ** = $p<0.05$, two-sample t-test.

3.5.6. SIBR2.0-Cas12a editing plasmids can be readily cured

Following successful genomic manipulation, SIBR plasmids must be cured (removed) from the host strain to enable iterative editing or transformation of other plasmids of interest. To determine an optimal protocol for plasmid curing from edited cells, we tested three different culturing conditions for which we could find precedent in the *C. necator* literature[207,182,219]. These involved growing cells in non-selective rich medium (LB) at 30 ºC with or without rifampicin, or at 37 ºC without any additional antibiotics. As a control, cells were forced to retain the editing plasmid by culturing in selective LB medium (LB with 100 µg/ml kanamycin). Cells were grown to saturation in each condition and serially passaged. At each passage, the kanamycin-resistant fraction of the cell population was quantified (see Section 3.3 Materials and Methods for details and Figure 3.8 B for an example). We found that culturing deletion mutants in LB at 37 ºC was the optimal condition for plasmid curing, with ≥98% of the cell population becoming sensitive to kanamycin after a single overnight (16-hour) incubation (Figure 3.8 A).

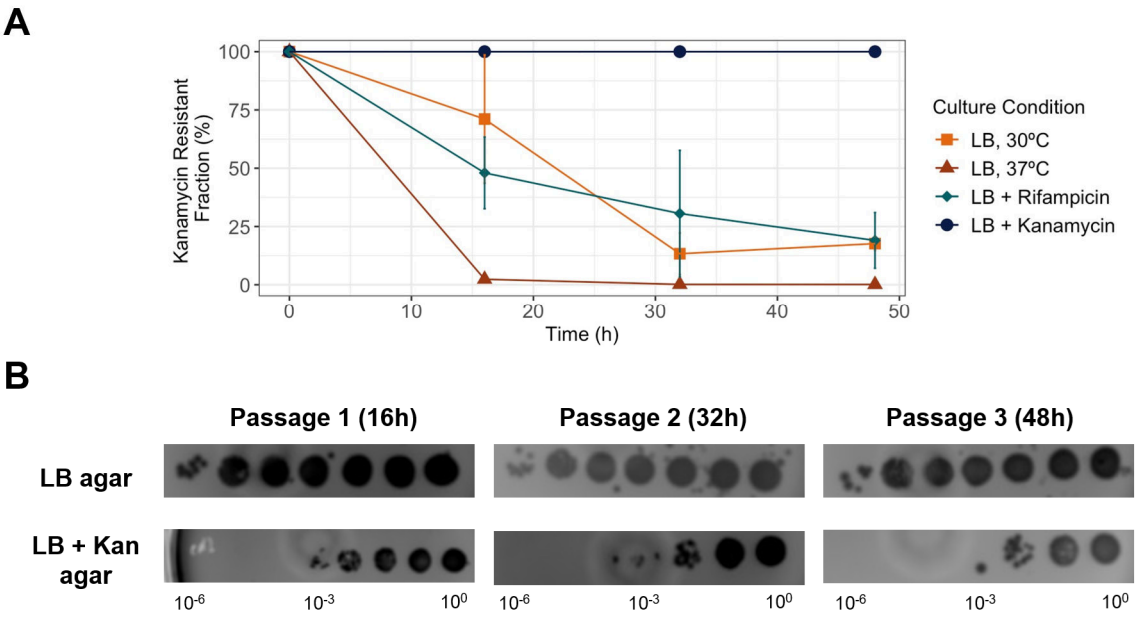


Figure 3. 8. Curing the SIBR-Cas12a editing plasmid.

(A) Evolution of plasmid loss over culture time in test and control conditions. For each specified culture condition, each data point corresponds to a passage (culture dilution) step. At each passage, the fraction of the population (%) which had retained the plasmid-encoded genetic marker conferring kanamycin resistance was quantified by serial dilution and colony counting. Individual points indicate the average of *n* = 3 replicates, ± one standard deviation. **(B)** Representative serial dilutions of cultures grown in LB at 37 ºC at the time of each passage. Dilutions on permissive and selective solid agar medium (LB and LB with 100 µg/ml kanamycin, respectively) were used to quantify the total number of cells and the kanamycin-resistant number of cells in the population, respectively. The first spot (rightmost) corresponds to undiluted culture samples (dilution factor 10^0), with every subsequent spot (right to left) being a ten-fold dilution of the preceding sample, as indicated.

3.6. Discussion and Conclusions

We successfully adapted the standardised SIBR-Cas system for use in *C. necator*, demonstrating that it can be applied to mediate high-efficiency genome editing through homologous recombination and CRISPR-Cas counterselection. In doing so, we provide the first report of an active CRISPR-Cas12a system in this bacterium, and a thorough characterisation of its (inducible) targeting efficiency. Though CRISPR-Cas

mediated counterselection has been previously demonstrated for *C. necator*[207], our report is the first to investigate its targeting kinetics and efficiency, and to provide evidence for tightly inducible control of both *Cas9* and *Cas12a* expression.

Our study delivers two significant innovations of the SIBR framework. SIBR-Cas was originally shown to tightly control the expression of *Cas12a*[209]. Here, we successfully transferred the SIBR intron to control the expression of a different Cas nuclease gene (*Cas9*). This demonstrates the system's portability, which stems from its organism- and sequence-independent design. However, we have also shown that this portability is limited. The canonical SIBR-Cas12a plasmids were shown to be defective in *C. necator*, highlighting the influence of both genomic and genetic contexts on inducible control of gene expression by the SIBR intron. Though the SIBR system is self-contained and can be easily transferred to non-model bacteria, our results suggest that its implementation in a new species ought to be tested and optimised on a case-by-case basis. We demonstrate a complete workflow that should be followed for this purpose, i.e. when implementing the SIBR-Cas system in a new genomic context. We show that, where the original SIBR intron fails due to alternative translation from within its sequence, SIBR2.0 can instead be applied to mediate tight induction. This new iteration of the SIBR technology, which explores alternative placement of the SIBR intron, could theoretically also be used to tightly regulate the expression of any gene(s) of interest, as the intron can be placed within any CDS (provided that suitable in-frame intron flanking sequences can be identified, and that adequate inducible splicing behaviour can be attained in the target genetic and genomic context). Moreover, the SIBR2.0 intron could be used to implement more complex regulatory architectures, as explored in Chapter 4.

We successfully applied SIBR2.0-Cas12a to perform a deletion of the *acoC* locus with high efficiency (~70%, Figure 3.7). Every step in the editing protocol is both streamlined and standardised. The design of targeting crRNA spacers can be conducted automatically using a custom Python software (Section 3.8). Once synthesised, spacers can be easily assembled into the SIBR plasmids via a one-pot Golden Gate reaction. Assembly of the homology arms requires only two steps: PCR amplification of the appropriate genomic DNA sequences, and Gibson assembly into the SIBR backbone, which can be linearised via restriction-digestion and does not require PCR amplification. The resulting editing plasmids can be efficiently transformed into *C. necator* via the optimised electroporation protocol outlined in Chapter 2. Efficient editing can be attained by extending the recovery time to eight hours, meaning the protocol can be executed in a single working day (approximately nine hours when factoring in ample time for electroporation and plating). Edited colonies are recovered on transformation plates after 48 to 72 hours, giving a total editing timeline of three to four days, most of which is determined by the organism's growth rate on solid (LB agar) medium. This represents a substantial improvement over previous methods, which require at least two weeks[206,207].

Nevertheless, the protocol has some important limitations. Firstly, editing was only demonstrated in the *C. necator*Δ0006 strain, as plasmids containing the SIBR intron sequence could not be efficiently transformed into the WT. Though the mutant is commonly used in the literature, this limitation means that SIBR may not be suitable to edit existing host strains that do not have a Δ0006 background. Where editing of the WT strain is required, users may circumvent this limitation by isolating SIBR

plasmids from *C. necator* and re-transforming them to attain higher electroporation efficiency. In this case, plasmids should be sequenced prior to electroporation to confirm that no mutations have been accumulated during propagation in *C. necator*. Another option would be to clone the SIBR editing plasmids in an *E. coli* strain which expresses *C. necator* restriction-associated methyltransferases. Any plasmid DNA isolated from these recombinant *E. coli* strains would carry a methylation pattern analogous to that of *C. necator*, and would therefore not be recognised as foreign. This is a common approach to circumvent restriction-modification systems in non-model microbes[47]. Interestingly we note that using the Δ0006 strain does not increase absolute transformation efficiency, for example transforming pSIBR005 or pSEVA231 plasmids into this strain or the WT leads to similar efficiency (Figure S3.1.5.D). Rather, the strain is more tolerant to recombinant plasmids that include sequences with specific features (such as methylation patterns), which would otherwise be recognised as foreign.

Another limitation of the protocol is the low transformation efficiency obtained with the targeting versions of pSIBR818 editing plasmids. Though editing efficiency follows the expected pattern (a high percentage of edited colonies is only recovered in the presence of the inducer), transformation efficiency for these plasmids was low in both induced and uninduced conditions (Figures 3.7 and S3.1.8). This result was both significant and reproducible, and to the best of our knowledge has no precedent in the literature. The only other CRISPR-Cas genome editing report available for *C. necator* does not provide details on editing or transformation efficiency in both induced and uninduced conditions[207]. Instead, the authors described culturing the strains in the presence of the inducer for 96 to 168 hours to attain high editing efficiency, a time-

consuming protocol which does not provide insights into the mechanisms underlying efficient editing. Here, we identified a trade-off between editing and transformation efficiency, and propose that a high percentage of edited colonies can be obtained simply by extending recovery time in non-selective rich medium. Despite the low transformation efficiency, we recovered at least one edited colony in all cases.

Nevertheless, this important limitation may impede successful editing at other loci and should be investigated further. Leaky expression of the *Cas12a* nuclease would generally be considered the most likely cause of low transformation efficiency. However, we did not observe leaky expression when conducting inducible targeting assays with SIBR2.0-Cas12a constructs (Figure 3.6C) . Another possible explanation for the observed decrease in transformation efficiency would be disruption of the plasmid by the host defence machinery, facilitated by the extended recovery time in non-selective medium. Bacteria may use transposable elements such as insertion sequences (ISs) to disrupt burdensome recombinant plasmids[220]. Moreover, the recombinant plasmid may be targeted by the host's restriction-modification systems. It is unclear, however, why this would be observed only for pSIBR818 editing plasmids and not for other constructs, including constructs for constitutive *Cas12a* expression, which pose a larger burden on the bacterium. This could be investigated further by sequencing the SIBR editing plasmids over time, both those maintained as episomes and chromosomally-integrated copies. This would elucidate whether plasmid disruption is a key cause of decreased transformation efficiency and provide further insights on the dynamics of homology-based editing in *C. necator*. These insights could be harnessed to further optimise protocol parameters (chiefly recovery time, induction time, and selective medium composition), to obtain a more robust method.

It might also be necessary to domesticate the host strain further, for example by creating an IS-free mutant. Restrictive enzymes may also be removed from the host genome to increase its genetic tractability. Such domesticated strains would additionally be useful for metabolic engineering, as they are likely to yield higher evolutionary stability for any engineered functions such as plasmid-encoded biosynthetic programs[85].

Despite the limitations discussed above, we have demonstrated that SIBR-Cas is functional in *C. necator*, and have provided a standardised toolkit which is primed for widespread use and optimisation in the broader research community (upon publication). Due to its modular and standardised architecture, SIBR-Cas could be easily modified to implement other CRISPR-based applications, such as genomic insertions (knock-ins), CRISPR interference (CRISPRi) or activation (CRISPRa). Moreover, the SIBR and SIBR2.0 DNA parts could be applied beyond CRISPR-Cas tools as useful regulatory modules for precise regulation of gene expression.

3.7. Chapter 3 Appendix 1 – Supplementary Information

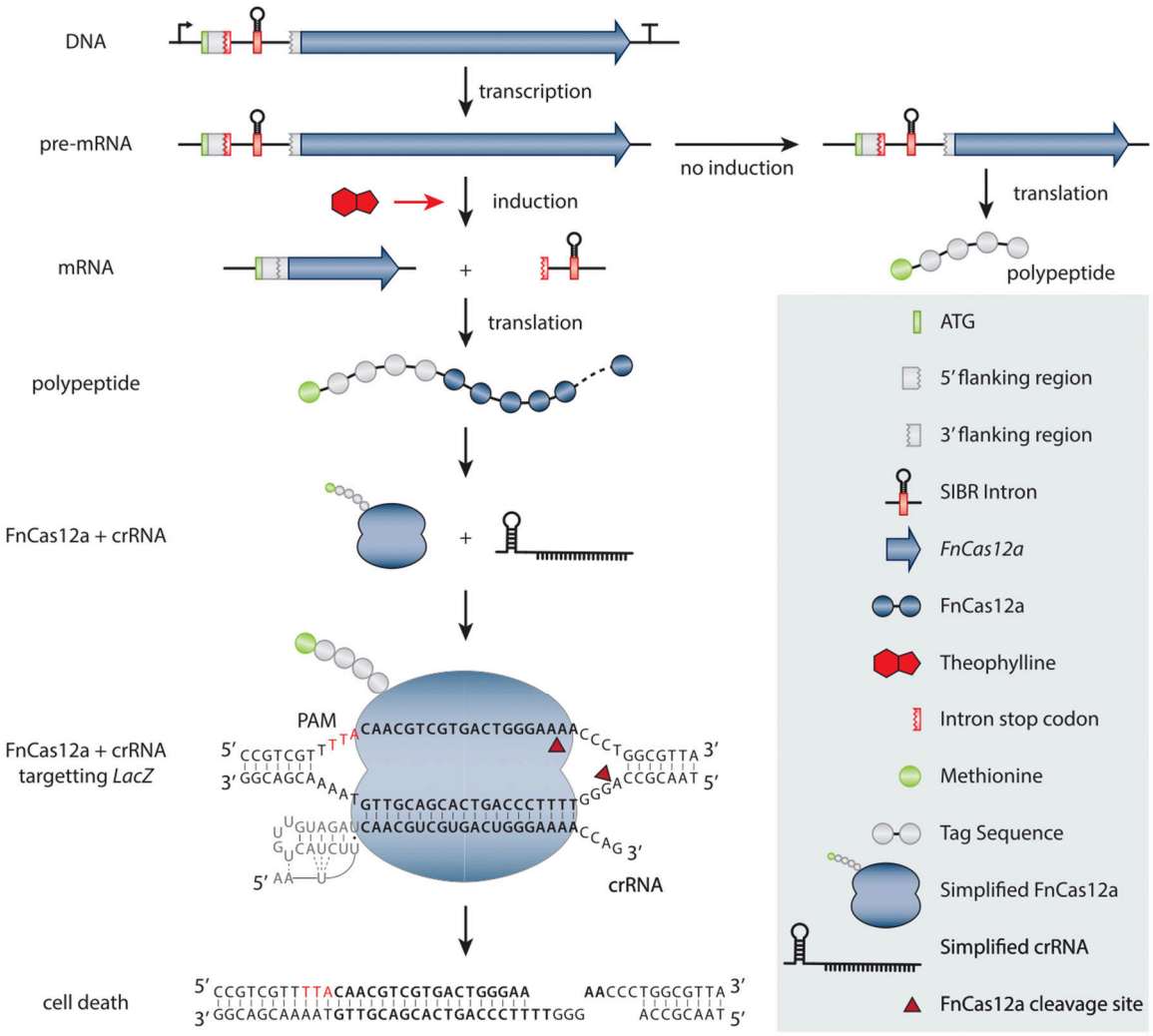


Figure S3.1. 1. SIBR-Cas enables tightly inducible control of Cas12a expression.

Schematic illustration of the SIBR-Cas technology, reproduced from Patinios *et al*. (ref. [209]). The SIBR intron, which encodes a stop codon, is placed at the 5' end of the *Cas12a* CDS, with its 5' flanking region (exon) immediately following the ATG start codon. In the absence of theophylline (inducer), the pre-mRNA is not processed, and translation is terminated when the ribosome encounters the stop codon encoded within the SIBR intron sequence. When the inducer is added, intron self-splicing is induced at the pre-mRNA level, resulting in processed transcripts with no premature stop codons, which can be translated into full-length Cas12a by the ribosome. The full-length nuclease then pairs with its cognate crRNA, which is constitutively expressed, to mediate cleavage of genomic DNA at the specified locus.

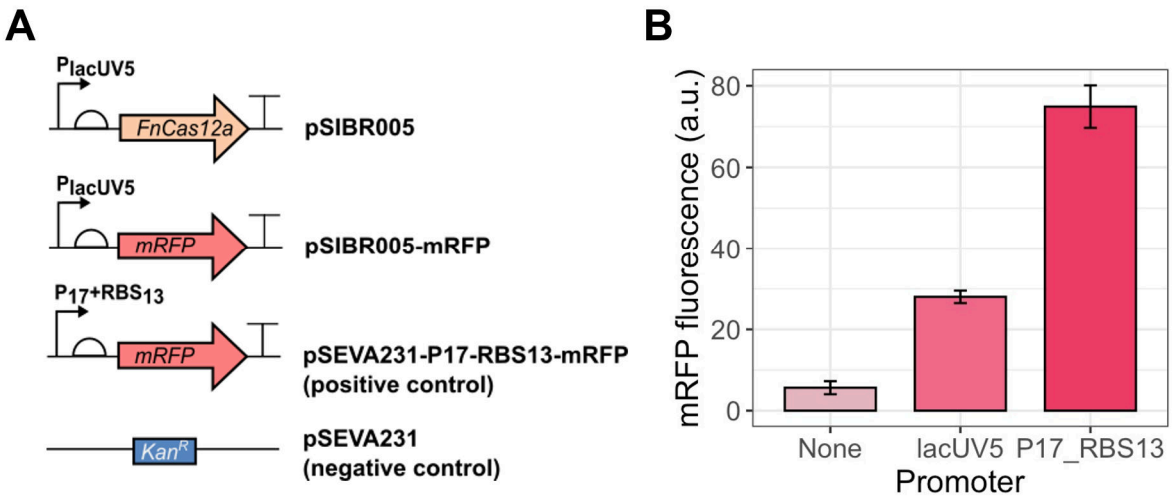


Figure S3.1. 2. The P_{lacUV5} promoter directs constitutive expression of mRFP in *C. necator*.

(A) In the SIBR framework, P_{lacUV5} is used to drive constitutive expression of the *Cas12a* nuclease as shown for plasmid pSIBR005. Plasmid pSIBR005-mRFP was built to check whether this promoter could drive constitutive gene expression in *C. necator*. Plasmid pSEVA231-P_{17}-RBS_{13}-mRFP was used as a positive control. P_{17} and RBS_{13} are synthetic parts which have been previously designed for use in *C. necator*[221]. They were implemented in the control plasmid to drive constitutive expression of mRFP. An empty Kan^R backbone plasmid (pSEVA231) was used as a negative control. **(B)** mRFP fluorescence of recombinant *C. necator* cells carrying each constructs shown in (A), as quantified by flow cytometry. Bars show the mean of *n* = 3 measurements ± one standard deviation.

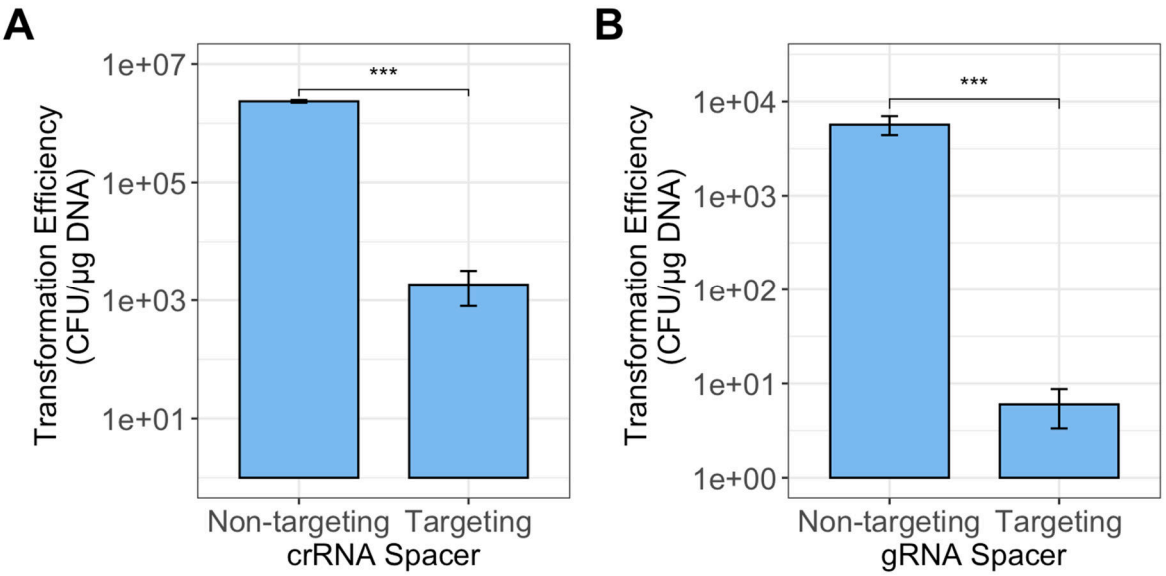


Figure S3.1. 3. Constitutive targeting at the *phaC* locus.

Targeting efficiency for **(A)** CRISPR-Cas12a and **(B)** CRISPR-Cas9 complexes, quantified as the transformation efficiency for plasmids that direct the constitutive expression of *Cas9* or *Cas12a*, paired with either a non-targeting or a targeting gRNA or crRNA spacer. Bars show the mean of *n* = 3, ± one standard deviation. *** = p<0.01, two-sample t-test.

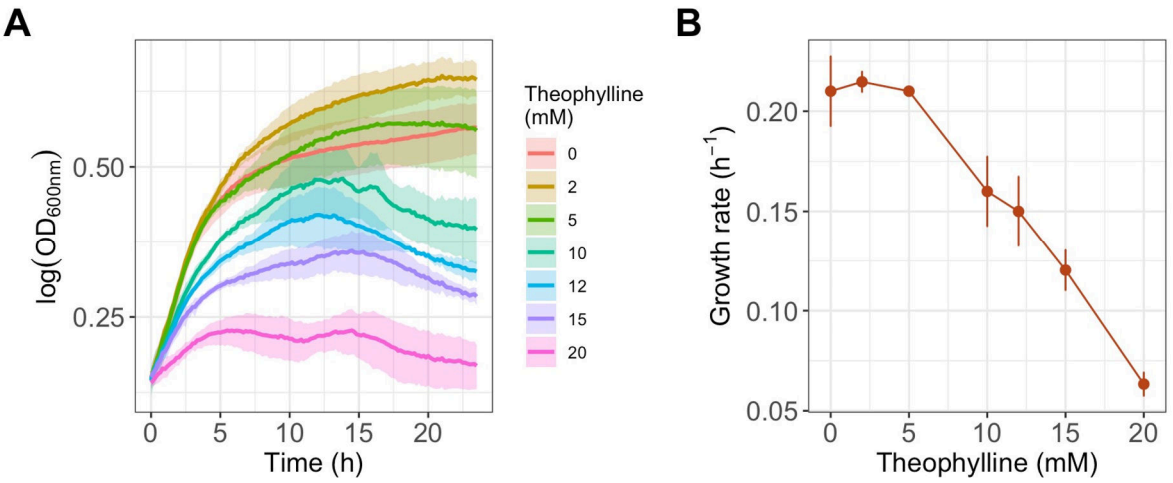


Figure S3.1. 4. Theophylline toxicity in *C. necator*.

(A) Growth curves from the theophylline minimum inhibitory concentration assay. Cells were cultured in rich medium (LB) with varying concentrations of theophylline, at the microplate scale. For each medium condition, solid lines show the average of *n* = 3 technical replicates, with the standard deviation indicated by the shaded region. **(B)** Corresponding growth rates, calculated as described previously[186]. Individual points show the mean of *n* = 3 technical replicates ± one standard deviation.

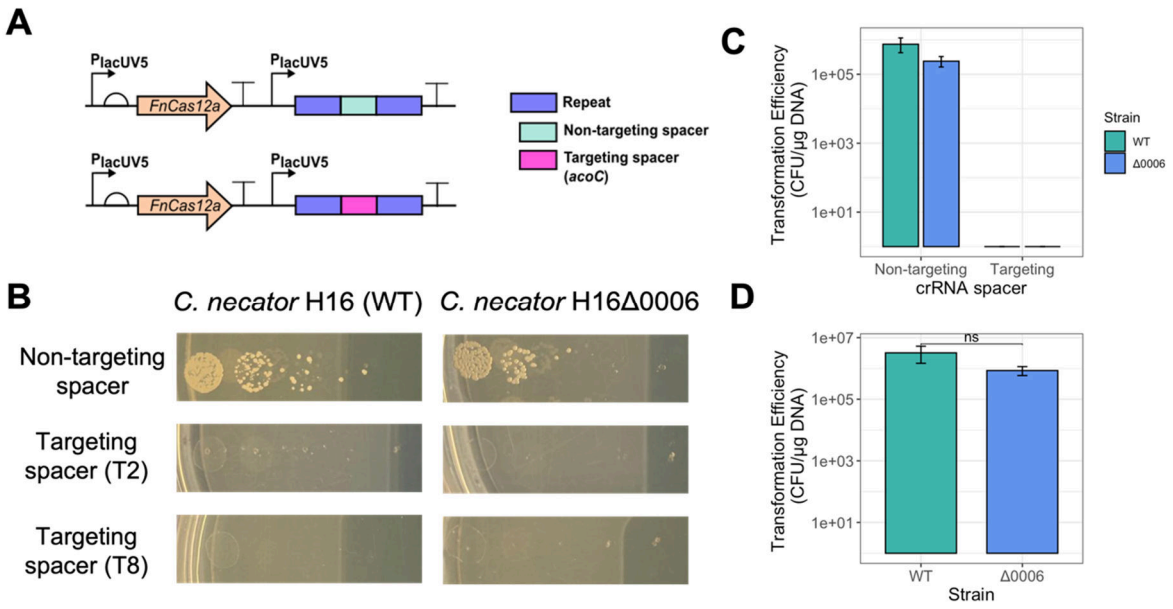


Figure S3.1. 5. Constitutive targeting of the *acoC* locus in *C. necator*Δ0006 by an active CRISPR-Cas12a complex.

(A) Schematic diagram of the pSIBR005 plasmids used in the constitutive targeting assay, which direct constitutive expression of both *Cas12a* and a crRNA cassette, harbouring either a non-targeting or an *acoC*-targeting spacer. Two versions of the latter were tested (T2 and T8). **(B)** Spot microdilutions of cultures transformed with the pSIBR005 plasmids. Each plasmid was transformed into both WT and Δ0006 *C. necator* strains. Each spot represents a 10-fold dilution over the previous column. Spacers T2 and T8 resulted in equivalent targeting efficiencies, with no CFUs being recovered for either strain following transformation (indicative of a targeting efficiency of 100%). **(C)** Transformation efficiency, quantified as CFU/µg DNA, for the non-targeting and *acoC*-targeting (T2) versions of pSIBR005. Bars indicate the mean of *n* = 3, ± one standard deviation.**(D)** Transformation efficiency of control plasmid pSEVA231 in WT and Δ0006 *C. necator*. ns = not significant, two-sample t-test.

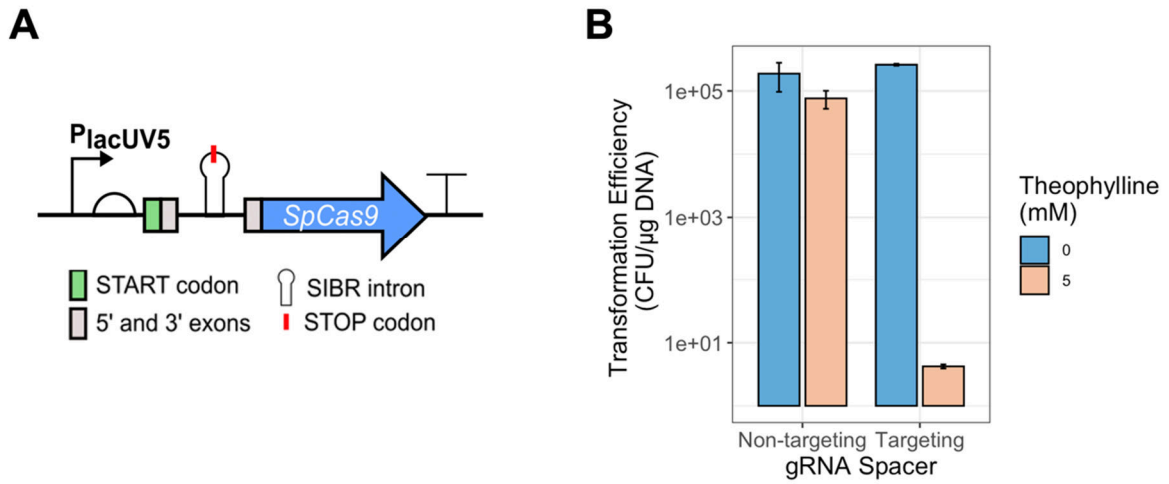


Figure S3.1. 6. SIBR-Cas9 enables inducible targeting.

(A) Schematic representation of the SIBR-Cas9 plasmids for inducible targeting, containing the SIBR intron (any of its four variants) at the canonical location immediately following the start codon. **(B)** Inducible targeting efficiency. Transformation efficiency for the non-targeting and *glcF*-targeting versions of pSIBR004-Cas9 (containing SIBR intron variant 4) across induction conditions (0 or 5mM theophylline). Bars indicate the mean of n = 3, ± one standard deviation.

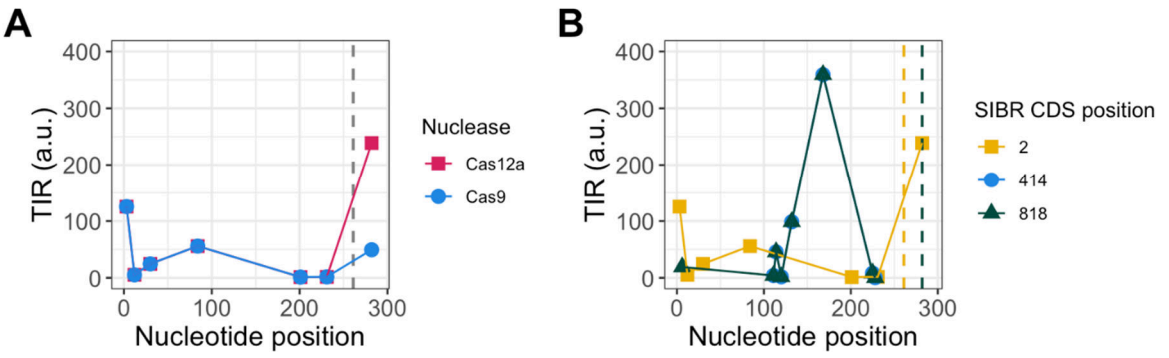


Figure S3.1. 7. Bioinformatic analysis of alternative RBSs within the SIBR intron.

(A) Predicted translation initiation rate (TIR) over the canonical SIBR intron sequence in SIBR-Cas12a and SIBR-Cas9 constructs, obtained using the RBS calculator[53,216,217]. The location of the final STOP codon is common to both constructs and shown with a grey dashed line. **(B)** Comparison of the predicted TIR over the intron sequence in SIBR-Cas12a and SIBR2.0-Cas12a constructs. The position of the final STOP codon is indicated with yellow (SIBR) or dark green (SIBR2.0) dashed lines.

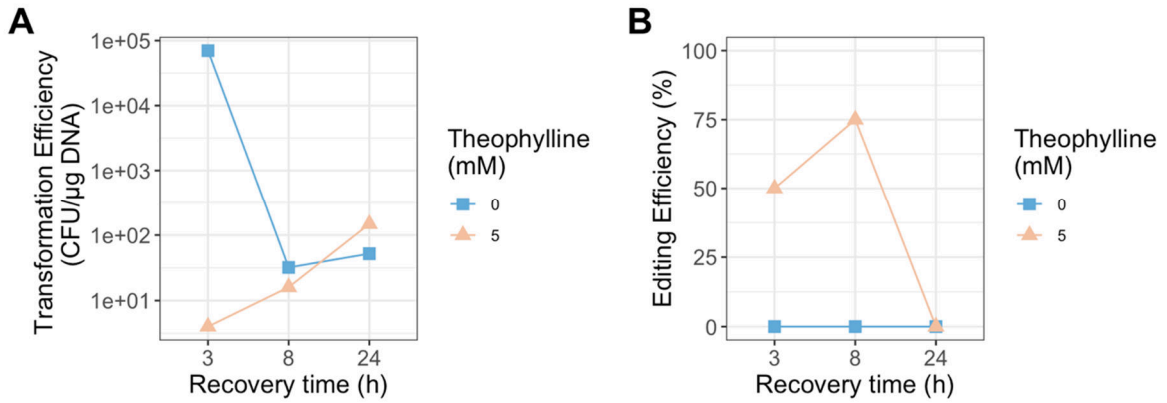


Figure S3.1. 8. Preliminary optimisation of recovery time in editing assays.

Plots show the **(A)** transformation efficiency and **(B)** editing efficiency obtained with the *acoC*-targeting version of editing plasmid pSIBR818, as a function of recovery time.

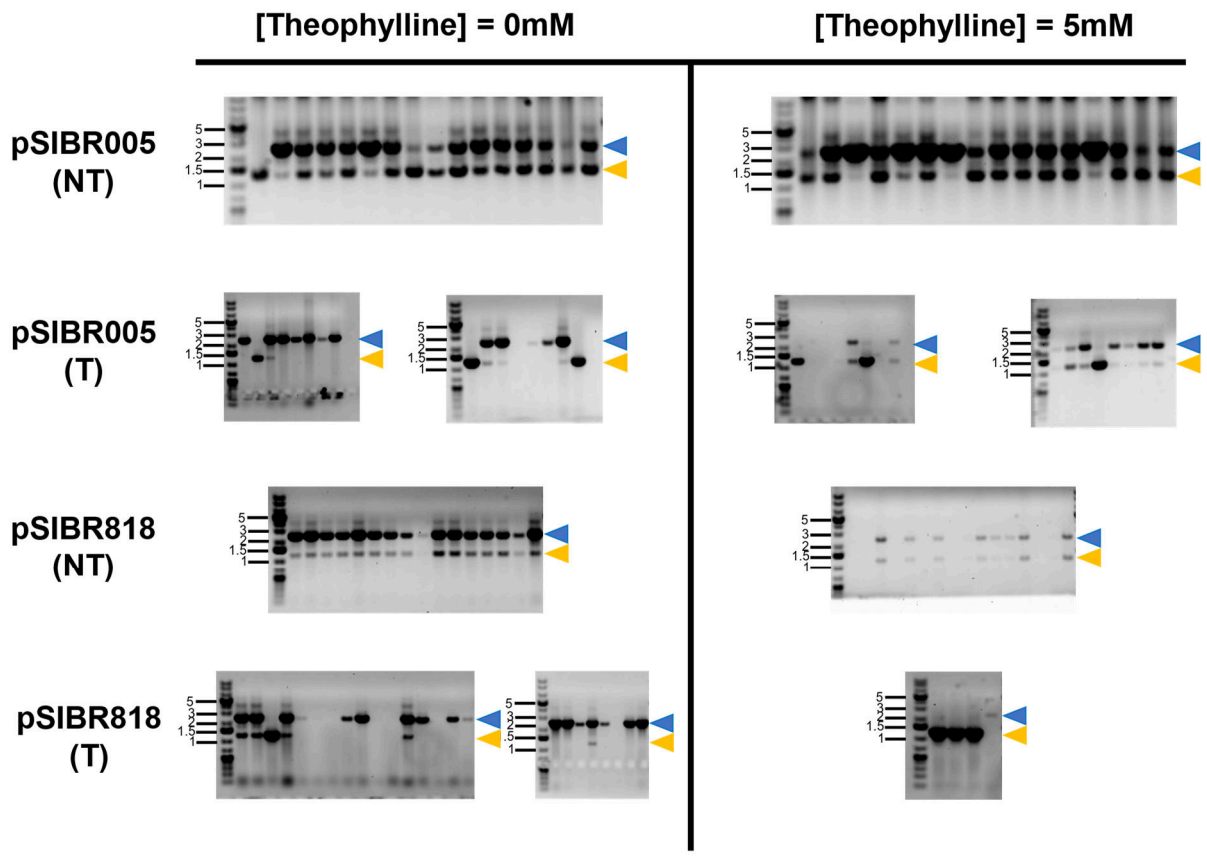


Figure S3.1. 9. cPCRs for quantification of editing efficiency.

For each plasmid and condition, representative agarose gels are shown, corresponding to one of three replicates of the editing experiment. Editing plasmids are labelled as NT = non-targeting or T=*acoC*-targeting. GeneRuler 1kb plus was used as the DNA ladder for all agarose gels. The size of relevant

bands in the ladder is indicated in kb. cPCR products corresponding to the WT (blue arrow) and edited (*acoC*-deletion, yellow arrow) loci are expected to be 2.3kb and 1.2kb in length, respectively.

Table S3.1. 1. Bacterial strains used in this study.

Strain	**Purpose**	**Genotype**
E. coli DH5α	Plasmid DNA storage and isolation	F– φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(rK^{-}, mK^{+}) *pho*A *sup*E44 λ–*thi*-1 *gyr*A96 *rel*A1
C. necator H16	Constitutive and inducible targeting assays	Wild-type *C. necator* strain
*C. necator*ΔphaC	Constitutive targeting assays	phaC deletion mutant (KEGG locus ID H16_A1437)
*C. necator*Δ0006ΔphaC	Constitutive and inducible targeting assays; editing assays.	Type I restriction modification system mutant (KEGG locus ID H16_A0006); *phaC* deletion mutant (KEGG locus ID H16_A1437).

Table S3.1. 2. Plasmids used in this study.

*Plasmid maps and sequences can be accessed via hyperlink. Sanger sequencing results are available in the linked files.

Plasmid*	Description	Ref.
pSEVA231	pBBR1 OriV-Rep; MCS; Kan^R	[177,178]
pSIBR005-mRFP	pBBR1 OriV-Rep; P_{lacUV5}; mRFP	This study
pSEVA231-P17-RBS13-mRFP	pBBR1 OriV-Rep; P_{17}; RBS_{13}; mRFP	This study
pSIBR005-NT	pSIBR005; *C. necator*-compatible non-targeting spacer.	This study
pSIBR005-phaC-T1	pSIBR005; *phaC*-targeting spacer T1	This study
pSIBR005-phaC-T2	pSIBR005; *phaC*-targeting spacer T2	This study
pSIBR005-phaC-T3	pSIBR005; *phaC*-targeting spacer T3	This study
pSIBR005-acoC-T2	pSIBR005; aco*C*-targeting spacer T2	This study
pSIBR005-acoC-T8	pSIBR005; aco*C*-targeting spacer T8	This study
pSIBR001-NT	pSIBR001; *C. necator*-compatible non-targeting spacer.	This study
pSIBR002-NT	pSIBR002; *C. necator*-compatible non-targeting spacer.	This study
pSIBR003-NT	pSIBR003; *C. necator*-compatible non-targeting spacer.	This study
pSIBR004-NT	pSIBR004; *C. necator*-compatible non-targeting spacer.	This study
pSIBR001-phaC-T3	pSIBR001; *phaC*-targeting spacer T3.	This study
pSIBR002-phaC-T3	pSIBR002; *phaC*-targeting spacer T3.	This study
pSIBR003-phaC-T3	pSIBR003; *phaC*-targeting spacer T3.	This study
pSIBR004-phaC-T3	pSIBR004; *phaC*-targeting spacer T3.	This study
pSIBR001-acoC-T2	pSIBR001; aco*C*-targeting spacer T2.	This study
pSIBR002-acoC-T2	pSIBR002; *acoC-targeting spacer T2.*	This study
pSIBR003-acoC-T2	pSIBR003; *acoC-targeting spacer T2.*	This study
pSIBR004-acoC-T2	pSIBR004; *acoC-targeting spacer T2.*	This study
pSIBR00X-NT	pSIBR001 backbone with additional STOP codon in 5' exon; *C. necator*-compatible non-targeting spacer.	This study
pSIBR00X-T2	pSIBR001 backbone with additional STOP codon in 5' exon; aco*C*-targeting spacer T2.	This study
pSIBR414-NT	pSIBR005 backbone; SIBR2.0 intron at amino acid position 414; *C. necator*-compatible non-targeting spacer.	This study
pSIBR818-NT	pSIBR005 backbone; SIBR2.0 intron at amino acid position 818; *C. necator*-compatible non-targeting spacer.	This study
pSIBR414-acoC-T2	pSIBR005 backbone; SIBR2.0 intron at amino acid position 414; *acoC*-targeting spacer T2.	This study
pSIBR818-acoC-T2	pSIBR005 backbone; SIBR2.0 intron at amino acid position 818; acoC-targeting spacer T2.	This study
pSIBR005-NT-acoC_edit	pSIBR005; *C. necator*-compatible non-targeting spacer; *acoC* homology arms.	This study
pSIBR005-T2-acoC_edit	pSIBR005; aco*C*-targeting spacer T2; acoC homology arms.	This study
pSIBR818-NT-acoC_edit	pSIBR005 backbone; SIBR2.0 intron at amino acid position 818; C. necator-compatible non-targeting spacer; acoC homology arms.	This study
pSIBR818-T2-acoC_edit	pSIBR005 backbone; SIBR2.0 intron at amino acid position 818; acoC-targeting spacer T2; acoC homology arms.	This study

3.8. Chapter 3 Appendix 2 – Python Software for Spacer Design

All SIBR-Cas plasmids follow a standardised architecture, which is compatible with Golden-Gate assembly of crRNA spacers for targeting and editing assays. This requires new SIBR users to be familiar with both CRISPR-Cas9 or CRISPR-Cas12a spacer design and Golden-Gate assembly. To lower the barrier for entry and facilitate the use of SIBR-Cas in the wider *C. necator* research community, we developed a simple software that enables users to identify suitable spacer sequences within their target locus. For each spacer, the Python script outputs the oligonucleotide sequences that users should synthesise to assemble spacers, such that they can be cloned into SIBR plasmids via a streamlined Golden-Gate protocol. The script can be used to generate spacers for either Cas9 or Cas12a, implementing the following algorithmic steps:

1. Parse DNA sequence of target locus from a FASTA file provided by the user.
2. Identify all Cas9 or Cas12a (as appropriate) protospacer adjacent motif (PAM) sequences within the target locus, on either the forward or reverse DNA strands.
3. Use the PAM locations to extract all possible spacer sequences.
4. Exclude any spacer sequences which are shorter than 20bp.
5. Exclude any spacer sequences which contain BbsI restriction sites (this is the enzyme that is used to assemble spacers into the SIBR plasmid backbones).

6. For Cas12a spacers, exclude any spacers that start with a ‘T’, to reflect known sequence biases which affect targeting efficiency[210].
7. For each spacer, design the Golden Gate-compatible oligonucleotide sequences (forward and reverse) by adding the necessary DNA prefix and suffix containing BbsI restriction sites and standardised overhangs.
8. For each spacer, count the occurrences of the full spacer sequence (including the PAM) on both *E. coli* and *C. necator* genomes.
9. For each spacer, count the occurrences of the spacer’s seed sequence (PAM+10bp of the spacer sequence closest to the PAM) on both *E. coli* and *C. necator* genomes.
10. Assemble all sequences and counts into a single dataframe (table).
11. Output the final table to a CSV file.

This versatile script is designed to give users a choice of spacers, from which they can identify a suitable set of sequences to be implemented experimentally. Where possible, users should exclude any spacer sequences which occur at multiple sites on the *C. necator* genome. The absence of additional occurrences on the *E. coli* genome may also facilitate cloning. A myriad of algorithms have been developed to compute optimal Cas9 or Cas12a spacers for any given DNA sequence, which users may choose to consult before selecting a set of spacers for their targeting or editing experiments. The main benefit of our software is that, for any given spacer, it provides the oligonucleotide sequences that users can directly copy and paste into their favourite DNA synthesis company’s website.

The software is open source and can be easily modified by users, for example to enable the design of spacers for multiple loci at once. At present, two different implementations of the Python script are available: one can be run entirely from the command line (Figure S3.2.1 A) and another encodes a graphical user interface (Figure S3.2.1 B). The output table generated by either script is identical, with an example shown in Figure S3.2.1 C. All script files will be made publicly available on GitHub upon publication. Currently, the scripts are available at the following address: https://drive.google.com/drive/folders/1kGWEv39TSMa0EUDDOsGr9E3r-zw0NgK2?usp=share_link.

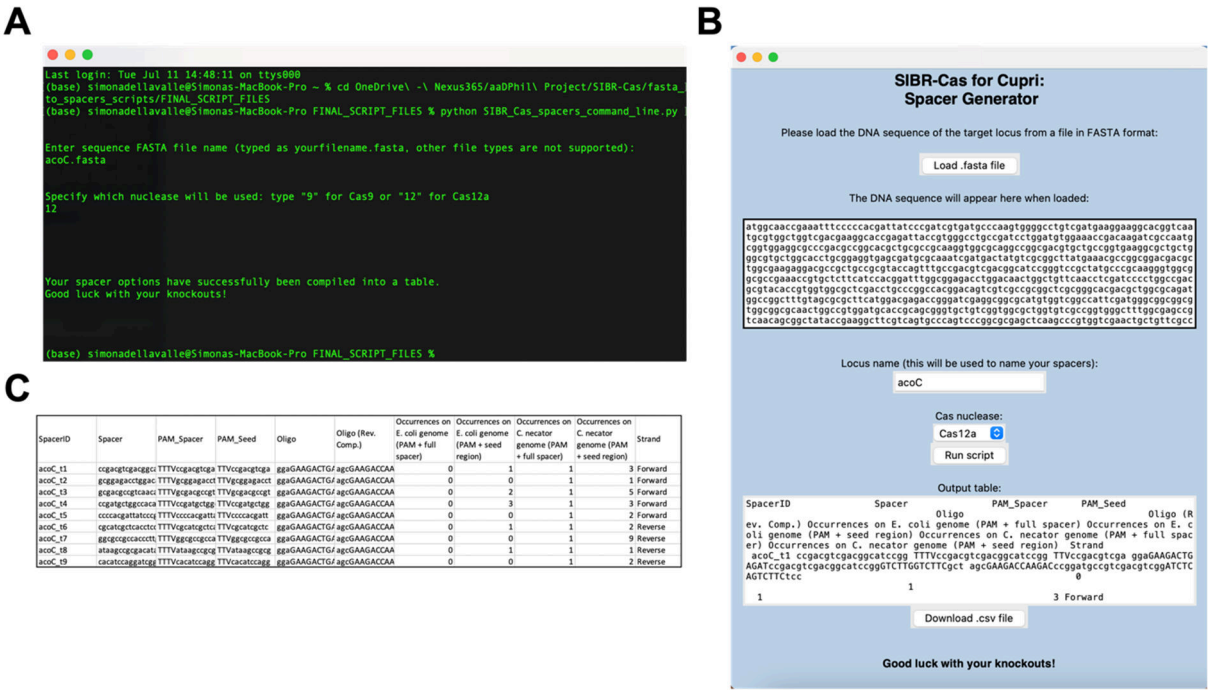


Figure S3.2. 1. Implementations of the Python software.

(A) The script can be executed interactively from the command line or **(B)** from a purpose-built graphical user interface. **(C)** The output of the script is a CSV file containing a table of all possible spacers with their corresponding features and oligonucleotide sequences.

Table S3.2. 1. Script-generated crRNA spacers for *phaC* targeting.

Spacer	Sequence	Oligo	Oligo (Rev. Comp.)	PAM + full spacer (*E. coli*)	PAM + seed region (*E. coli*)	PAM + full spacer (*C. necator*)	PAM + seed region (*C. necator*)
phaC_t1	aggtcggccg caatgtcgcg	ggaGAAGACTGAGATaggtcggccg caatgtcgcgGTCTTGGTCTTCgct	agcGAAGACCAAGACcgcgacattg cggccgacctATCTCAGTCTTCtcc	0	0	1	7
phaC_t2	ccgacacgg gcatcctcgac	ggaGAAGACTGAGATccgacacggg catcctcgacGTCTTGGTCTTCgct	agcGAAGACCAAGACgtcgaggatg cccgtgtcggATCTCAGTCTTCtcc	0	0	1	1
phaC_t3	tggtgtcgtgg cgcaatccg	ggaGAAGACTGAGATtggtgtcgtgg cgcaatccgGTCTTGGTCTTCgct	agcGAAGACCAAGACcggattgcgc cacgacaccaATCTCAGTCTTCtcc	0	0	1	1
phaC_t4	acgtatcgccc aggcgcggg	ggaGAAGACTGAGATacgtatcgccc aggcgcgggGTCTTGGTCTTCgct	agcGAAGACCAAGACcccgcgcctg ggcgatacgtATCTCAGTCTTCtcc	0	1	1	1
phaC_t5	gcgccggcct gcccggccag	ggaGAAGACTGAGATgcgccggcct gcccggccagGTCTTGGTCTTCgct	agcGAAGACCAAGACctggccgggc aggccggcgcATCTCAGTCTTCtcc	0	0	1	16

Table S3.2. 2. Script-generated crRNA spacers for *acoC* targeting.

Spacer	Sequence	Oligo	Oligo (Rev. Comp.)	PAM + full spacer (*E. coli*)	PAM + seed region (*E. coli*)	PAM + full spacer (*C. necator*)	PAM + seed region (*C. necator*)
acoC_t1	ccgacgtcga cggcatccgg	ggaGAAGACTGAGATccgacgtcga cggcatccggGTCTTGGTCTTCgct	agcGAAGACCAAGACccggatgccg tcgacgtcggATCTCAGTCTTCtcc	0	1	1	3
acoC_t2	gcggagacct ggacaactgg	ggaGAAGACTGAGATgcggagacct ggacaactggGTCTTGGTCTTCgct	agcGAAGACCAAGACccagttgtcca ggtctccgcATCTCAGTCTTCtcc	0	0	1	1
acoC_t3	gcgacgccgt caacagcggc	ggaGAAGACTGAGATgcgacgccgt caacagcggcGTCTTGGTCTTCgct	agcGAAGACCAAGACgccgctgttga cggcgtcgcATCTCAGTCTTCtcc	0	2	1	5
acoC_t4	ccgatgctggc cacatgagc	ggaGAAGACTGAGATccgatgctggc cacatgagcGTCTTGGTCTTCgct	agcGAAGACCAAGACgctcatgtggc cagcatcggATCTCAGTCTTCtcc	0	3	1	3
acoC_t5	ccccacgatta tcccgatcg	ggaGAAGACTGAGATccccacgatta tcccgatcgGTCTTGGTCTTCgct	agcGAAGACCAAGACcgatcgggat aatcgtggggATCTCAGTCTTCtcc	0	0	1	2
acoC_t6	cgcatcgctca cctccgcag	ggaGAAGACTGAGATcgcatcgctca cctccgcagGTCTTGGTCTTCgct	agcGAAGACCAAGACctgcggaggt gagcgatgcgATCTCAGTCTTCtcc	0	1	1	2
acoC_t7	ggcgccgcca cccttgcggg	ggaGAAGACTGAGATggcgccgcca cccttgcgggGTCTTGGTCTTCgct	agcGAAGACCAAGACcccgcaagg gtggcggcgccATCTCAGTCTTCtcc	0	0	1	9
acoC_t8	ataagccgcg acatagtcat	ggaGAAGACTGAGATataagccgcg acatagtcatGTCTTGGTCTTCgct	agcGAAGACCAAGACatgactatgtc gcggcttatATCTCAGTCTTCtcc	0	1	1	1
acoC_t9	cacatccagg atcggcaggc	ggaGAAGACTGAGATcacatccagg atcggcaggcGTCTTGGTCTTCgct	agcGAAGACCAAGACgcctgccgat cctggatgtgATCTCAGTCTTCtcc	0	0	1	2

4. A SIBR2.0-based genetic AND gate for tight dual-level control of gene expression in *Cupriavidus necator*

4.1. Aims and Objectives

During the development of the SIBR-Cas toolkit for *C. necator*, we identified that SIBR-mediated control of *Cas12a* expression was defective (Chapter 3, Section 3.5.2). In light of this, we explored alternative options for inducible expression of *Cas12a*, including the Jungle Express system, which controls gene expression at the transcription level. We found that the system operates with a high basal transcription rate in *C. necator*, precluding its use in genome editing applications. Later, having resolved the defective regulation of *Cas12a* expression through the development of SIBR2.0, we investigated whether SIBR2.0 could be used in conjunction with Jungle Express to implement a genetic AND logic gate. In doing so, we aim to (i) demonstrate that the SIBR2.0 intron can be used to lower the basal transcription rate from leaky inducible promoters, and (ii) validate a generalisable regulatory architecture that enables tight control of gene expression at the transcriptional and translational level. This has not been previously demonstrated in *C. necator*, and could be useful to install more complex genetic circuits. The present Chapter details the characterisation of the JEx system in *C. necator*, as well as the pursuit of the aims stated above.

4.2. Introduction

Implementing novel biological functions in any given host organism requires precise control over gene expression. Naturally occurring regulatory modules, such as bacterial allosteric transcription factors and their cognate promoters, have been extensively developed to deliver inducible gene expression systems, exerting control at the level of transcription or translation. Once extracted from their environmental context, such modules can be refactored to create DNA parts which recognise specific inputs and deliver fine-tuned response functions. Though reports of optimised inducible regulatory modules abound in the literature, their performance is often host-dependent and requires thorough characterisation to be effectively applied in genomic contexts other than the one in which they were first developed.

Inducible expression systems are fundamental in strain engineering. Precise temporal control over the expression of genes along a biosynthetic pathway is most often required to attain bioproduction goals[222]. Promoters which are selectively activated or repressed by the pathway's substrates, intermediates and products can be implemented[223,224]. The strength of their gene expression output may also be modulated to fine-tune the relative stoichiometries of pathway enzymes and their reaction products, in an effort to optimise yields and limit metabolic burden[225,226]. Where pathway intermediates or products are toxic to the host organism, regulatory modules must enable tightly inducible control of gene expression, operating with minimal leakiness[227]. Minimising or eliminating basal expression rates is also important in the context of bacterial genome engineering. For instance, tools that rely on homologous recombination and CRISPR-Cas counterselection require strict control

over the synthesis of Cas nucleases, since leaky expression may prove cytotoxic or lead to inefficient editing[209].

In Chapter 3, we describe how the SIBR-Cas tool enables tight control of *Cas12a* and *Cas9* expression in *C. necator*, and apply it to perform gene knockouts with high efficiency. During the development of this genome editing protocol, we discovered that the canonical SIBR intron was not suitable to mediate tightly inducible control of *Cas12a* expression (Section 3.5.2). Though we subsequently identified both a likely cause (Section 3.5.3) and a useful solution (Section 3.5.4) for this defective regulation, we simultaneously explored the use of alternative genetic regulatory modules.

Specifically, we investigated whether the Jungle Express (JEx) system would be a suitable alternative to tightly control the expression of *Cas12a* for genome editing applications. In JEx, the transcriptional repressor EilR mediates tightly inducible gene expression. First identified in the rainforest-dwelling bacterium *Enterobacter lignolyticus*[228,229], the EilR repressor has been developed as a versatile and portable regulatory component. Its functionality has been demonstrated in *E. coli*, as well as the non-enteric proteobacteria *Pseudomonas putida, Sinorhizobium meliloti* Rm1021 and *Caulobacter crescentus* NA1000[230]. Several cationic dyes, such as crystal violet, can be used as low-cost inducers to release EilR repression. Owing to its reported low basal transcription rate and demonstrable functionality in several proteobacteria (i.e. portability), we identified JEx as a promising regulatory framework to mediate inducible expression of CRISPR-Cas12a complexes in non-model bacteria such as *C. necator*.

Here, we provide a detailed characterisation of the JEx regulatory module in *C. necator*. We demonstrate that the system can deliver inducible gene expression in this organism, thereby adding a useful DNA part to the bacterium's small but expanding synthetic biology toolkit. However, our investigation revealed a high basal transcription rate, indicating that JEx may not be suitable to mediate the expression of cytotoxic genes (such as *Cas12a*), which are applied in genome editing tools. Using *Cas12a* expression as a test system, we explore two options minimse the leakiness of JEx: modulating RBS strength and implementing a genetic AND gate. For the latter, we used JEx in conjunction with the SIBR2.0 intron developed in Chapter 3 to achieve tight control of gene expression at both the transcription and translation level.

4.3. Materials and Methods

Bacterial strains and culture conditions

All bacterial strains used in this study are listed in Table S4.1. Plates for bacterial growth were LB agar. Liquid media was LB, Super Optimal Broth supplemented with 20 mM fructose (SOF), or M9 minimal medium supplemented with 20 mM fructose, as indicated. Where appropriate, kanamycin was added to solid or liquid growth media at 100 µg/ml (selective medium). All *E. coli* strains were incubated at 37 °C with vigorous shaking (200 rpm). *C. necator* strains were incubated at 30 °C with moderate shaking (150 rpm). Optical density (OD) of bacterial cultures was measured using a UV-1800 UV/Vis spectrophotometer (Shimadzu).

Assembly of JEx-Cas plasmids

All plasmids used in this study are listed in Table S4.2. Plasmid pJC580 was a gift from the group of Michael P. Thelen[230]. Unless otherwise stated, plasmids were isolated from *E.* coli DH5α, using the NEB Monarch® Miniprep Kit according to the manufacturer's specifications. Plasmid pJEx00S was assembled by amplifying the backbone of the non-targeting version of plasmid pSIBR005[209] (excluding the P_{lacUV5} promoter region immediately upstream of the *Cas12a* CDS) and the *eilR*-P_{DE20} cassette from plasmid pJC580. All PCRs were conducted using Q5® High-Fidelity DNA polymerase (New England BioLabs, NEB), following the manufacturer's instructions. PCR products were checked by agarose gel electrophoresis and purified using the Monarch® PCR & DNA Cleanup Kit (NEB). Gibson assembly of double-stranded DNA fragments was carried out using the NEBuilder® HiFi DNA Assembly (NEB). To construct the pJEx00S RBS variant library, the backbone of pJEx00S (excluding the RBS immediately preceding the *Cas12a* CDS) was amplified by PCR. The synthetic RBS variants were encoded in PCR primer tails (approximately half of the RBS sequence on the forward primer tail and the rest on the reverse primer tail). Following gel electrophoresis, PCR products were purified and circularised via blunt-end ligation using the KLD Enzyme Mix (NEB). To construct plasmid pJEx00N, the backbone of plasmid pJEx00S (excluding the *Cas12a* CDS) and the SIBR2.0-containing *Cas12a* CDS from plasmid pSIBR818 (*see* Chapter 3) were amplified and assembled (Gibson). For all plasmids constructed in this study, correct assembly was confirmed via Sanger sequencing.

Flow cytometry for single-timepoint fluorescence measurements

Strains carrying test and control plasmids were cultured overnight (16-18 h) in selective LB medium. All cultures used in these experiments had a total volume of 5 ml, in 50 ml conical centrifuge tubes. Overnight precultures were used to inoculate fresh cultures in selective mineral medium (M9), with 20 mM fructose as sole carbon source, at an OD of 0.05-0.1. Cells were grown to mid-exponential phase (OD 0.2) before adding the inducer (crystal violet) at a range of concentrations between 0 and 200 nM. Cultures were further incubated to an OD of 0.4-0.6, at which point 1 ml of each culture was transferred to 1.5 ml microcentrifuge tubes. Cells were pelleted by centrifugation and washed twice in phosphate buffered saline (PBS). After the final wash, cell pellets were resuspended in 1 ml PBS, then diluted in PBS to an OD of 0.01. The cells were analysed using a BD FACSCalibur flow cytometer (BD Biosciences). mRFP fluorescence was measured with a 488 nm laser and a 585/42 nm emission band-pass filter (corresponding to the instrument's FL2 channel). The voltage of the FL2 detector was set to 705 V and the amplitude gain was adjusted to 1.0. At least 100,000 events were collected for each sample. Flow cytometry data was analysed using the proprietary FlowJo software (BD Biosciences).

CRISPR-Cas targeting assays

To measure the targeting efficiency of inducibly-expressed CRISPR-Cas12a complexes, the transformation efficiency of non-targeting and targeting versions of JEx-Cas plasmids was quantified. Targeting crRNA spacers for the *acoC* locus were assembled as described in Chapter 3. Plasmids were electroporated into *C. necator*.

Electrocompetent cells were prepared and transformed following the protocol outlined in Chapter 2. Recovery was performed in 1 ml cultures within 1.5 ml microcentrifuge tubes for 2-4h. Transformation efficiency was quantified as CFU/µg DNA via spot microdilution on selective LB agar plates (100 µg/ml kanamycin) containing the appropriate inducer. Where relevant, crystal violet was added at concentrations in the range of 0-1 µM (as indicated), and theophylline was added invariably at 5 mM. Transformation plates were incubated at 30 ºC for 48 h before single colonies could be counted.

4.4. Results

4.4.1. JEx induction kinetics in *C. necator*

First, we characterised the induction kinetics of the JEx system in *C. necator* to determine which crystal violet concentrations could be used to induce gene expression in this specific genomic context. We used plasmid pJC580 for this purpose, which encodes JEx-controlled expression of the fluorescent reporter gene mRFP (Figure 4.1 A). By quantifying mRFP fluorescence over a range of crystal violet concentrations, we determined that the JEx system is suitable to mediate inducible gene expression in *C. necator* (Figure 4.1 B). The dose-response curve obtained in this organism is similar to that reported for *E. coli* (*see* ref. [230] Figure 2). Specifically, we observed that concentrations of crystal violet of ≥ 50 nM led to saturation of the gene expression output. At inducer concentrations <20 nM, no fluorescence could be detected; the probability distribution of mRFP fluorescence for cells in these test conditions was

superimposable to the negative control (Figure S4.1), confirming the reported low basal transcription rate of the JEx system[230].

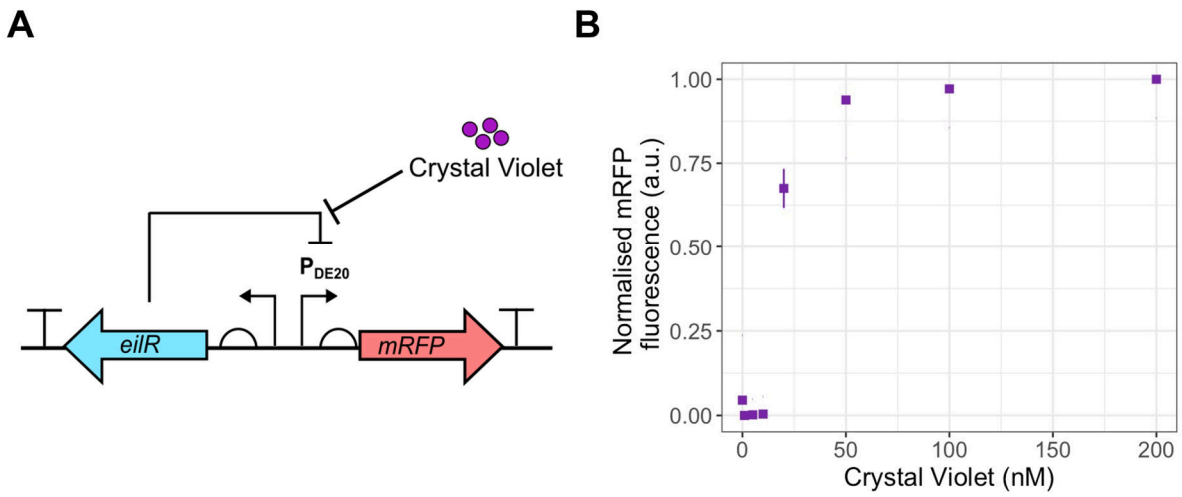


Figure 4. 1. JEx induction kinetics in *C. necator.*

(A) Plasmid pJC580 implements the JEx transcriptional regulation system for control of *mRFP* expression. **(B)** Normalised median mRFP fluorescence measurements over a range of crystal violet concentrations. Individual points show the mean of *n* = 3, ± one standard deviation.

4.4.2. Design of JEx-Cas: JEx constructs for inducible *Cas12a* expression

Having characterised the induction kinetics of JEx in *C. necator*, we next investigated whether this regulatory module could be used to tightly control the expression of *Cas12a*, i.e. whether it would be a suitable DNA part to replace the defective SIBR intron in SIBR-Cas. Plasmids to implement JEx-mediated *Cas12a* expression (JEx-Cas) were designed to closely mirror the SIBR-Cas architecture (Figure 4.2). Briefly, the universal pSIBR005 plasmid backbone was directly inherited, including the origin of replication, kanamycin-resistance marker, *Cas12a* CDS and terminator, crRNA

cassette, and Esp3I restriction sites for plasmid linearisation and Gibson assembly of homology arms[209]. The region of the pSIBR005 plasmid containing the P_{lacUV5} promoter for constitutive *Cas12a* expression was replaced by the JEx system. This comprises a cassette for constitutive expression of *eilR*, as well as its cognate P_{DE20} promoter and accompanying RBS. The resulting plasmid, pJEx00S, can be used for crystal violet-inducible *Cas12a* expression.

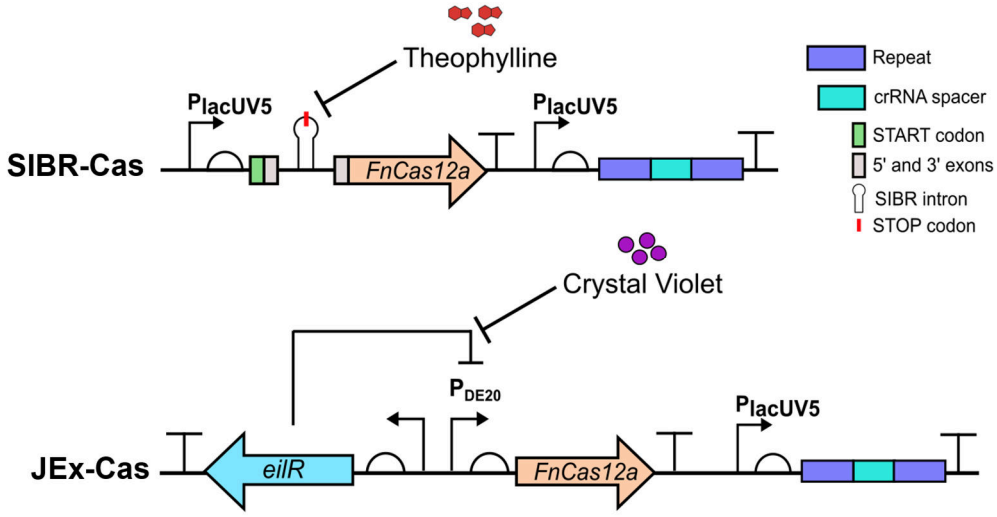


Figure 4. 2. The genetic architecture of SIBR-Cas and JEx-Cas.

In SIBR-Cas, the P_{lacUV5} promoter drives constitutive expression of both *Cas12a* and the crRNA cassette. Regulation of *Cas12a* expression is mediated at the translation level by the self-splicing, theophylline-inducible SIBR intron. No additional genes, such as transcription factors, are needed. In JEx-Cas, expression of *Cas12a* is driven by the P_{DE20} promoter, which is tightly controlled by the crystal violet-inducible repressor EilR. Both *eilR* and the crRNA cassette are constitutively expressed.

4.4.3. Leaky control of *Cas12a* expression by JEx-Cas

Full JEx-Cas induction is cytotoxic

We designed versions of the JEx-Cas constructs (pJEx00S) to mediate inducible CRISPR-Cas12a targeting at the *acoC* locus. These plasmids pair EilR-controlled *Cas12a* expression with either a non-targeting or *acoC*-targeting spacer sequence within the crRNA cassette (*see* Chapter 3 for details on spacer sequences). Both versions of the pJEX00S plasmid were electroporated into *C. necator* to conduct an inducible targeting assay, where transformants were recovered on selective LB agar plates (100 µg/ml kanamycin) with and without crystal violet. The inducer was added at a final working concentration of 1 µM, as previously reported for *E. coli*[230]. For all induction conditions, CRISPR-Cas12a targeting was quantified as the difference in transformation efficiency between non-targeting and *acoC*-targeting versions of the pJEx00s plasmid (Figure 4.3 A).

In uninduced conditions (0 µM crystal violet), the targeting construct was transformed with lower efficiency than its non-targeting counterpart, suggesting leaky *Cas12a* expression. At 1 µM crystal violet, a ~10^1 reduction in transformation efficiency was observed for the targeting construct relative to the uninduced condition, consistent with CRISPR-Cas12a activity. However, this drop in transformation efficiency was also observed for the non-targeting construct. This suggests that the measured “targeting” efficiency may not only be due to CRISPR-Cas12a cleavage at the specified genomic locus, but that JEx-Cas constructs may become cytotoxic upon crystal violet addition.

A crystal violet induction gradient reveals leaky EilR-mediated repression in JEx-Cas

Consistent with the dose-response curve shown in Figure 4.1 B, we expect that EilR repression would be fully relieved when 1 µM crystal violet is added, i.e. that transcriptional output from the P_{DE20} promoter would be maximal. Full induction would therefore lead to a large number of mRNA transcripts and Cas12a protein molecules being produced, possibly exceeding the number of crRNA molecules in the cell. It is known that, when large amounts of Cas12a are being produced such that some nuclease molecules are lacking their cognate crRNA, off-target genomic cleavage is likely to occur[231]. We hypothesised that if JEx-Cas cytotoxicity were due to an overwhelming number of Cas12a molecules being produced, this could be avoided by modulating the strength of induction. We assume that there is a threshold crystal violet concentration, below which the release of EilR from its cognate operator sequence in the P_{DE20} promoter is not stable enough to allow for a too-large number of transcripts and protein molecules to be produced.

We tested this hypothesis (and its underlying assumption) by conducting an extended targeting assay, using different concentrations of crystal violet to implement an induction gradient (Figure 4.3 B). The results confirmed that at 0 mM crystal violet the transformation efficiency of the targeting constructs is substantially lower than that of its non-targeting counterpart, consistent with leaky expression of *Cas12a*. Nevertheless, a reduction in transformation efficiency for the targeting plasmid was observed at inducer concentrations ≥ 5 nM, indicating that a small amount of crystal violet is sufficient to induce CRISPR-Cas12a targeting.

For non-targeting constructs, transformation efficiency remained high at inducer concentrations ≤50 nM, and decreased only when higher amounts of crystal violet were added. This confirms that, in line with our hypothesis, modulating transcriptional regulation by the EilR repressor is important to avoid cytotoxicity of JEx-Cas constructs.

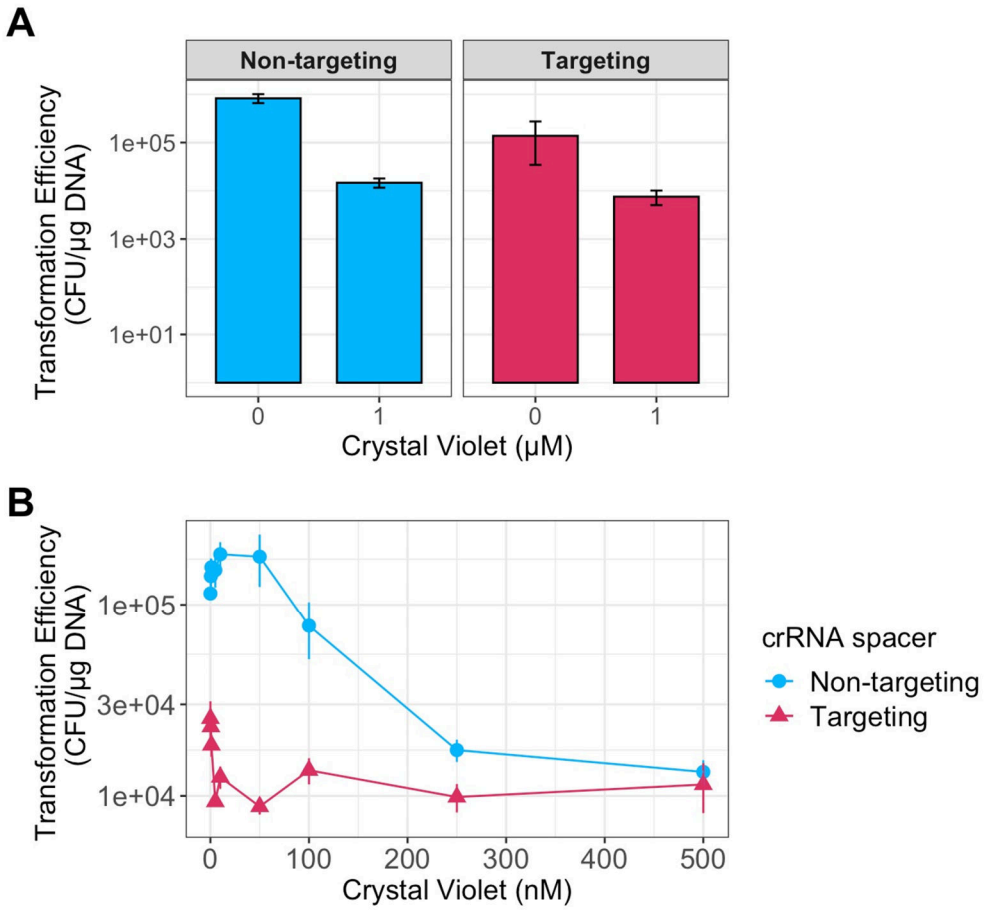


Figure 4. 3. JEx-Cas inducible targeting assays.

(A) Transformation efficiency for non-targeting and *acoC*-targeting versions of pJEx00S, quantified as CFU/µg DNA under binary induction conditions (0 or 1 µM crystal violet). **(B)** Transformation efficiency for pJEx00S constructs, as quantified over an induction gradient (0-500 nM crystal violet). Bars and individual points represent the mean of *n* = 3, ± one standard deviation.

4.4.4. Reducing RBS strength does not alleviate leaky EilR repression

The results obtained for the JEx-Cas induction gradient showed that cytotoxicity of non-targeting constructs could be overcome by using lower concentrations of the inducer (≤50 nM), which still allow for inducible targeting with the targeting version of pJEx00S. However, leaky *Cas12a* expression must also be addressed in order to successfully apply the JEx-Cas system to genomic manipulation. Other studies have used a range of tactics to tighten transcriptional control of Cas nuclease genes, including increasing the strength of constitutive promoter controlling expression of repressor, adding degradation tags (*ssrA*), or reducing the strength of the RBS preceding the Cas nuclease CDS[232,233].

We first investigated whether we could tighten *Cas12a* expression in the JEx-Cas system by optimising the strength of the RBS preceding the *Cas12a* CDS in pJEx00S. We designed a library of RBS variants, all of which were predicted to have maximal translation initiation rates (TIR) below that of the P_{DE20}-associated RBS (65800 arbitrary units). The RBS variant library comprised a total of ten synthetic parts, labelled A-J. RBS variants A, C and D correspond to synthetic sequences which had been previously characterised in *C. necator*[221]. The remaining variants (B, E, F, G, H, I, J) were designed using the RBS calculator[53,56,216,217]. These RBS sequence variants were specifically selected to cover a range of TIRs over two orders of magnitude (Figure 4.4 A). Redundancy was built into the library. Namely, several RBS variants with TIRs that were ten-fold lower than that of the original construct were included. A reduction of this magnitude has previously been shown to successfully tighten

transcriptional regulation[232]. We constructed both non-targeting and *acoC*-targeting versions of the pJEx00S plasmid containing the selected synthetic RBS sequences. The final library only contained RBS variants A, B, C, D, G and H, as we failed to correctly assemble variants E, F, I and J into the pJEx00S backbone.

All plasmid variants were transformed into *C. necator* to conduct an inducible targeting assay over a crystal violet gradient (Figure 4.4 B). For non-targeting constructs, only RBS variants C and D successfully reduced the strength of expression, as no drop in transformation efficiency was observed with increasing concentrations of the inducer (>50 nM). All other RBS variants followed the same induction pattern as the original pJEx00S RBS, suggesting no correlation between *in silico* predictions for TIR and *in vivo* performance. For targeting constructs, none of the RBS variants alleviated leaky *Cas12a* expression in uninduced conditions. These observations suggest that modulating RBS strength is not sufficient to overcome the leakiness of EilR repression, which results in the cytotoxicity of targeting pJEx00S constructs even in the absence of the inducer. A more detailed explanation of this paradigm is provided in Figure S4.2.

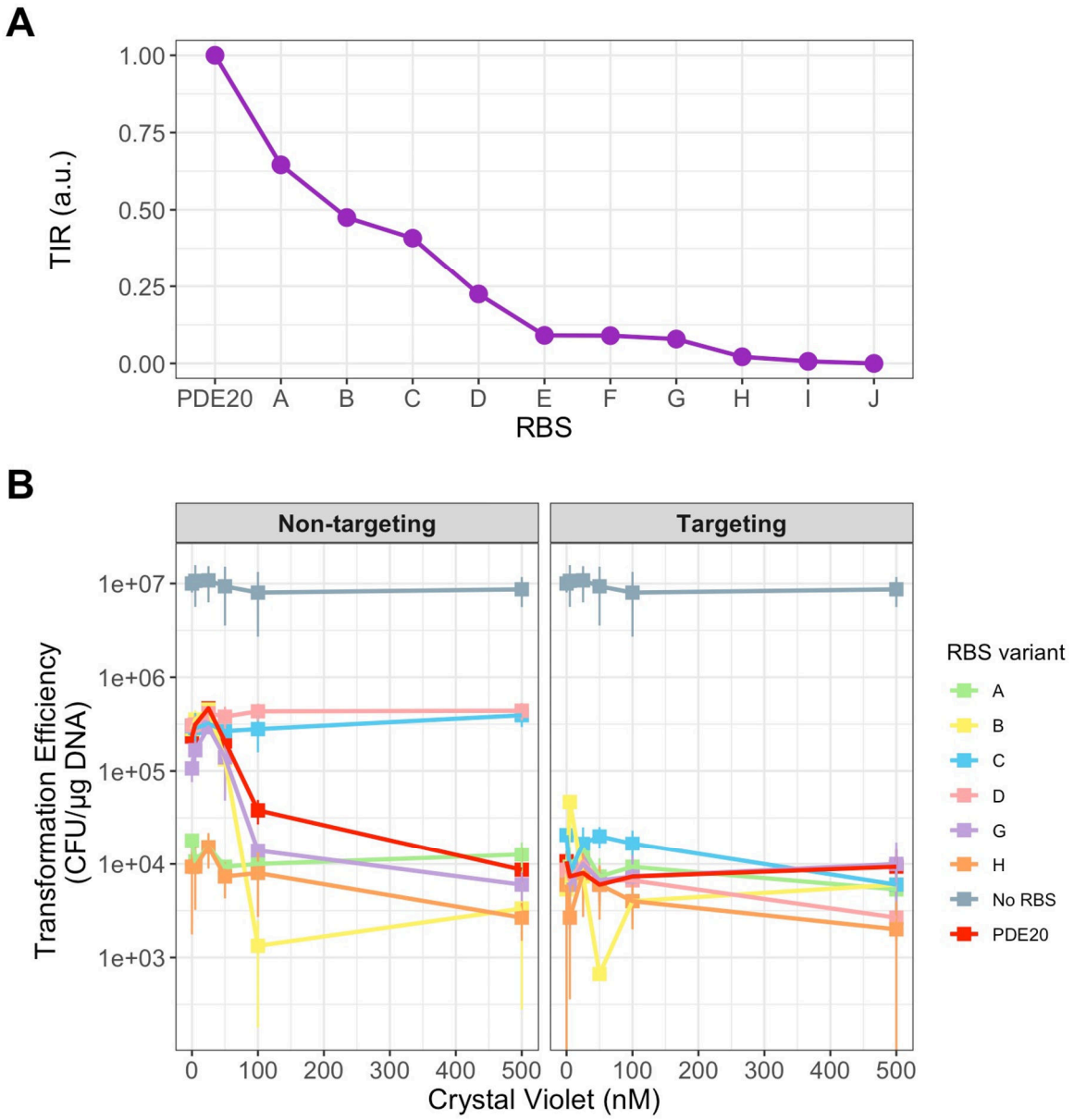


Figure 4. 4. Modulating RBS strength in the JEx-Cas system.

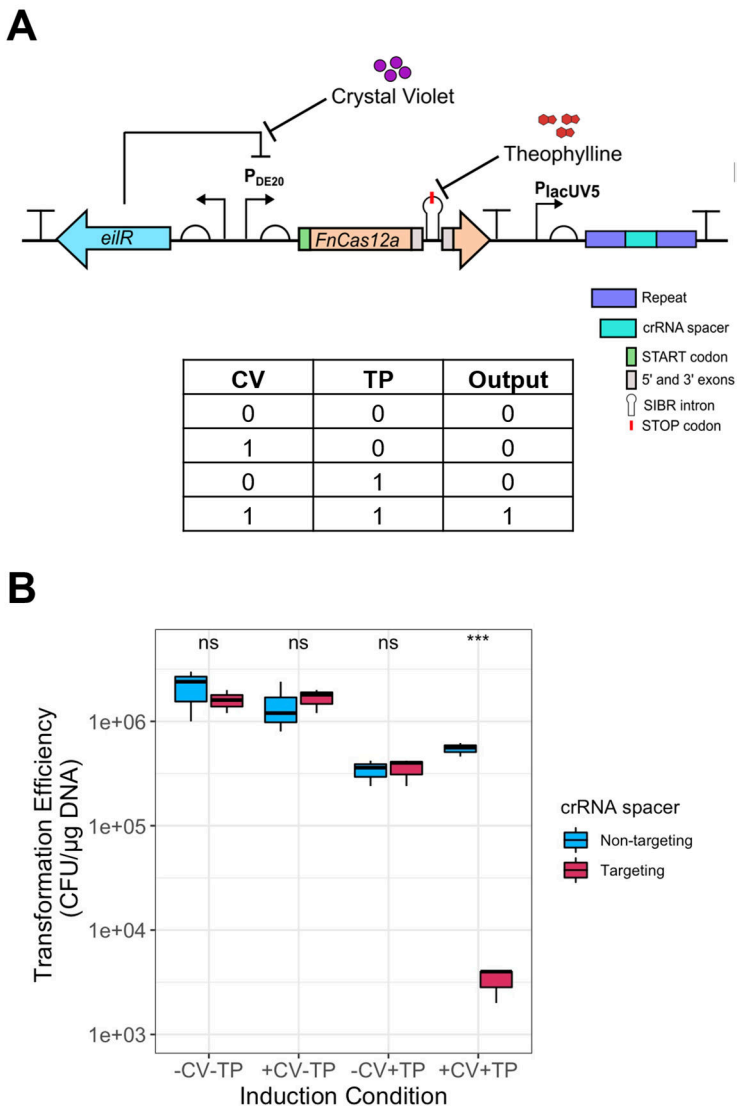
(A) Predicted maximal translation initiation rate (TIR) for all RBS variants, normalised to the strength of the RBS following the P_{DE20} in pJEx00S (labelled as “PDE20”, with a TIR of 65800 arbitrary units). **(B)** Transformation efficiency for pJEx00S (red, labelled as “PDE20”) and its RBS library derivatives, as quantified over an induction gradient (0-500 nM crystal violet). Transformation efficiency for the empty backbone plasmid pSEVA231 is shown in grey (labelled as "No RBS"). Individual points represent the mean of *n* = 3, ± one standard deviation.

4.4.5. A SIBR2.0-based AND gate enables tight control of *Cas12a* expression

The data reported above indicate that further optimisation of the JEx-Cas system is required to achieve tight transcriptional control of *Cas12a* expression and reduce the cytotoxicity of targeting pJEx00S plasmids. To do so, we opted to implement additional layers of gene expression control by combining JEx-Cas with the SIBR2.0 intron detailed in Chapter 3 (once it became available). The resulting plasmid (pJEx00N) encodes a genetic AND logic gate, regulating gene expression at both the transcription and translation level (Figure 4.5 A). When crystal violet is added, EilR-mediated repression of transcription from the P_{DE20} promoter is relieved, enabling the production of pre-mRNA transcripts that contain the SIBR2.0 intron sequence within the *Cas12a* CDS. If no theophylline is added, pre-mRNA transcripts will not be further processed, and translation of these transcripts will terminate once the ribosome encounters the stop codon sequence within the SIBR2.0 intron. Only when theophylline is also added to induce self-splicing of the SIBR2.0 intron can full-length, functional Cas12a be translated from the processed mRNA transcripts. The addition of either inducer in isolation will therefore not result in the expression of a functional gene product.

To test the performance of the AND gate experimentally, non-targeting and *acoC*-targeting versions of the pJEx00N plasmid were constructed and transformed into *C. necator* for inducible targeting assays. All pairwise combinations of the inducers were tested (0 or 50nM crystal violet, combined with 0 or 5mM theophylline). A significant difference in transformation efficiency between the non-targeting and targeting plasmids (~10^2 reduction) was only observed when both inducers were present, indicating successful implementation of the genetic AND gate (Figure 4.5 B). We note

that, for both conditions where theophylline was added, transformation efficiency was lowered for all plasmids relative to crystal violet-only conditions, suggesting that though AND gate induction is tight, theophylline itself may be slightly cytotoxic.



CV	TP	Output
0	0	0
1	0	0
0	1	0
1	1	1

Figure 4. 5. A genetic AND logic gate in *C. necator*.

(A) Architecture and truth table of the JEx- and SIBR2.0-based AND gate encoded on plasmid pJEx00N. CV = crystal violet, TP = theophylline. **(B)** Transformation efficiency for pJEx00N constructs, recorded over all induction conditions, as specified: -CV = 0nM crystal violet, +CV = 50nM crystal violet, -TP = 0mM theophylline, +TP = 5mM theophylline. ns = not significant; *** = p<0.01, two-sample t-test.

4.5. Discussion and Conclusions

The work presented in this Chapter comprises a thorough characterisation of the JEx system in *C. necator*. Using a fluorescent reporter system, we describe its induction kinetics in this specific genomic context, providing additional evidence for its broad applicability in diverse proteobacterial hosts. By applying JEx to the expression of *Cas12a* as a test system, we demonstrate how its gene expression output can be tuned by adjusting the inducer concentration, or by replacing individual DNA parts within the regulatory module.

JEx has been previously reported to be suitable for the expression of toxic genes in *E. coli* due to its low basal transcription rate[230]. Namely, the system was applied to inducibly express the *sacB* gene, which is commonly used as a counterselection marker due to its conditional lethality in the presence of sucrose. *E. coli* strains carrying JEx-*sacB* constructs were shown to grow normally in the presence of sucrose when no inducer was added to the growth medium, consistent with a minimal expression rate in uninduced conditions.

In our study, we similarly demonstrate that JEx operates with low basal transcription rate. In uninduced conditions, basal RFP expression was undetectable, in line with the data previously reported for *E. coli* (Figures 4.1 and S4.1)[230]. However, we found that the basal expression rate of the JEx system in *C. necator* is high enough to preclude the expression of cytotoxic genes (namely *Cas12a*, as shown in Figure 4.3). We overcame this limitation by implementing additional layers of gene regulation, combining JEx with the SIBR2.0 intron described in Chapter 3. In doing so, we

demonstrate the versatility of SIBR2.0, showing that it can be implemented to modulate (tighten) the gene expression output of other regulatory modules. The resulting regulatory architecture, a genetic AND logic gate, is generalisable, and could in theory be applied to tightly control the expression of any gene(s) of interest.

Though we have previously shown that SIBR2.0 alone is sufficient to enable tight control of gene expression, intron splicing can at present only be controlled by theophylline. Certain biotechnological applications may require sensing different inputs by implementing regulatory modules which are responsive to particular inducers. Modifying the SIBR2.0 system for use with a different inducer may not be possible, and at the very least would require re-engineering the entire intron sequence and re-optimising inducible self-splicing. Instead, implementing SIBR2.0 within an AND gate regulatory module ensures tight gene regulation whilst allowing for flexibility in the nature of the input. This regulatory architecture, which has not been previously demonstrated in *C. necator*, enables the implementation of more complex response functions. For instance, instead of using the JEx repressor and promoter, the SIBR2.0 intron could be coupled to an intracellular biosensing module, enabling condition-dependent genome editing (if the AND gate is used to control a CRISPR-Cas counterselection module) or modulation of gene expression (if the AND gate is used, for instance, to control a CRISPRi or CRISPRa module).

The inability to implement intricate genetic circuitry was recently identified as a limitation to advanced synthetic biology and metabolic engineering in *C. necator*[133]. By characterising the JEx inducible promoter system and the SIBR2.0 intron (Chapter 3), and demonstrating that these regulatory modules can be combined to construct a

genetic AND gate, we have successfully expanded the molecular toolkit available for this organism, supporting its development towards more complex biotechnological applications.

4.6. Chapter 4 Appendix – Supplementary Information

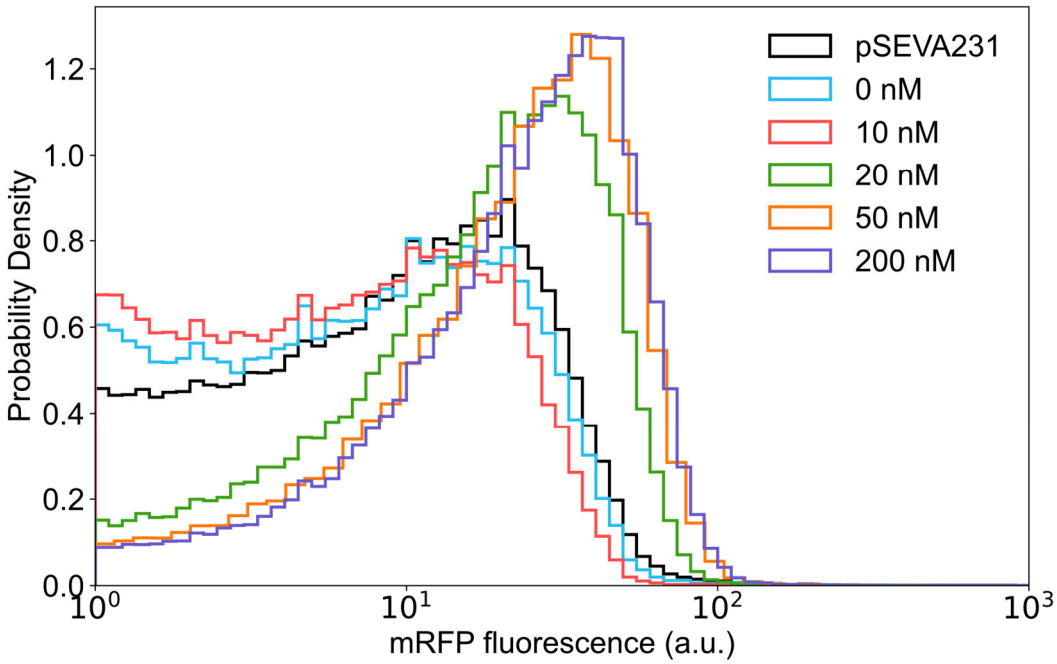


Figure S4. 1. JEx induction kinetics.

Individual histograms correspond to representative populations of cells carrying plasmid pJC580 at each inducer concentration. Plasmid pSEVA231, which does not direct *mRFP* expression, was used as a negative control.

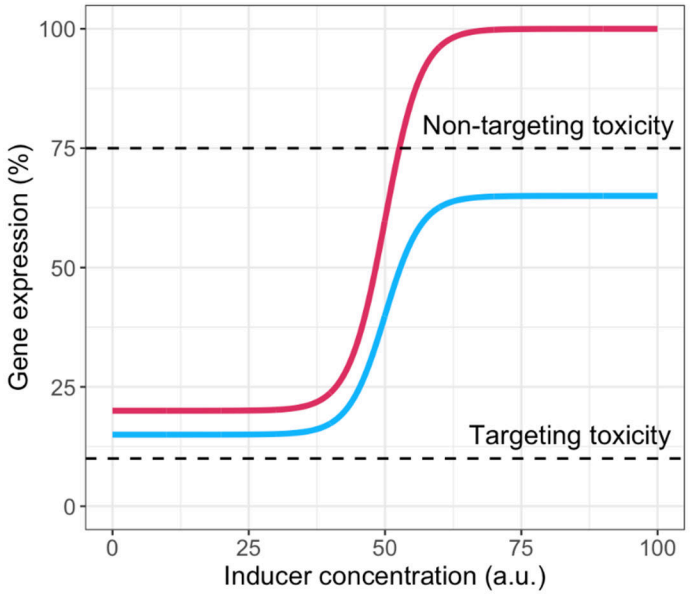


Figure S4. 2. Modulating induction kinetics in JEx-Cas.

The red sigmoidal curve corresponds to a stylised version of the crystal violet dose-response (induction) curve for the EilR-P_{DE20} transcriptional regulation system on pJEx00S. When no inducer is present, basal gene expression output is above the threshold at which targeting pJEx00S constructs are toxic. At high inducer concentrations, maximal gene expression surpasses the threshold at which non-targeting pJEx00S constructs are also toxic, likely due to off-target cleavage of genomic DNA by Cas12a. By reducing the strength of the RBS preceding the *Cas12a* CDS, we built pJEx00S variants with modified induction kinetics. Some of these variants, namely pJEx00C and pJEx00D, had a maximal gene expression output below the non-targeting toxicity threshold (blue curve). Though their basal gene expression output in uninduced conditions may have been reduced relative to pJEx00S, it remained above the targeting toxicity level.

Table S4. 1. Bacterial strains used in this study.

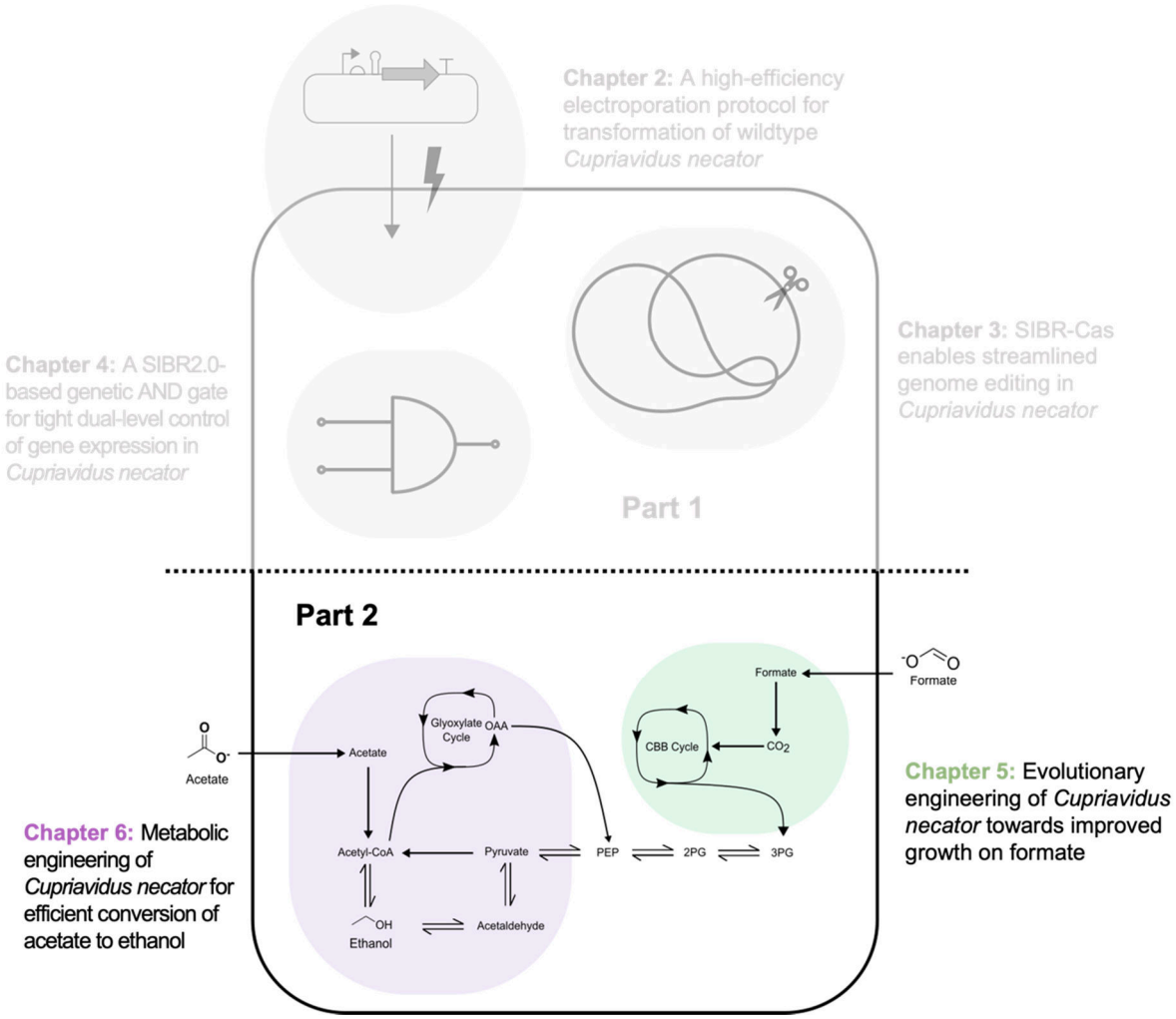
Strain	Purpose	Genotype
E. coli DH5α	Plasmid DNA storage and isolation	F– φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(rK⁻, mK⁺) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1
*C. necator*Δ0006Δ*phaC*	Inducible targeting assays.	Type I restriction modification system mutant (KEGG locus ID H16_A0006); *phaC* deletion mutant (KEGG locus ID H16_A1437).

Table S4. 2. Plasmids used in this study.

*Plasmid maps and sequences can be accessed via hyperlink. Sanger sequencing results are available in the linked files.

Plasmid*	Description	Ref.
pSEVA231	pBBR1 OriV-Rep; MCS; Kan^R	177,178
pJC580	RK2; ColE1; *eilR*; P_{DE20}; *mRFP*; Kan^R	230
pJEx00S-NT	pBBR1 OriV-Rep; *eilR;* P_{DE20}*; Cas12*; Kan^R; PlacUV5; crRNA cassette with *C. necator*-compatible non-targeting spacer.	This study
pJEx00S-acoC-T2	pBBR1 OriV-Rep; *eilR;* P_{DE20}*Cas12*; Kan^R; PlacUV5; crRNA cassette with *acoC*-targeting spacer T2.	This study
pJEx00A-NT	pJEx00S-NT with RBS variant A following P_{DE20}	This study
pJEx00A-acoC-T2	pJEx00S-acoC-T2 with RBS variant A following P_{DE20}.	This study
pJEx00B-NT	pJEx00S-NT with RBS variant B following P_{DE20}.	This study
pJEx00B-acoC-T2	pJEx00S-acoC-T2 with RBS variant B following P_{DE20}	This study
pJEx00C-NT	pJEx00S-NT with RBS variant C following P_{DE20}.	This study
pJEx00C-acoC-T2	pJEx00S-acoC-T2 with RBS variant C following P_{DE20}.	This study
pJEx00D-NT	pJEx00S-NT with RBS variant D following P_{DE20}.	This study
pJEx00D-acoC-T2	pJEx00S-acoC-T2 with RBS variant D following P_{DE20}	This study
pJEx00G-NT	pJEx00S-NT with RBS variant G following P_{DE20}.	This study
pJEx00G-acoC-T2	pJEx00S-acoC-T2 with RBS variant G following P_{DE20}.	This study
pJEx00H-NT	pJEx00S-NT with RBS variant H following P_{DE20}.	This study
pJEx00H-acoC-T2	pJEx00S-acoC-T2 with RBS variant H following P_{DE20}	This study
pJEx00N-NT	pJEx00S-NT variant where the SIBR2.0 intron has been placed at amino acid position 818 in the *Cas12a* CDS.	This study
pJEx00N-acoC-T2	pJEx00S-acoC-T2 variant where the SIBR2.0 intron has been placed at amino acid position 818 in the *Cas12a* CDS.	This study

Part 2:
Improving growth and production from C_1 and C_2 feedstocks



5. Evolutionary engineering of *Cupriavidus necator* towards improved growth on formate

5.1. Aims and Objectives

The aim of the work presented in this Chapter was to engineer *C. necator* strains with improved growth on formate via adaptive laboratory evolution. In doing so, we sought to demonstrate the functionality of an integrated evolutionary platform implemented within Chi.Bio turbidostats, which could theoretically be applied to evolve any growth-related traits in *C. necator* and other microorganisms amenable to continuous cultivation.

5.2. Introduction

One-carbon (C_1) compounds are increasingly gaining traction as renewable feedstocks for industrial biotechnology[14,106]. Amongst these, formate has emerged as a promising substrate for the production of a wide range of chemicals, holding demonstrable potential to drive a transition towards more sustainable bioprocesses[234]. Formate can be produced by electrochemical (abiotic) reduction of CO_2, a process operating with low energy input requirements and remarkable Faradaic efficiency[235–238]. In microbial fermentation, this substrate can act as an effective electron mediator, providing a source of both energy and carbon. Formate-based bioprocesses deliver many advantages over gas fermentation on CO_2/H_2 or syngas mixtures. These include

improvements in safety, substrate solubility, ease of handling and transport, simplified bioreactor designs, and the potential to support higher biomass and product yields, all of which can reduce the complexity and operating costs of bioprocesses[107,234].

An outstanding limitation for implementing energetically and economically feasible formate-based bioprocesses is the scarce availability of specialised microbes which can be used as production hosts[106].The facultative chemolithoautotroph *C. necator* is one of only a few proficient formatotrophic organisms which could be applied in C_1 bioprocesses, and optimising its growth kinetics on formate would be beneficial to realise its potential for sustainable bioproduction at an industrial scale. *C. necator* strains with higher growth rates could deliver increased volumetric productivities (g/l/h), and strains exhibiting improved formate assimilation into biomass and desirable metabolites are likely to increase production yields in C_1 bioprocesses.

A recent proteomic study of *C. necator* grown on fructose, succinate and formate provided quantitative evidence for the versatile nature of its metabolism[239]. However, it was found that, in any given growth condition, a large portion of the *C. necator* proteome is allocated towards metabolic functions which are not utilised. For instance, expression of genes within the *cbb* operon, which are involved in CO_2 fixation, was found to be high even during growth on fructose. Its Rubisco enzyme accounted for 3% of the proteome (by mass) in these conditions, without conferring any measurable fitness advantage through CO_2-reassimilation. Considering these findings, the bacterium can be described as a metabolic bet-hedger, favouring investments of metabolic resources towards readiness for changing environmental conditions, rather than more frugal expenditure of cellular machinery on any specific growth mode. As

C. necator is a soil bacterium, this metabolic strategy is likely to confer a selective advantage in its natural environment, where the precise composition and abundance of different feedstocks is likely to be in constant flux. However, this strategy is maladaptive in laboratory- or industrial-scale cultivation, where host organisms are required to grow optimally on a single substrate in an unchanging environment.

These quantitative observations demonstrate that there is ample room for optimisation of *C. necator*'s growth on formate. Broadly, there are two distinct strategies towards improving the growth of a bacterium on a specific carbon substrate. Rational engineering achieves this by targeted rewiring or modulation of metabolic pathways. For instance, previous studies in *C. necator* have improved autotrophic growth by overexpressing a heterologous Rubisco enzyme[163], or by replacing the bacterium's endogenous pathway for carbon fixation (the CBB cycle) with the reductive glycine pathway, a more efficient synthetic alternative[111,112].

As an alternative, adaptive laboratory evolution (ALE) can deliver global metabolic optimisation by replicating the process of genetic diversification and natural selection. Bacterial populations can be cultured for many generations in a defined selective environment, where mutations conferring a fitness advantage can rapidly arise and fix. Strains with desirable phenotypes can then be isolated from the evolving population[91,240]. Phenotypic and genotypic characterisation of these evolved isolates can provide insights into beneficial mutations, as well as global transcriptomic and proteomic changes underlying the observed traits. As a systems-level tool, ALE can optimise parameters which are inaccessible by rational design alone, particularly when knowledge of the host's metabolism is limited. Rational and evolutionary engineering

are often synergistically combined. For instance, ALE can be implemented to improve the *in vivo* performance of rationally-engineered metabolic modules. ALE can also provide insights into a microbe's global metabolic networks, which can then be retro-engineered into the host for more targeted strain optimisation, or extrapolated to other biotechnological applications.

In a study contemporaneous to the work presented in this Chapter, Calvey *et al.* successfully isolated *C. necator* strains with improved growth on formate using ALE[92]. Implementing a serial batch culture method over a total of 400 generations, several mutant strains with improved formatotrophic growth kinetics were isolated. Common mutations underlying the superior growth phenotype were identified and re-introduced in the WT ancestor, yielding engineered strains with a 24% faster maximum growth rate on this C_1 substrate. Here, we describe a parallel ALE experiment, evolving *C. necator* towards improved growth on formate using an alternative evolutionary platform. Namely, we conducted the evolutionary experiment within Chi.Bio turbidostats[183], which enabled continuous cultivation of *C. necator* on formate over ~150 generations (25 days of continuous exponential growth). First, we calibrated the turbidostat-based evolutionary platform, identifying effective parameters for continuous formatotrophic cultivation and *in situ* UV mutagenesis. When performing the ALE experiment, we evolved three parallel lineages, and conducted periodic phenotypic characterisation of individual isolates from each lineage along the evolutionary trajectory. We provide a detailed analysis of the growth kinetics of the best evolved isolates at both permissive and restrictive formate concentrations. By conducting genomic analysis of the most proficient formatotrophs isolated in the ALE

experiment, we revealed common mutations underlying the improved growth phenotypes, which could be harnessed for strain engineering.

5.3. Materials and Methods

Bacterial strains and culture conditions

Wild-type *C. necator* H16 was used as the ancestor for the evolution experiment. *E. coli* DH5α was used as a reference strain for the nalidixic acid toxicity assay as described below. Rich growth medium was LB. Minimal medium was J minimal medium (JMM), previously optimised for formatotrophic cultivation of *C. necator*[154]. JMM was supplemented with 80 mM or 160 mM sodium formate, or 20 mM fructose, as indicated, and the pH of the medium was adjusted to 7.2. Plates for bacterial growth were LB agar or JMM agar with 80 mM sodium formate, as appropriate. Routine cultivation was performed in a total volume of 5 ml, in 50 ml conical centrifuge tubes, incubated at 30 ºC with 150 rpm shaking (*C. necator*) or 37 ºC with 200 rpm shaking (*E. coli*).

Batch- and turbidostat-mode formatotrophic cultivation of *C. necator* in Chi.Bio reactors

To inoculate cultures within Chi.Bio devices, cells were first revived from glycerol stocks onto LB agar plates and incubated for 48 h at 30 ºC. Individual colonies were then picked and cultured overnight (16-18 h) in LB medium. A small volume (≤200 µl) of these saturated LB cultures was then used to inoculate a second round of overnight

cultures in JMM with 20 mM fructose as sole carbon source, to pre-adapt cells to growth in minimal medium. Cells from these minimal medium precultures were then used to inoculate cultures in sterile 30 ml flat-bottom glass vials, to be placed within Chi.Bio devices. Batch-mode cultures of *C. necator* in formate medium were inoculated at a starting optical density (OD) of 0.05-0.075 in a total volume of 20 ml, and incubated at 30 ºC with continuous magnetic stirring (stir rate set-point = 0.8). Turbidostat-mode cultures were conducted analogously, with the addition of a target OD set-point and programmed dithering, as specified. OD measurements were recorded every minute using the devices' built-in 650 nm laser. Recommended hardware and consumables were used for all Chi.Bio components (see ref. [183] and https://chi.bio/hardware/ for details).

Nalidixic acid toxicity assay

Nalidixic toxicity in *E.* coli and *C. necator* was quantified via a minimum inhibitory concentration assay at the microplate scale. Bacterial strains were first revived from glycerol stocks onto solid medium (LB agar) and plates were incubated at 30 ºC for 48 h (*C. necator*) or 37 ºC for 16 h (*E. coli*). Individual colonies were then picked and used to inoculate overnight (16-18 h) pre-cultures in rich medium. Following incubation, cells from overnight cultures were pelleted by centrifugation. Spent medium was discarded, and the cell pellet was resuspended in fresh LB medium, adjusting the cell density to an OD of 1. Cells were used to inoculate fresh cultures in a 96-well microplate by diluting them 1:10 into the plate wells, giving a starting OD of 0.1. The total volume of each well was 200 µl. Nalidixic acid was added to microplate wells by diluting a stock solution to give working concentrations in the range of 0-400 µg/ml, as indicated. The

microplate was incubated at 30 °C with double orbital shaking in a Spark microplate reader (Tecan), with OD measurements taken automatically at 15min intervals over a period of 24 h.

Measurement of UV-mediated genetic diversification

To quantify UV-mediated genetic diversification, turbidostat cultures were inoculated as described above. The bacterial population within the turbidostats was cultured at the target OD set-point (0.25±0.025) for ~10 h, after which UV irradiation was performed *in situ* using the Chi.Bio 280 nm LED set to different intensities (0, 0.005 or 0.01, arbitrary units) in parallel reactors, for a total of 8 h. Turbidostat cultures were sampled periodically following irradiation, specifically after 0, 12, 24 and 42 h. The culture sample was pelleted by centrifugation, most of the supernatant was discarded, and the cell pellet was resuspended in the residual supernatant (~100 µl) and spread onto LB agar plates supplemented with nalidixic acid at a working concentration of 80 µg/ml. Plates were incubated at 30 ºC for 48 h prior to colony counting.

Adaptive laboratory evolution in Chi.Bio turbidostats

Turbidostat cultures were inoculated as described above in three separate Chi.Bio reactors (M0, M1, M3), operated in parallel. The bacterial population within each turbidostat was cultured at the target OD set-point (0.25±0.025) for a total of ~150 generations (~600 h, or 25 days). Where indicated, UV mutagenesis was applied by irradiating the cultures with the built-in UV LED set to an intensity of 0.005 (chosen based on UV dosage calibration methodology above) for a fixed period of 8 h. Inflowing

minimal medium was JMM with 80 mM sodium formate as sole carbon source. The turbidostats were operated at ambient CO_2 and O_2 concentrations. The pH of the medium was not adjusted during cultivation and was measured in the range of 7.2-7.4 for turbidostat cultures at the target OD set-point (sterile JMM pH had been adjusted to 7.2, as stated above). Where necessary, turbidostat cultures were re-started either by dilution to the target OD set-point or to the starting OD (0.05-0.075), depending on the cause and duration of the stoppage. In cases where biofilm formation was observed, the cultures were transferred to fresh sterilised glass vials. Cultures were manually sampled at the specified timepoints using a sterile syringe with a 12 cm needle, which could be threaded through the lids of the glass vials in the reactors, such that the experiment did not have to be stopped for sampling. A total of 3 ml were extracted from each turbidostat at each sampling time-point. The liquid samples were pelleted by centrifugation, and the supernatant was discarded. Cell pellets were resuspended in residual supernatant and spread onto JMM agar plates, with 80 mM sodium formate as sole carbon source.

Microplate-scale formate growth assays

For each reactor and at each sampling time-point, ten individual colonies were picked from JMM + formate agar plates and used to inoculate precultures in JMM medium supplemented with 20 mM fructose. The ancestral WT *C. necator* strain was revived from a frozen glycerol stock and precultured in the same conditions. Cells from these overnight precultures were then used to inoculate formatotrophic growth assays in microplate format, as follows: cells were pelleted by centrifugation, washed twice in PBS to remove residual fructose, and resuspended in fresh JMM medium with 80 mM

or 160 mM sodium formate as sole carbon source, as indicated. The density of these cell stocks was adjusted to an OD of 0.5. Cells were then diluted 1:10 into microplate wells to inoculate fresh cultures at a starting OD of 0.05 in a total volume of 180 µl. Each microplate well was topped with 20 µl of clear mineral oil to prevent evaporation, as described previously[111]. The microplate was then incubated at 30 °C with double orbital shaking in a Spark microplate reader (Tecan), at ambient CO_2 and O_2 concentrations, with OD measurements taken automatically at 15 min intervals over a period of at least 48 h. Data analysis was performed using custom R scripts, implementing a previously described algorithm for growth rate estimation[186].

Whole-genome sequencing and analysis

Genomic DNA extraction, DNA library preparation, Illumina sequencing (2x250bp reads, 60x depth of coverage), and preliminary bioinformatic analyses (quality control, Illumina read trimming) were performed by MicrobesNG (United Kingdom). Downstream analysis of sequencing data was performed using breseq[241] version 0.38.1. Briefly, trimmed reads provided by MicrobesNG were aligned to the reference genome sequence to identify possible mutations, using the standard breseq configuration in consensus mode (frequency cutoff = 70%). NCBI accession numbers for the reference genome files were NC_008313.1 (Chromosome 1), NC_008314.1 (Chromosome 2), and NC_005241.1 (pHG1 megaplasmid)[137]. Comparative analysis of the resulting GenomeDiff-formatted files (containing mutations identified for each evolved isolate) was performed using the gdtools utility program within the breseq package. Mutations present in the ancestral wildtype strain were subtracted from the GenomeDiff files of evolved isolates prior to this analysis.

5.4. Results

5.4.1. Overview of the evolutionary platform

Continuous cultivation in Chi.Bio reactors[183] was chosen as the basis of the evolutionary platform (Figure 5.1). The reactors were operated in turbidostat mode, such that a bacterial population of fixed size could be grown continuously in the exponential phase at a high growth rate, thereby maximising the number of cell divisions. This has been proposed to be an important parameter in ALE experiments, particularly when evolving cell populations towards improved growth rate phenotypes[242]. Moreover, relative to serial batch culture the turbidostat platform reduces the complexity of the selective environment, ensuring that selection is directly applied towards the evolution of faster growth rates in the exponential phase. This limits the selective advantage of phenotypic traits that may emerge during other growth phases, such as a shortened lag phase, or a decreased likelihood of becoming quiescent in the stationary phase (i.e. when nutrients are depleted). Additionally, the absence of bottlenecks is one of the main advantages of continuous cultivation over serial batch transfer for ALE. In a turbidostat, all genotypes have an equal chance of being represented in the next generation. Their abundance along the evolutionary trajectory is then primarily determined by their associated fitness, and not by being randomly removed from the population during serial passaging. Note that random removal of new mutants from the evolving population is not completely eliminated, as this may still occur during dilution or manual sampling of the turbidostat cultures.

The turbidostats also offer the advantage of largely hands-off operation. Bacterial growth within the devices can be automatically monitored *in situ* and manipulated remotely. The experimenter's input is limited to ensuring that sufficient fresh (sterile) medium is available, manually sampling the reactors where necessary, and disposing of waste (spent medium). The minimal requirement for hands-on interventions is particularly advantageous for long-term evolution experiments, as it reduces both the likelihood of contamination and potential for human error.

Mutagenesis is routinely used in ALE experiments to increase the genetic diversity of the population and accelerate the rate of evolution[89]. The turbidostats offer the possibility of performing genome-wide, random mutagenesis via *in situ* ultra-violet (UV) irradiation, providing a streamlined protocol for genetic diversification of the evolving bacterial population.

Prior to the ALE experiment, we conducted preliminary analyses of the proposed evolutionary platform. These included validating the Chi.Bio reactors for both batch and continuous formatotrophic cultivation of *C. necator* (Section 5.4.2), as well as investigating how UV irradiation can be harnessed to induce genetic diversification in the bacterium (Section 5.4.3).

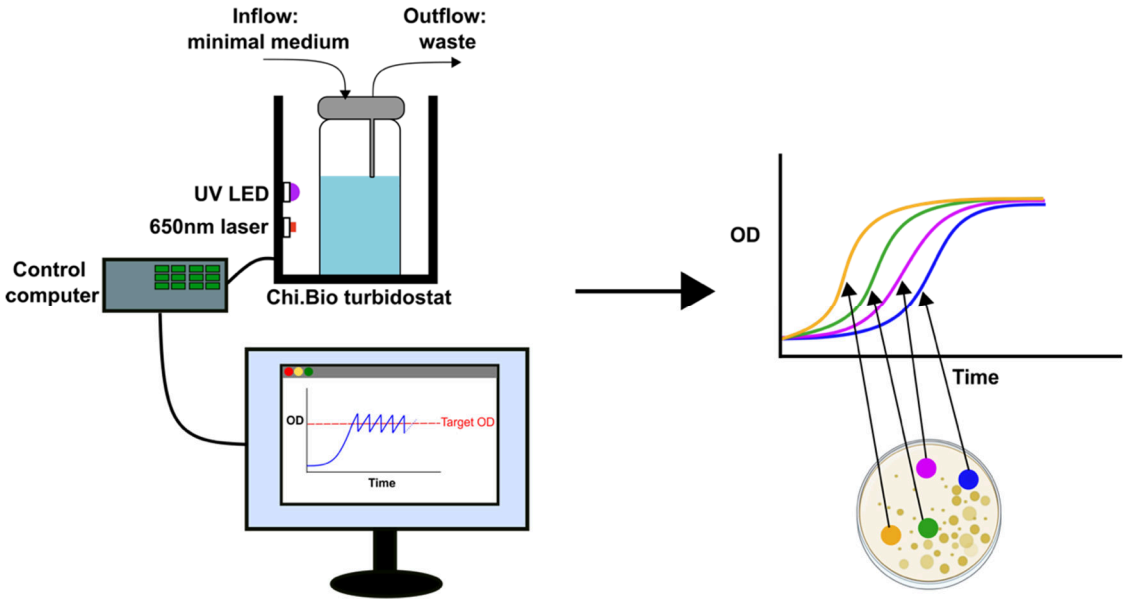


Figure 5. 1. Continuous evolution experiment.

C. necator is grown continuously within a Chi.Bio turbidostat. Here, the culture volume is kept constant by balancing the inflow of formate-containing minimal medium and the outflow of waste. The bacterial population is kept in the exponential growth phase by controlling the optical density (OD) within the bioreactor, allowing for small fluctuations around the target OD set point to enable growth rate estimation. OD is measured every minute using a 650 nm laser. Mutagenesis is performed *in situ* using a built-in UV LED. The bioreactor culture is sampled periodically. Individual clones are screened using high-throughput formatotrophic growth assays in microplate format, to identify evolved strains with improved growth rate and formate assimilation.

5.4.2. Continuous formatotrophic cultivation of *C. necator* in Chi.Bio turbidostats

First, we characterised the formate growth kinetics for *C. necator* in Chi.Bio reactors to determine suitable parameters (chiefly target OD) for continuous cultivation. We performed all experiments using 80 mM sodium formate as sole carbon source in the growth medium, at ambient CO_2. This formate concentration is routinely used for cultivation of *C. necator*, providing sufficient carbon for growth whilst remaining

permissive, i.e. not resulting in restrictive growth inhibition. A thorough characterisation of formatotrophic substrate inhibition kinetics is provided in ref. [153].

We profiled the bacterium's growth in batch cultures within the reactors to determine a suitable target OD set-point for turbidostat cultivation. Growth and growth rate curves for a representative batch culture are provided in Figure 5.2 A and B, respectively. A maximal specific growth rate of ~0.21 h^{-1} was recorded in early exponential phase (OD ~0.15), a value in line with previous literature reports which have estimated maximal growth rates in the range of 0.17-0.23 h^{-1}, depending on the choice of growth scale and mode[153,243–246]. Harnessing these results, we selected a target OD of 0.25±0.025, corresponding to a growth rate of ~0.18 h^{-1}. Though populations at this density do not grow at the maximal rate, increasing the number of cells within the culture has two main advantages. Firstly, using a larger bacterial population allows for more genetic diversity to emerge over the course of the evolution experiment and is expected to yield a higher rate of phenotypic and genotypic adaptation, as well as more robust experimental outcomes[247–250]. Secondly, from a more practical perspective, *in situ* growth rate estimation on the Chi.Bio platform requires small fluctuations (dithering) of the cell density around the target OD. Even in the absence of programmed dithering, dilution by the turbidostats' peristaltic pumps is imprecise, making it difficult to maintain the culture at the exact target OD set-point. If the set-point was low (such as OD 0.1-0.15, where the growth rate was estimated to be maximal), cultures may be diluted too far due to dithering or imprecise pumping, or a combination thereof. This may return the bacterial cultures to the lag phase, and indeed any dilution would reduce the growth rate, negating the effects of continuous exponential phase culture. By raising the target OD set-point to 0.25±0.025, we minimise the likelihood of the population's

growth rate being drastically reduced due to dilution and ensure that turbidostat cultures remain in exponential phase throughout the evolution experiment. At this target OD, the population size was (8±3)x10^8 cells (as quantified by serial dilution on LB agar plates).

We demonstrated that continuous formatotrophic cultivation of *C. necator* in Chi.Bio turbidostats is feasible, thereby validating the continuous cultivation set-up for experimental evolution. Growth curves recorded in three independent Chi.Bio turbidostats are shown in Figure 5.2 C. The corresponding growth rate curves, estimated from programmed dithering around the target OD set-point, are shown in Figure 5.2 D. The average growth rate recorded for the turbidostat cultures after was 0.18±0.01 h^{-1} after 24 hours and 0.19±0.01 h^{-1} after 48 hours. These measurements demonstrate low inter-reactor variability, as well as low intra-reactor growth rate variability over time.

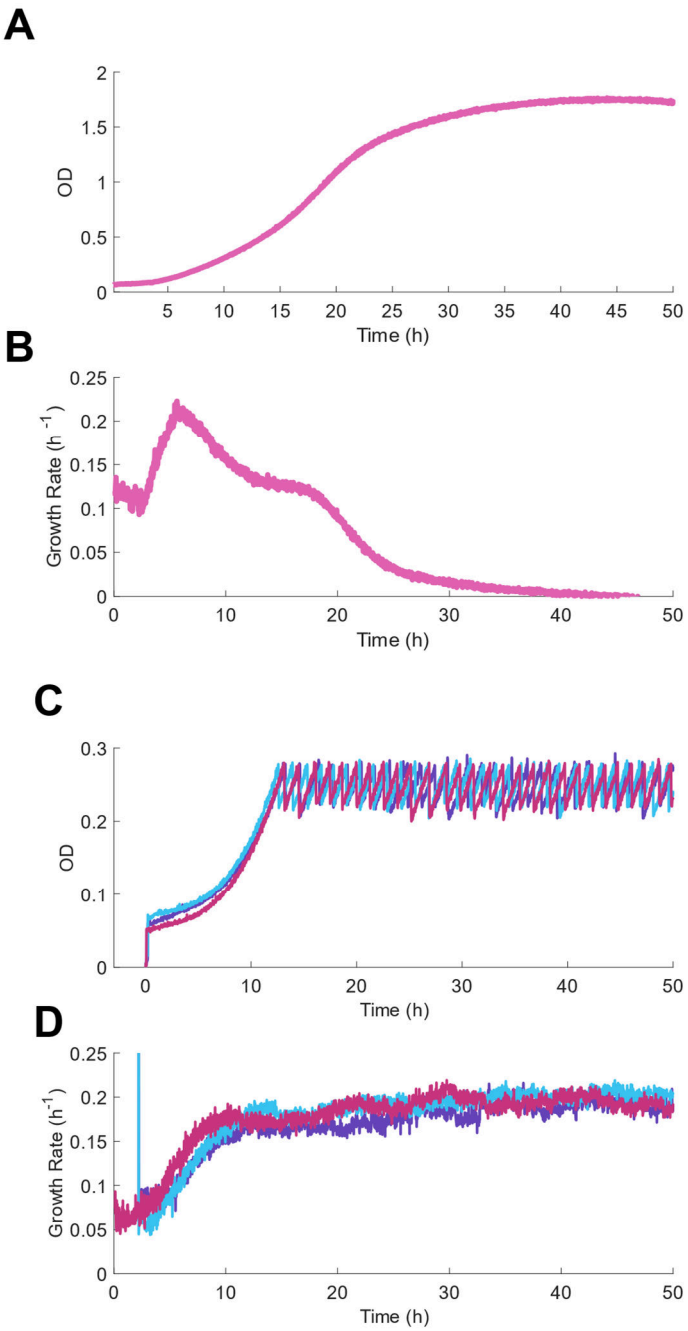


Figure 5. 2. Formatotrophic cultivation of *C. necator* in Chi.Bio reactors.

(A) OD and **(B)** growth rate of *C. necator* when cultured in batch mode within a Chi.Bio device. **(C)** Turbidostat-mode Chi.Bio cultures of *C. necator*. Each line represents the OD trajectory of an individual Chi.Bio device, where the target density of the bacterial cultures was set in mid-exponential phase (target OD 0.25 ± 0.025). **(D)** Growth rate in turbidostat cultures, as estimated from programmed dithering around the target OD.

5.4.3. UV-mediated genetic diversification

When a given physico-chemical mutagen, such as UV irradiation, is applied to a bacterial culture, there is a trade-off between genetic diversification and growth inhibition. It is important to fine-tune the mutagen dosage to balance these two effects and perform mutagenesis most effectively. Considering this, we investigated which UV dose should be applied to enable maximal genetic diversification of *C. necator* populations within the Chi.Bio turbidostats whilst minimising growth inhibition.

The turbidostats' built-in UV LED (280 nm) can be set to varying intensities, and activated for varying time periods. By modulating the intensity parameter, we exposed turbidostat cultures to different UV dosages and measured their effect on the bacteria's growth rate. We define a UV dose as the product of LED intensity and time of irradiation, where the latter was kept constant (eight hours, roughly corresponding to two doubling times). We characterised the UV dose-response by using eight different UV LED intensity set-points in the range of 0-0.1. We observed that UV intensities ≥ 0.005 caused a reversible, dose-dependent growth rate reduction, consistent with temporary growth inhibition (Figure 5.3). The recovery time, i.e. the time needed for the growth rate to return to its pre-irradiation value, was also dose-dependent. However, full recovery time for some larger UV doses fell outside of the timeline of the experiment and therefore could not be accurately quantified.

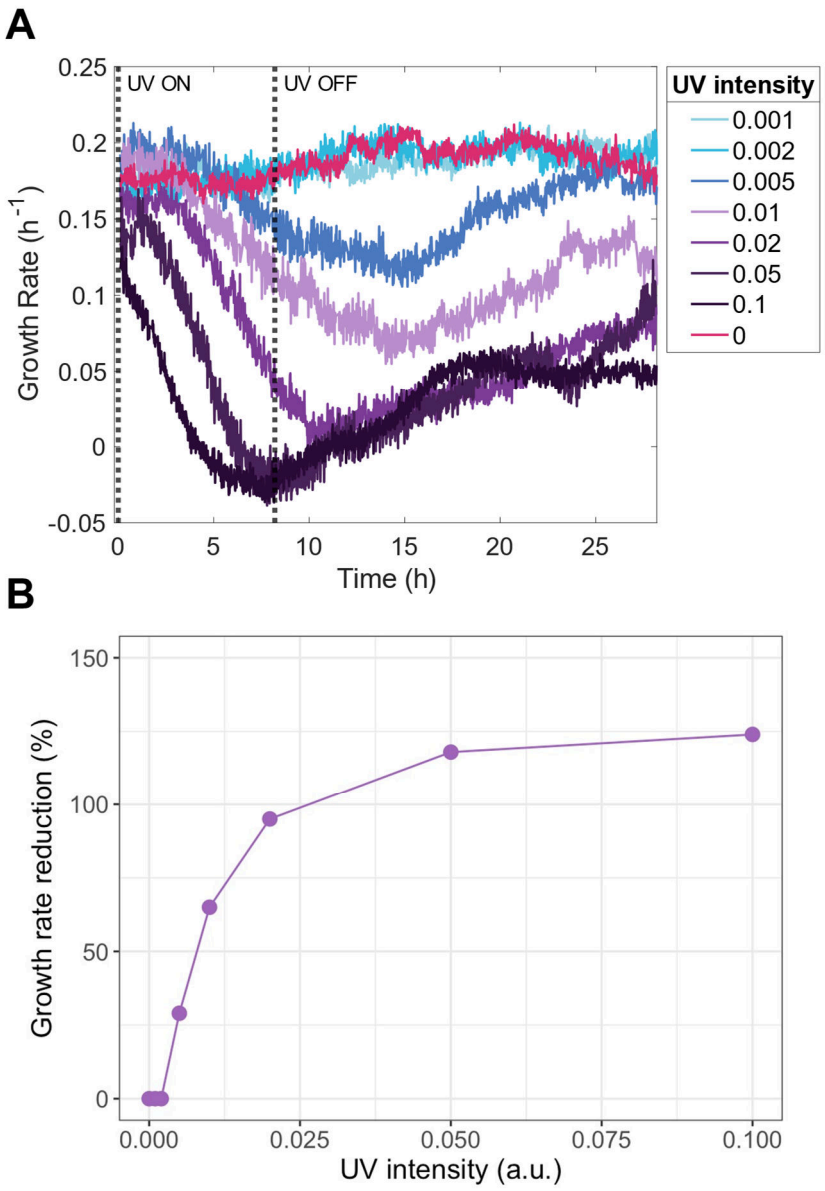


Figure 5. 3. UV dose-response in turbidostat cultures.

(A) Growth rate in turbidostat cultures during and following UV irradiation at varying LED intensities, as indicated. **(B)** Maximum reduction in growth rate recorded for each UV dose, expressed as a percentage of the growth rate at the time when UV irradiation began (Time = 0 h in panel A).

We assumed that UV doses which result in measurable growth rate reductions are likely to generate more genetic diversity than lower doses, which do not inhibit growth. Yet using a UV dosage which leads to drastic reductions in growth rate may be disadvantageous due to the resulting magnitude of cell death and the extended recovery time. In line with this reasoning, we selected only two UV doses for further

investigation, corresponding to LED intensities of 0.005 and 0.01. Specifically, we determined whether exposure to these UV doses would not only impact the growth rate, but also lead to more genetic diversity being generated relative to no-UV conditions. We used the spontaneous emergence of antibiotic-resistant clones following UV irradiation as a proxy measure for the population's genetic diversity. Nalidixic acid (NalX), a synthetic quinolone antibiotic, was chosen for this purpose, as it has been previously used for fluctuation studies in *E. coli*[251]. In Gram-negative bacteria, resistance to NalX is most often acquired by a single point mutation in the quinolone resistance-determining region of the *E. coli gyrA* gene, which encodes the A subunit of DNA gyrase[252,253]. First, we confirmed that this was also the case for the *C. necator gyrA* orthologue (Figure S5.1 A). We then characterised NalX toxicity in the bacterium via a minimum inhibitory concentration assay and determined it to be similar to its toxicity in *E. coli* (Figure S5.1 B). Hence, we used a NalX dose of 80 µg/ml to quantify the spontaneous emergence of antibiotic resistance in *C. necator* populations, as previously done for *E. coli*[251].

We applied three distinct UV doses (0, 0.005 and 0.01 UV LED intensity over eight hours) to independent turbidostat cultures. Following irradiation, we quantified the emergence of NalX-resistant clones over time by sampling the cultures periodically and spreading the cells on selective agar plates (Figure 5.4 B). We observed that both 0.005 and 0.01 UV intensities led to genetic diversification over time, whilst the no-UV control dose did not yield measurable benefit (Figure 5.4 A). The number of NalX-resistant mutants increased steadily following UV exposure, peaking ~24 hours after irradiation. This is consistent with mutations arising due to the action of DNA repair machinery, which requires significant time for activation and subsequent introduction

of mutations[254]. After 24 hours, we recorded a small drop in the population's genetic diversity (i.e. a reduction in the number of spontaneously-emerging NalX-resistant mutants), likely due to some NalX-resistant mutants being counterselected due to having accumulated deleterious mutations that conferred a growth disadvantage. This reduction in genetic diversity was expected; only clones with neutral or advantageous mutations are likely to be successful and remain in the population once the mutagen is removed. In general, cultures undergoing continuous growth at fixed population sizes will maintain some level of standing diversity dependent upon mutation rates and selective coefficients of accessible mutations, and as such we expect culture diversity to reduce over time following the artificial increase in mutation. Since we did not observe significant differences in population genetic diversity between the 0.005 and 0.01 UV doses, we chose to apply the former within the evolution platform, as its inhibitory effect on bacterial growth is smaller. Though quantified, the timing of the genetic diversity peak was not central to the evolution experiment. Nevertheless, it provides insights into the dynamics of the mutation process.

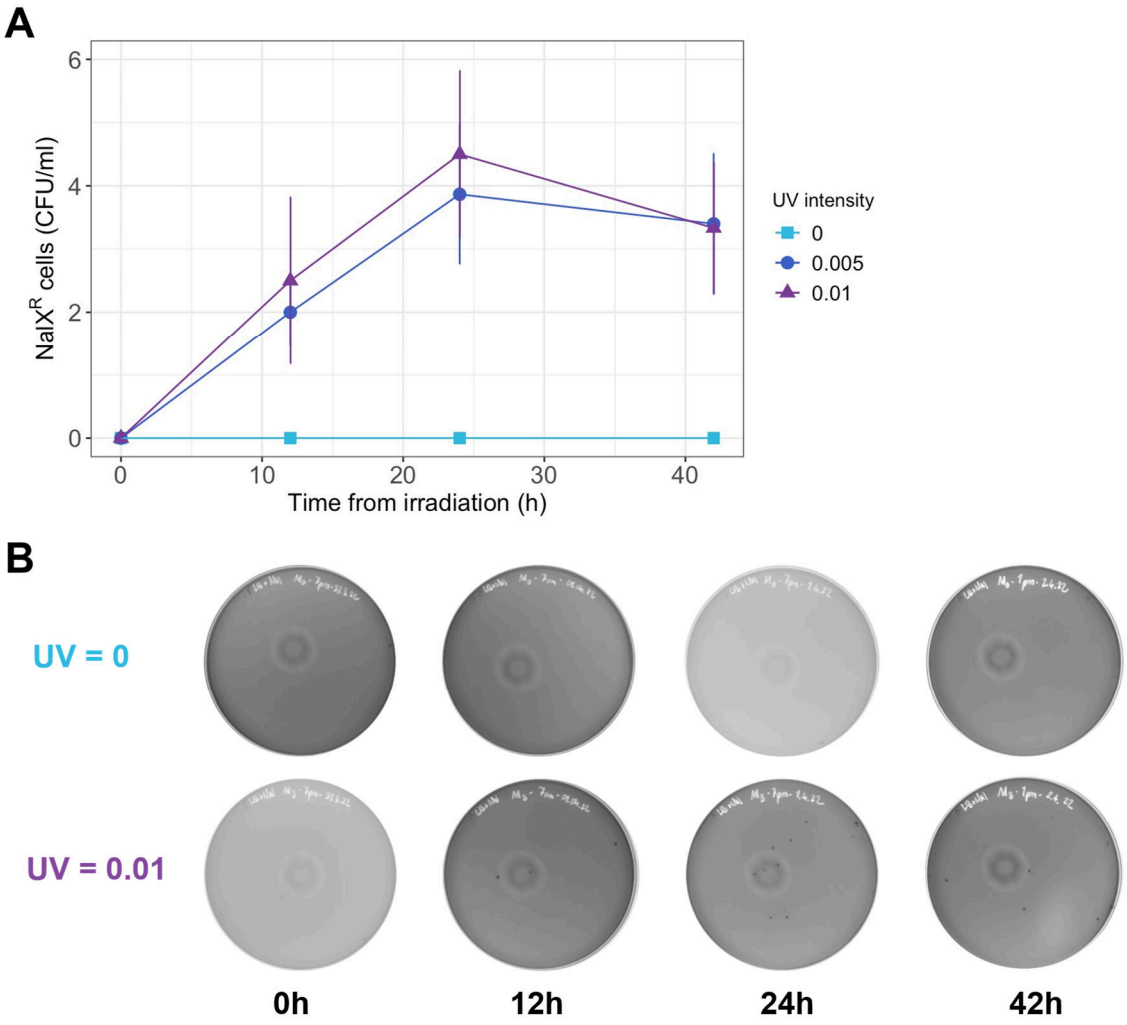


Figure 5. 4. UV-mediated genetic diversification.

(A) Emergence of NalX-resistance ($NalX^R$) in turbidostat cultures following UV exposure (time 0 corresponds to the time at which the UV LED was switched off). Individual points represent the mean of *n* = 3 ± one standard deviation. **(B)** Representative selective agar plates (LB agar + NalX) for two of the tested UV intensities at each time-point. Note that plates for a single replicate are shown for simplicity.

5.4.4. Formatotrophic growth kinetics of evolved isolates

To perform the evolutionary experiment, we inoculated three independent lineages from a single ancestral WT strain of *C. necator*. Specifically, each parallel lineage was grown to the target OD set-point in a separate Chi.Bio turbidostat (labelled as M0, M1,

or M3), then cultivated continuously in exponential phase for a total of ~150 generations. Note that, in this evolutionary context, we defined a generation as a single doubling time (~four hours as calculated using the ancestral strain's growth rate on 80 mM sodium formate in the Chi.Bio reactors). Further details on the ALE set-up are provided in Section 5.3.

Growth and growth rate curves measured *in situ* for all turbidostats are shown in Figure 5.5. UV mutagenesis was applied twice over the course of ALE, at time-points near the start and the midpoint of the experiment (~10 and ~80 generations). The turbidostat cultures had to be restarted several times along the ALE timeline, due to unforeseen events such as liquid spills or stirring issues. Where possible, dilution was avoided upon restart to keep the bacterial population in the exponential phase and not introduce random population bottlenecks which may influence evolutionary outcomes.

For all devices, we found that *in situ* growth rate estimates were subject to random fluctuations and may not accurately reflect real growth rate improvements. These fluctuations were at least partially due to experimental limitations, such as biofilm formation or spilt growth medium obfuscating the path of the 650 nm laser through the reactors' glass vials. Nevertheless, some patterns could be discerned in the growth rate trajectories for each turbidostat. For M0, the estimated growth rate remained broadly stable over time (Figure 5.5B), with some improvements recorded from ~80 generations onwards. For reactors M1 (Figure 5.5D) and M3 (Figure 5.5F), larger growth rates improvements were recorded, particularly from 100 generations onwards.

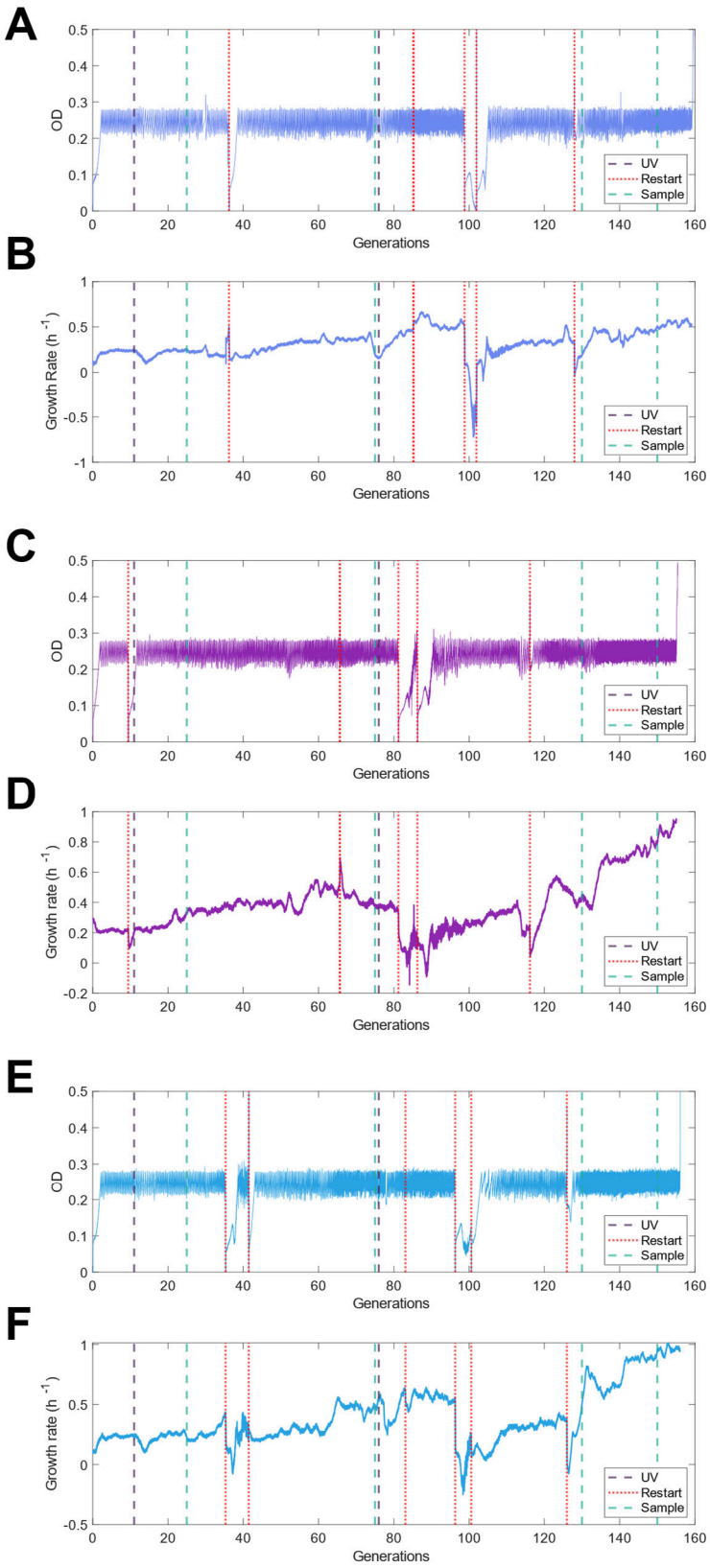


Figure 5. 5. Growth in the Chi.Bio turbidostats along the evolutionary timeline.

Optical density and growth rate curves for M0 (**A and B**), M1 (**C and D**), and M3 (**E and F**). Growth rate was estimated from programmed dithering around the target OD set-point (0.25±0.025). Experimental interventions (UV irradiation, turbidostat restarts, and sampling time-points) are indicated by vertical lines in all panels for ease of reference. Stoppage times (red dashed lines) are identical within the individual turbidostats, but not between them.

Manual sampling of the culture and characterisation of individual isolates would ideally be performed once improvements in the population growth rate are observed. Due to the aforementioned difficulties associated with *in situ* growth rate estimation, we sampled the turbidostat cultures periodically at arbitrary time-points, which were not entirely determined by observed growth rate improvements. Specifically, all three turbidostat cultures were manually sampled after 25, 75, 130 and 150 generations. The growth kinetics of individual isolates from each lineage were characterised at each time-point following an identical sampling protocol. Briefly, a small volume of each Chi.Bio culture was plated on solid minimal medium (JMM agar with 80 mM sodium formate as sole carbon source). Ten individual colonies were picked for each reactor, precultured, and then used to inoculate microplate cultures for formatotrophic growth assays. A WT (unevolved ancestor) control was identically pre-cultured and included in the microplate growth assay. Note that precultures were performed in fructose-containing minimal medium, with no added formate. This step was designed to bias against evolved isolates in the formatotrophic growth assays. We hypothesised that it would increase the likelihood that any observed growth improvements would be due to genotypic differences between evolved isolates and the WT, rather than transient phenotypic changes induced by cultivation in formate medium.

Kinetic growth parameters (growth rate, lag phase duration, and carrying capacity) measured for all reactors, time-points and isolates are shown in Figure 5.6. Interestingly, we observed that some of the evolved isolates had shortened lag phases and improved carrying capacities relative to the WT, neither of which were deliberately selected for in the ALE platform. Clones isolated from the M3 lineage after 150 generations exhibited the most promising growth advantages, with mean absolute

improvements of 14%, 21% and 59% for growth rate, lag phase duration and carrying capacity, respectively.

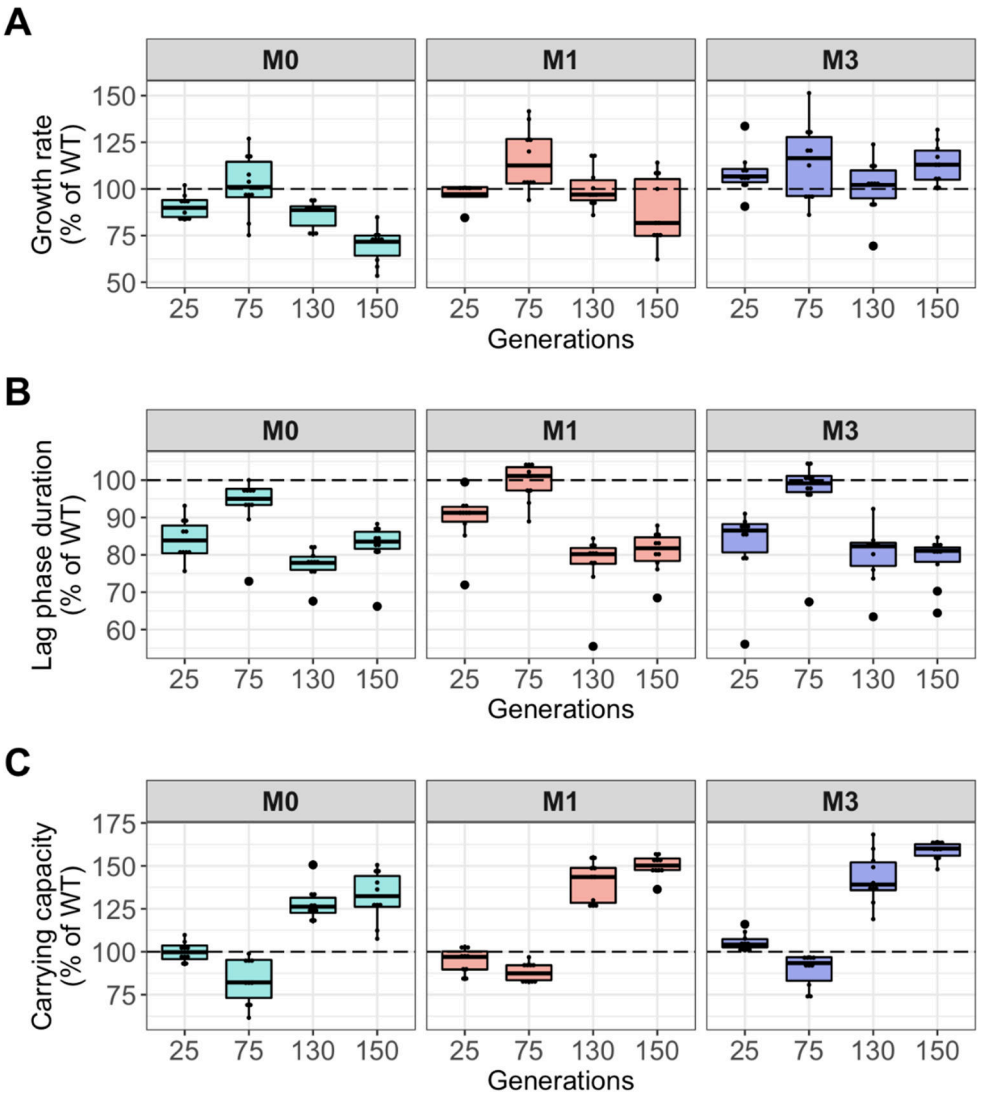


Figure 5. 6. Formatotrophic growth parameters in evolved isolates.

(A) Growth rate, **(B)** lag phase duration, and **(C)** carrying capacity measured for evolved isolates from each turbidostat at each sampling time-point, as determined via microplate growth assays. All values are expressed as a percentage of the WT reference. The individual points shown on the boxplots correspond to the average values of each tested isolate (*n* = 3 technical replicates per isolate, 10 isolates per turbidostat, per sampling timepoint).

Growth curves and kinetic parameters for the three best-performing isolates from this lineage (M307, M306, M304) are shown in Figure 5.7. These isolates were determined to have significantly higher growth rates than the WT reference ($p \leq 0.05$, two-sample t-test). Though individual clones with promising growth parameters were identified through the sampling process, no consistent pattern of growth rate improvement was recorded over the evolutionary trajectory of each lineage. This was partly due to noise in the microplate growth assays, evidenced by large differences in kinetic growth parameters between the unevolved WT references (controls) included at different sampling time-points (Figure S5.2). Due to the nature of the sampling protocol, evolved isolates could only be compared to the WT reference included within the same microplate growth assay. Comparisons between different assays may be misleading due, for instance, to small variations in the composition of growth media prepared on different days.

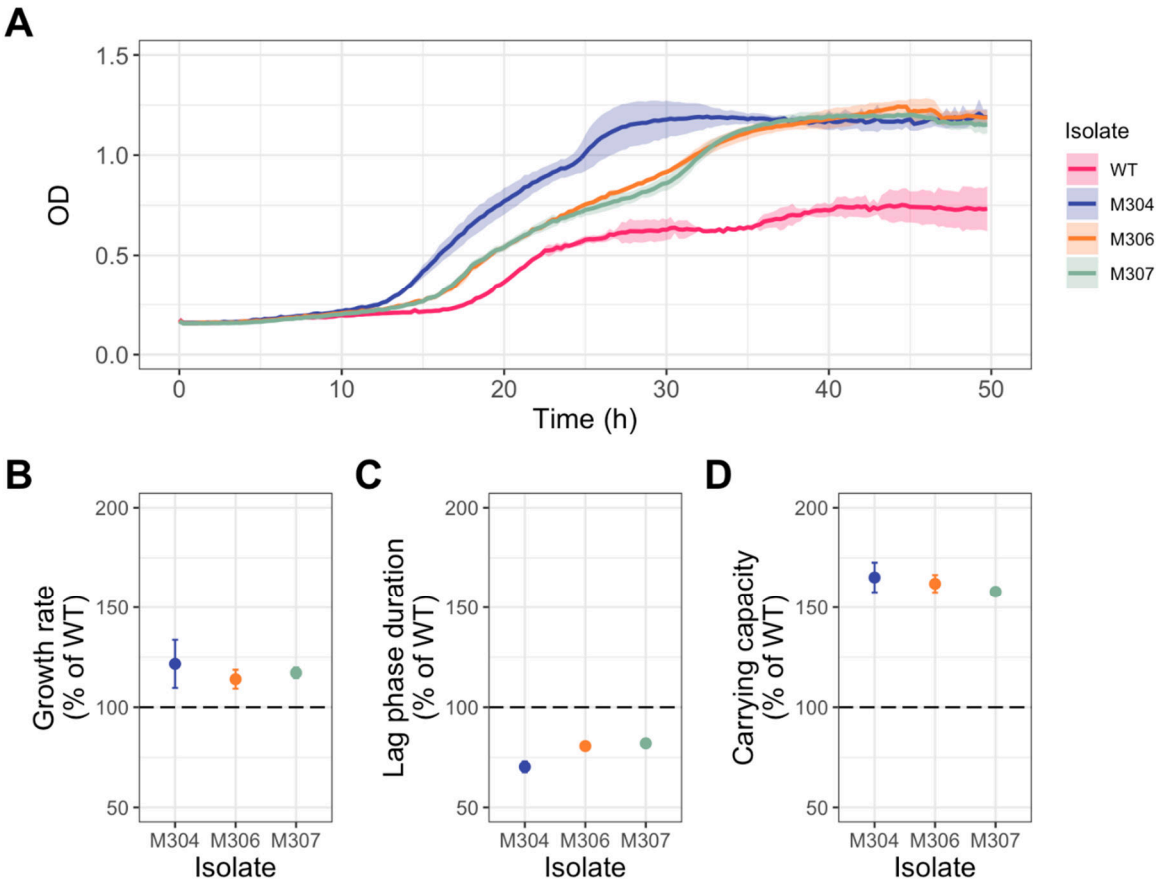


Figure 5. 7. Isolates from the M3 lineage exhibit improved growth kinetics.

(A) Growth curves for the top three M3 isolates from the 150-generation time-point. Solid lines show the average of *n* = 3 technical replicates, with the standard deviation indicated by the shaded region. **(B)** Growth rate, **(C)** lag phase duration, and **(D)** carrying capacity measured for each isolate as a percentage of the WT reference. Individual points show the mean of *n* = 3 technical replicates, ± one standard deviation.

We conducted an additional growth assay to address this experimental limitation. The best-performing isolates from each lineage and sampling time-point were simultaneously revived from frozen glycerol stocks, and their growth on formate was assayed on a single microplate, i.e. with a shared WT reference. Further, we assayed the selected isolates on both 80 mM and 160 mM sodium formate, to examine whether these evolved isolates exhibited higher formate tolerance and utilisation relative to the WT strain.

At 80 mM formate, a substantially shortened lag phase was observed for all evolved isolates relative to the WT (Figure 5.8 A and D). M3 isolates from the 150-generation time-point additionally exhibited a 39% improvement in carrying capacity (Figure 5.8E). We failed to replicate the improvements in growth rate and carrying capacity that had been previously observed for other 150-generation isolates, highlighting the variable nature of these results. At 160 mM formate, particularly for the M1 and M3 lineages, strains isolated after 150 generations had a shorter lag phase and higher carrying capacity relative to the WT control, indicating that these strains may have a higher formate tolerance (Figure 5.8B, D and E). No clear improvements in growth rate could be discerned at either 80 mM or 160 mM formate for evolved strains isolated after 150 generations (Figure 5.8 C). We note that a large variability between technical replicates was recorded for many strains during this growth assay, which may confound the observed results (Figure S5.3). Nevertheless, strains isolated from the M3 lineage after 150 generations consistently exhibited improvements in kinetic growth parameters, warranting further investigation.

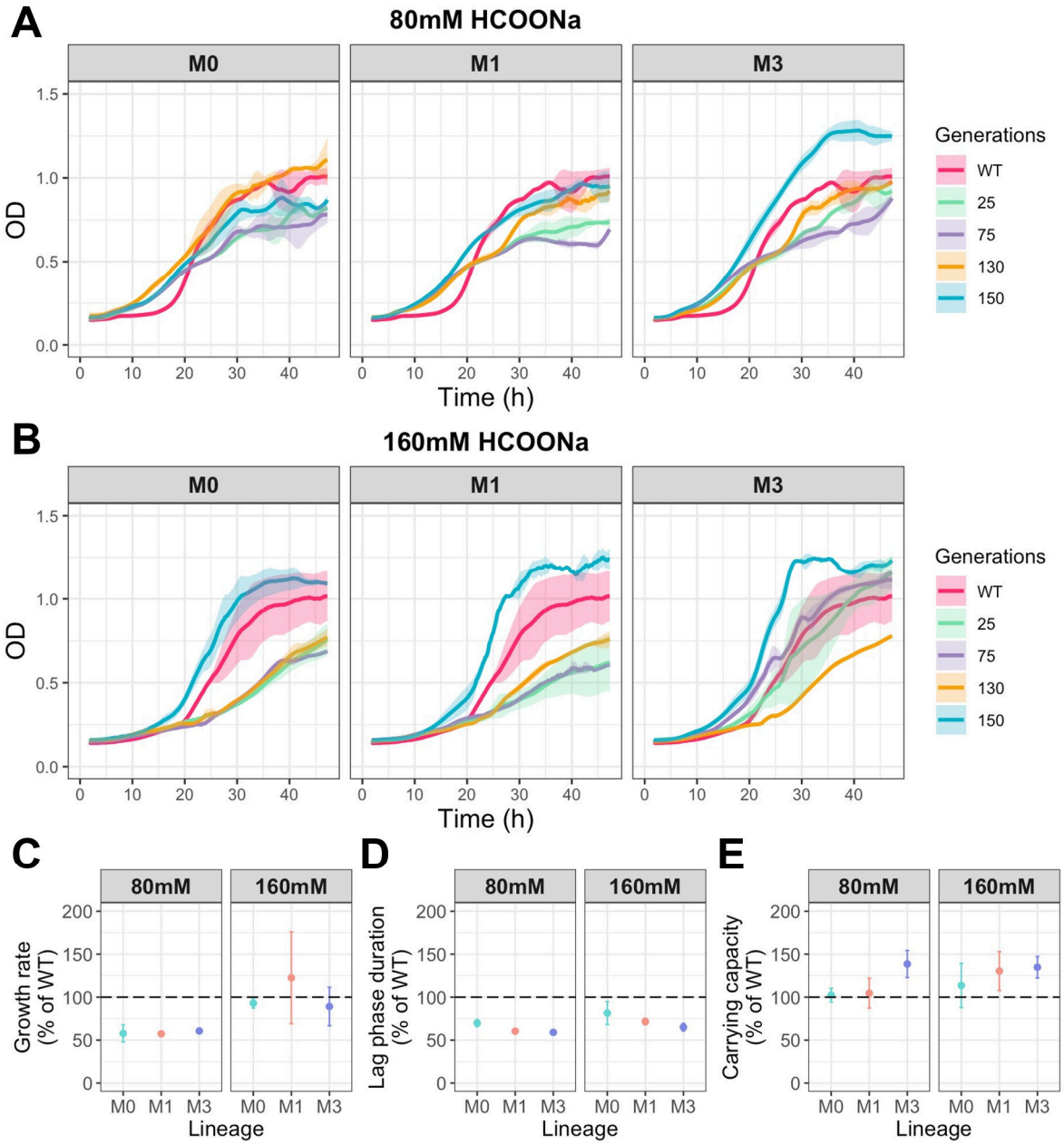


Figure 5. 8. Further analysis of evolved isolates.

Growth curves for the top isolate from each sampling time-point and turbidostat at **(A)** 80mM and **(B)** 160mM sodium formate. For each isolate, solid lines show the average of *n* = 3 technical replicates, with the standard deviation indicated by the shaded region. **(C)** Growth rate, **(D)** lag phase duration, and **(E)** carrying capacity for 150-generation isolates at both formate concentrations, as a percentage of the WT reference value. Individual points show the mean of *n* = 3 technical replicates, ± one standard deviation.

5.4.5. Comparative genomics of evolved isolates

We hypothesised that evolved isolates from the M3 lineage may share unique adaptive mutations underlying their improved growth phenotypes. To test this, we sequenced the genomes of the three best-performing evolved isolates from each lineage at the final time-point (150 generations). The ancestral WT strain was also sequenced and found to have 136 mutations that were not present in the NCBI database's reference genome for *C. necator*[137].

Surprisingly, only three of the mutations identified in evolved isolates were not present in their shared ancestor, and all sequenced isolates from the M0 lineage were found to be genetically identical to the WT (Figure 5.9). Strains isolated from the M1 and M3 lineages all contained a deletion within the CDS at the H16_A3118 locus, which encodes a hybrid sensor histidine kinase/response regulator with a CheY-like receiver domain (Pfam ID PF00072, see Figures S5.4), homologous to the *phcR* gene from the related organism *Ralstonia solanacearum* [255,256] (Figure S5.5). Strain M108 additionally carried a point mutation (transversion) at the H16_B2204 locus, which encodes a bifunctional glycoside hydrolase 114/polysaccharide deacetylase family protein. This gene product bears homology to the PelA extracellular matrix proteins from the closely related bacterium *Cupriavidus basilensis*[257] and the opportunistic pathogen *Pseudomonas aeruginosa*[258,259] (Figure S5.6), enzymes which are essential for the biosynthesis of the pellicle polysaccharide found within bacterial biofilms.

Interestingly, all three M3-derived strains were genetically identical. In addition to the deletion at the *phcR* locus shared with M1 strains, they also carried a point mutation

(transversion) at the H16_A2725 locus, which results in a non-synonymous amino acid substitution. The locus encodes one of two RpoD paralogues in *C. necator*, an RNA polymerase sigma subunit (sigma factor σ^{70}) involved in the expression of housekeeping genes[137].

Figure 5. 9. Genomic analysis reveals shared mutations in evolved isolates.

For each mutation, the frequency is defined as the percentage of Illumina sequencing reads for which the mutant variant was detected. A frequency of 0% indicates the WT allele was present at that specific locus. Individual strains are labelled according to their lineage, with the first two characters indicating which turbidostat they were isolated from (M0, M1 or M3). The following two digits indicate their isolate number (01-10, arbitrary). For each gene, loci names and KEGG IDs are shown.

5.5. Discussion and Conclusions

In this work, we have calibrated and demonstrated the utility of a versatile platform for continuous evolution of improved growth phenotypes, and applied it to engineer more proficient formatotrophic strains of *C. necator*. By conducting phenotypic and genotypic analyses of evolved isolates, we identified strains with promising kinetic growth parameters, which could be further characterised and applied toward sustainable bioproduction using this attractive C_1 feedstock.

Analysis of expected and observed evolutionary outcomes

The fundamental objective of the ALE experiment was to identify evolved isolates with improved formatotrophic growth phenotypes (specifically a higher growth rate), and to investigate their underlying genotypes. Only a small set of (possibly causative) mutations were identified in evolved strains, and all sequenced isolates from the M0 lineage were found to be genetically identical to the WT ancestor. Below, this observation is analysed through a simplified, back-of-the envelope calculation.

Using a simple model, the magnitude of the fitness benefit conferred by single-site mutations arising during the evolutionary trajectory can be approximated as function of the population size, the expected number of mutants in the population, and the number of generations in the evolution experiment. This relationship can be described by the following equation:

$$n_{mut} \cdot (1 + s)^t = b \cdot p$$

Equation 5.1

Where n_{mut} is the number of cells in the population with a specific beneficial single-site mutation, s is the fitness benefit conferred by said mutation, t is the number of generations, b is the fraction of the population the mutant would make up after t generations (assuming that it arises against a homogenous background, and in the absence of clonal interference or epistatic interactions), and p is the population size.

Briefly, Equation 5.1. approximates the growth of a mutant population across the evolutionary timeline. Using this simple model, we can perform a back-of-the-envelope calculation of the expected evolutionary outcomes of the ALE experiment. We assume that, to be detectable by our genomic analysis, the population of mutants carrying a specific beneficial single-site mutation at the end of the ALE experiment should constitute a significant fraction of the total population size. According to the calibration data presented in Section 5.4.3., we expect a maximum n_{mut} value of ~77 cells following UV irradiation at an LED intensity of 0.005. This was estimated from the fact that, at the peak of genetic diversity, an average of 77 NalX-resistant mutants were detected in the turbidostat population following UV exposure (Figure 5.4A). This estimation assumes that each mutant carries a single-site mutation which confers NalX resistance, and that all resistant mutants carry identical mutations. The average size of the bacterial population in the turbidostats (p) was 8x10^8 cells (see Section 5.4.2). Thus, to detect single-site mutants that make up 10% of the total population in the turbidostat after ~150 generations (b = 0.1, t = 150), the minimum fitness benefit conferred by an underlying causative mutation would have to be at least 10%

(s≥0.097). In other words, for a mutant to reach ~10% of the total population at the end of the experiment (and thus be near-measurable by our sequencing analysis) it would need to have a growth rate increase greater than 10%. Anything less than this is thus unlikely to be detected by the genomic analysis conducted in this work, where only a small number of isolates from the final timepoint of the experiment were sequenced.

In line with this reasoning, it is possible that the parameters of the ALE experiment (specifically the values for n_{mut} and t) may have set a too-high threshold s value for adaptive mutations to fix in the population and be detected in our analysis. It is therefore not unexpected that only a small number of mutations were identified in evolved isolates. For instance, whilst no mutations were found in strains isolated from M0 after 150 generations, no growth rate improvements were recorded for these strains in microplate-scale growth assays either (Figure 5.6A and 5.8A and C). Growth rates estimates from *in situ* turbidostat data also suggest that limited phenotypic improvements were accrued during the ALE experiment in M0 relative to M1 and M3 (Figure 5.5). Interestingly, M0-derived strains did exhibit reproducible improvements in other kinetic parameters, namely a shortened lag phase duration and a higher carrying capacity. Since we could not identify mutations underlying these phenotypes, we hypothesise that they may result from inducible tolerance to formate, i.e. physiological rather than genotypic adaptation in the evolved isolates. This could be investigated further through transcriptomic and proteomic studies.

In contrast, the top three strains isolated at the 150 generation timepoint from the M1 and M3 lineages exhibited average growth rate improvements of 10±4% and 18±4%,

respectively (Figure 5.6A). Compare this to the predictions from Equation 5.1, where if we assume a mutant must reach an abundance of ~50% of the population (b = 0.5) after 150 generations (t = 150) we would anticipate a minimum required growth rate benefit of ~11% (s = 0.109). Though these growth rate measurements were not fully replicated in follow-up experiments (Figure 5.8 A and C), the magnitude of the fitness benefits observed would thus be consistent with a small number of (or one) high-impact mutations being fixed, as was determined by genotypic analysis. Recapitulation of individual mutations in a WT background would be necessary to determine causal relationships between observed mutations and growth phenotypes with absolute certainty, and to refute alternative explanations for their fixation, such as genetic hitchhiking. In the absence of this experimental data, a more detailed analysis of the possible effects of these mutations is provided below, identifying avenues for future research.

Elucidating the effects of mutations in evolved isolates

The three best-performing isolates identified in the experiment came from the M3 lineage, and were found to be genetically identical, carrying mutations at the *rpoD* and *phcR* loci which were not present in the WT ancestor. The *rpoD* locus encodes an RNA polymerase sigma subunit (σ^{70}), which is responsible for mediating the expression of housekeeping genes[137]. Mutations in RNA polymerase subunits such as sigma factors lead to global changes in gene expression patterns, and are commonly observed in ALE experiments as a means via which cells may significantly shift global gene expression patterns[240]. For instance, mutations in the *rpoB* gene have been found to underpin increases in growth rate and metabolite productivity of *E. coli* grown

on glucose minimal medium[260,261]. In another ALE study, a single mutation in *rpoA* improved growth of *E. coli* on acetate by 25% in chemostat cultures[262].

Two RpoD paralogues are present on the *C. necator* genome[137]. A study in the related organism *Cupriavidus metallidurans* identified that RpoD paralogues may control the expression of different sets of genes[263], yet little is known about the RpoD regulons in *C. necator*. Transcriptomic analysis of both WT and M3-evolved strains across different growth conditions may not only elucidate differential gene expression patterns underlying the observed phenotypes, but also provide insights into global regulatory networks in *C. necator*, with important implications for strain engineering.

Further, the emergence of this mutation in our ALE experiment suggests that more targeted engineering of the RpoD regulon may be beneficial to obtain strains with improved formate growth kinetics. Specifically, global transcription machinery engineering (gTME) could be applied[264]. This method couples the generation of a library of RpoD mutants with enrichment screening to isolate strains in which global RpoD-mediated changes in gene expression led to desirable phenotypes. gTME has been successfully applied to engineer the metabolism of diverse microbes, for instance to obtain strains of *S. cerevisiae* and *Z. mobilis* with improved ethanol tolerance and production[265,266], to boost L-tyrosine production in *E. coli*[267], or to increase xylose tolerance and butanediol accumulation in *Klebsiella pneumoniae*[268].

The *rpoD* mutation was only observed in strains derived from the M3 lineage. In contrast, the *phcR* mutation emerged in both M1 and M3 lineages. This is a biologically significant observation, which highlights the adaptive advantage that this mutation may

confer. PhcR functions as a response regulator within a quorum sensing mechanism, homologous to that of the related bacterium *Ralstonia solanacearum*[255,256], which is a plant pathogen. In *R. solanacearum*, PhcS and PhcR form a two-component signalling system which is responsible for activating the expression of diverse virulence factors in response to changes in cell density (quorum sensing). At low cell density, PhcR represses the expression of *phcA*, a LysrR-type transcriptional regulator responsible for mediating differential gene expression leading to pathogenicity. At high cell density, the accumulation of the extracellular signalling molecule 3-hydroxypalmitic acid methyl ester (3-OH-PAME) is sensed by the PhcS-PhcR system, leading to relief of repression on *phcA*. Alongside the upregulation of virulence factors, PhcA is also known to repress cell motility, twitching motility and surface adherence[269–271].

Though *C. necator* is a non-pathogenic organism, its homologous Phc regulatory network has been shown to mediate many of the same phenotypic changes, including exerting control over cell motility[92,256]. A recent study by Calvey *et al*. established that mutations within the Phc cell density sensing system (namely a full deletion of *phcA*) could improve growth of *C. necator* not only on formate, but also when growing heterotrophically on fructose or succinate[92]. In their study, mutations at the *phcA* locus were identified in several strains isolated from an ALE experiment (serial batch culture in shake flasks towards improved growth on formate). By recapitulating the observed *phcA* mutation in a WT background, the authors demonstrated that this deletion results in a reduction of cell motility and flagellar synthesis, which confers a selective advantage on diverse carbon sources by enabling a more frugal expenditure of cellular resources.

PhcR acts upstream of PhcA in the quorum-sensing regulatory network. Thus, it is unclear whether the observed deletion at the *phcR* locus in M3-derived strains would relieve or increase repression of *phcA* expression. In line with the results presented by Calvey *et al*., a relief of *phcA* repression would result in increased cell motility, whilst constitutive repression would have the opposite effect. Further phenotypic characterisation of the M3 evolved isolates, for instance transcriptomic and proteomic analysis, would be necessary to determine the effects of the observed *phcR* deletion and how these may result in improved growth on formate.

The need for further phenotypic analysis of evolved isolates

In their study, Calvey *et al*. demonstrated that the magnitude of the growth improvement conferred by the *phcA* deletion depended on the scale at which the bacteria were cultured. Namely, they observed that Δ*phcA* mutants of *C. necator* had a 32% higher growth rate than the WT when grown on formate in shake flasks (where the mutants first evolved), but that this advantage was reduced to 12% when the same strains were grown in pH-controlled bioreactors. This highlights how mutations that arise in ALE experiments may confer growth advantages which are specific to the environment in which the experiment was performed. In line with this, it would be interesting to assay the growth kinetics of the M3-derived strains at different scales, including batch and turbidostat cultures in Chi.Bio reactors, as well as batch cultures in shake flasks. Due to time constraints, we were only able to conduct phenotypic analyses (growth assays) of the evolved isolates in microplate format, where we observed a large degree of variability between the growth trajectories of technical replicates, as well as between experiments conducted on different days. This made it

difficult to accurately characterise kinetic growth parameters. Though noisy, the turbidostat data shown in Figure 5.5 suggest that growth rate advantages may be more pronounced at this growth scale, particularly for strains derived from the M1 and M3 lineages. Attempting to reproduce these observations in the turbidostats at the isolate (rather than population) level may elucidate whether improvements in growth rate are context-dependent, and yield a more accurate measurement of their magnitude. In general, assaying growth at different scales would provide a more comprehensive characterisation of the evolved isolates' formatotrophic growth kinetics.

Nevertheless, we were able to discern growth advantages in evolved isolates, specifically for M3-derived strains. Though we did not observe reproducible increases in growth rate on formate, the genetically identical M304, M406 and M307 isolates displayed shortened lag phases and higher final ODs (carrying capacities), with consistent improvements recorded across different experiments and at different formate concentrations. These measured advantages indicate that the evolved strains may have a faster formate consumption rate than the WT, as well as an improved ability to assimilate formate into biomass, both of which would lead to a higher formate tolerance. It is surprising, however, that these improvements were not accompanied by a significantly higher growth rate, especially when considering that this was the only trait that was directly selected for in the ALE experiment. Conducting growth assays of the M3 isolates at different scales (especially in Chi.Bio reactors, where they were first evolved) would be beneficial to shed light on this apparent inconsistency, as well as to accurately determine their potential for bioproduction processes. Detailed measurements of formate consumption rates and substrate inhibition kinetics would also be beneficial.

Practical limitations of the evolution experiment

Over the course of ALE, we identified several practical limitations and experimental design flaws, which should be addressed in future work to harness the evolutionary platform more productively. For instance, frequent unplanned stoppages of the turbidostat cultures made it difficult to maintain a constant selective pressure, and often resulted in the introduction of genetic bottlenecks through culture dilution. The causes of these stoppages were diverse, but most often they occurred due to liquid spills from the glass vials. Hardware re-design would likely be required to circumvent this limitation.

Difficulties were also encountered when recording *in situ* optical density measurements (and related growth rate estimations), due for instance to bacterial wall growth (biofilm formation) within the glass vials. The formation of biofilms during continuous cultures in turbidostats and chemostats is a well-documented limitation of these systems[272,273], and is particularly problematic for turbidostats, where “sticky” variants can trigger premature culture dilution, further amplifying the selective advantage conferred by biofilm formation. As a result of this, the turbidostat population may also experience unplanned fluctuations in size, which may affect the evolutionary outcomes of the experiment. In future work, this could be avoided by more frequent replacement of the glass vials used within Chi.Bio turbidostats, or by using an automated system to periodically cycle the cultures between different reactors or growth chambers, thereby negating the selective advantage of biofilm formation. Previous studies have implemented creative solutions to this problem, such as the

"Evolugator", where cultures are cycled between different growth chambers formed by sectioning off different portions of the same tubing[274].

Considerations for improving experimental design in future work

In addition to these practical limitations, flaws in the experimental design resulted in the utility of the automated evolution platform not being fully exploited. For instance, we opted to use a single permissive concentration of formate in the growth medium throughout the experiment to maintain a constant selective pressure. However, as strains adapt to this environment, the real selective pressure that they experience decreases. The turbidostat platform allows this limitation to be circumvented via feedback-controlled modulation of selective pressure. Namely, the concentration of sodium formate in the growth medium could have been dynamically adjusted as a function of the population's growth rate, thereby maintaining the same level of selective pressure upon growth rate evolution. Moreover, it is known that applying a strong selective pressure increases the likelihood of high-benefit adaptive mutations being fixed[250]. In line with this, increasing the level of formate stress on the bacterial populations along the evolutionary trajectory would have increased the likelihood of fixing highly beneficial mutations over a shorter timescale.

The evolution experiment could have also been modified to synergistically combine random mutagenesis with increases in selective pressure. By calibrating UV dosage in the Chi.Bio turbidostats, we determined that UV mutagenesis leads to genetic diversification of *C. necator* cultures, and identified a peak in the population's genetic diversity roughly 24 hours after irradiation. Formate stress could have been applied at

this point to modulate the selection coefficient, i.e. to rapidly select for high-benefit adaptive mutations that might have emerged as a result of UV exposure. This could be explored in future work.

The protocol for applying UV mutagenesis *in situ* could also be optimised further. In many ALE studies, a constant, low dose of mutagen is applied throughout the evolution experiment (*see* ref.[242] for an example). We opted against this method as it may lead to the emergence of UV resistance in the evolving population[275]. The mutagen itself can become part of the selective pressure, further complicating the strain optimisation landscape which the bacteria must traverse, and by extension making it difficult to identify the causative mutations underlying any improved phenotypes. Instead, we applied two rounds of UV mutagenesis to accelerate the rate of adaptation by periodically increasing the genetic diversity of the evolving bacterial populations.

However, genotypic analysis revealed a very small set of mutations amongst the best-performing evolved isolates. Amongst other causes, this could suggest that mutagenesis might not have been applied most productively to diversify the population. To investigate this, a no-UV control could be included in the experiment, to test whether and to what extent random mutagenesis shortens the timescale of adaptive evolution in this specific experimental context. The UV LED intensity, length of irradiation, and frequency of UV exposure could also be further optimised through a multifactorial process. Though we did consider this possibility, we were limited by the number of turbidostats which could be operated in parallel and the relatively long timescale of the experiment.

Ideally, ALE experiments should not be interrupted until a stable, desirable phenotype is observed for the specific trait which is being targeted. In our study, the ALE experiment was interrupted after ~150 generations due to time constraints. While the duration of ALE experiments varies widely, it is typically in the range of ~500 to 2,000 generations[247,242], with some notable exceptions such as the Lenski Long Term Evolution Experiment (LTEE)[276,277]. The timeline of our experiment is on the lower end of this scale. Though other studies have successfully evolved improved growth rates in as little as 100 generations[278], extending the timescale of ALE may have increased the likelihood of isolating strains with improved phenotypic traits, underpinned by either high-impact or low-impact adaptive mutations (see Equation 5.1). In future work, an extended timeline could be paired with more productive experimental design, as discussed above.

Conclusion

In conclusion, our study delivers two important contributions. Firstly, we have demonstrated how the versatile, programmable Chi.Bio platform can be applied to evolutionary engineering, identifying practical limitations associated with its operation and providing avenues for future research and optimisation. Secondly, we have identified *C. necator* strains with improved formatotrophic growth phenotypes which could be investigated and engineered further, or directly applied in C_1 bioproduction processes.

5.6. Chapter 5 Appendix – Supplementary Information

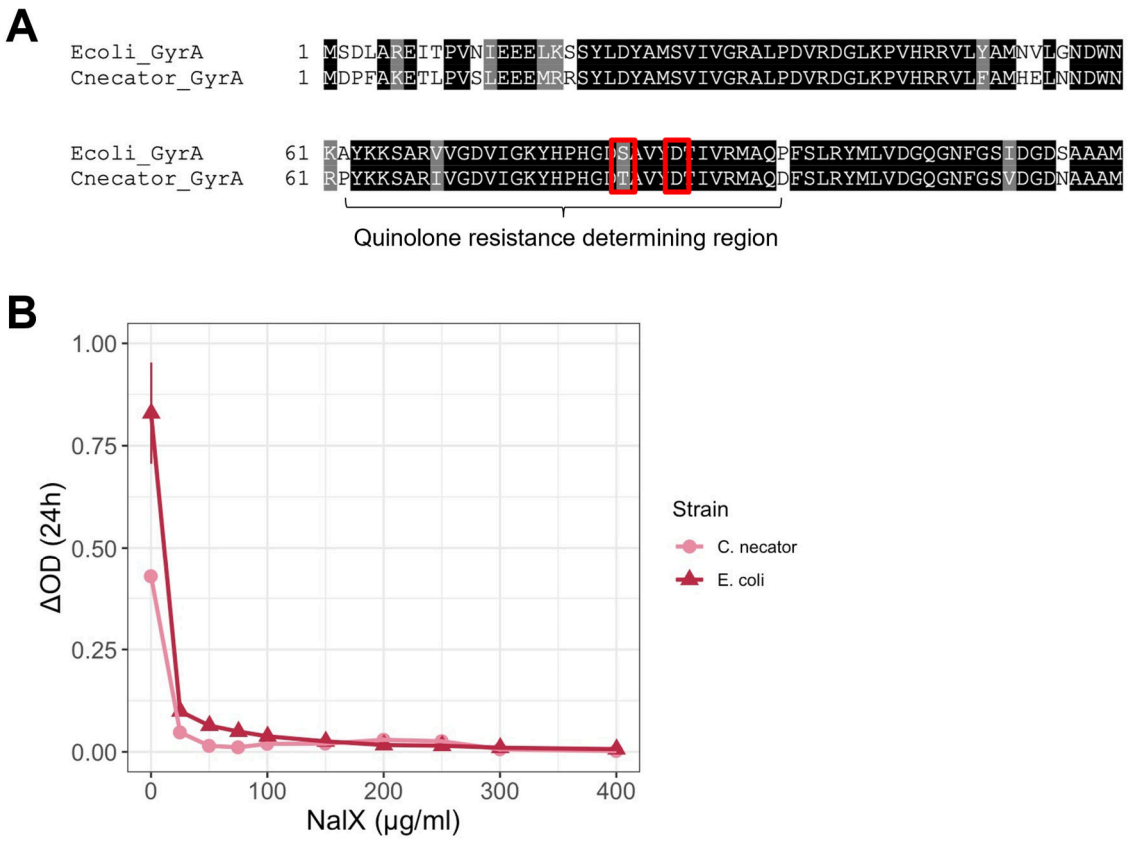


Figure S5. 1. Nalidixic acid tolerance in *C. necator.*

(A) Section of a multiple protein sequence alignment of the A subunit of GyrA from *E. coli* and *C. necator*. The alignment was performed using the Clustal Omega (ClustalW) tool[194]. Complete or partial residue identity is indicated by black or grey shading of relevant residues, respectively. Partial identity (or similarity) in this context is defined as the conservation of the physicochemical properties of the residue at any given position. Amino acid positions where substitutions can arise from point mutations and confer NalX resistance are indicated by red rectangles. In *E. coli* and other Enterobacteria, an S83I amino acid substitution may confer quinolone resistance. This substitution is also possible via a single point mutation in *C. necator* (ACC to ATC). Point mutations that result in amino acid substitutions at D87 may also confer resistance, with more flexibility allowed in the nature of the substituting residue. **(B)** NalX toxicity in *E. coli* and *C. necator*, as measured via a minimum inhibitory concentration assay. ΔOD (24h) is defined as the increase in optical density observed for bacterial populations cultured at each antibiotic concentration over a 24h period. Individual points show the mean of n = 3, ± one standard deviation.

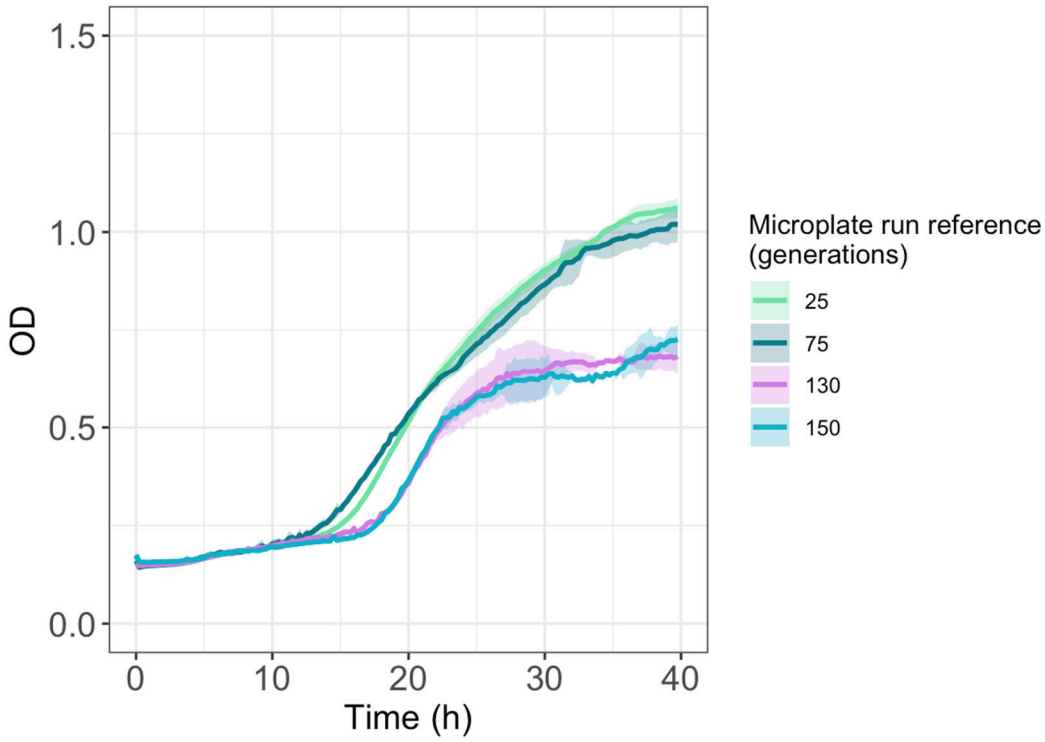


Figure S5. 2. Growth curves for WT references in formatotrophic growth assays.

Individual growth curves correspond to the unevolved WT control included in the microplate growth assays for each sampling time-point. Solid lines show the average of *n* = 3 technical replicates, with the standard deviation indicated by the shaded region.

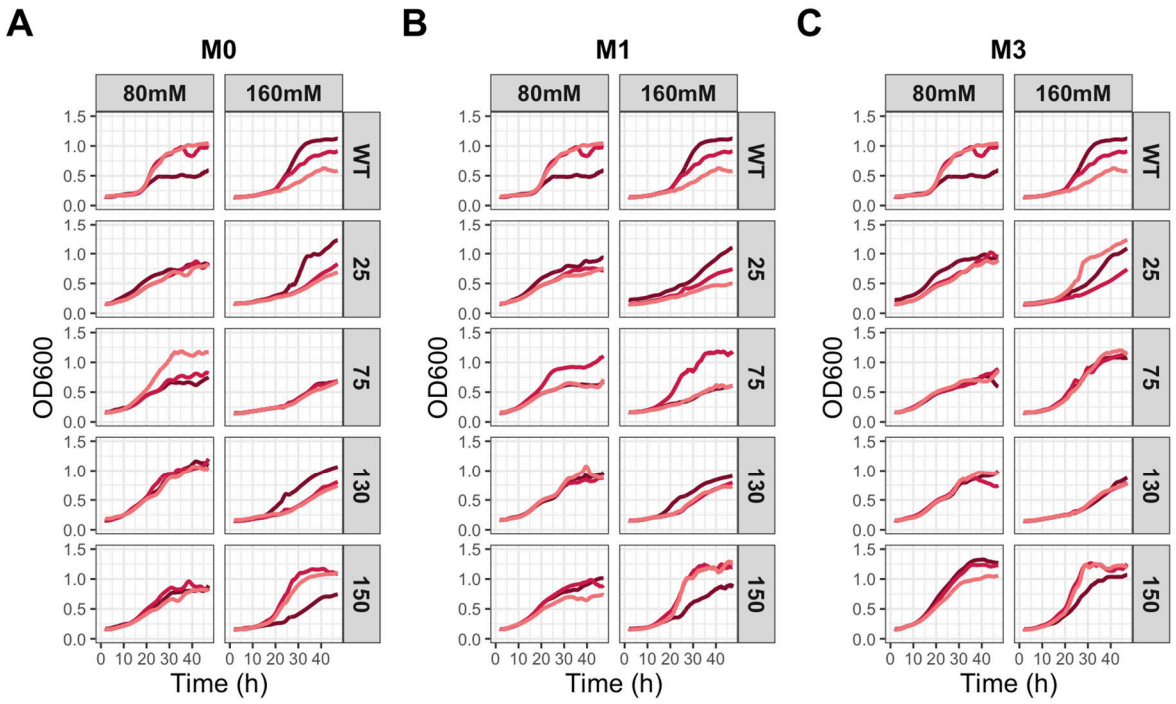
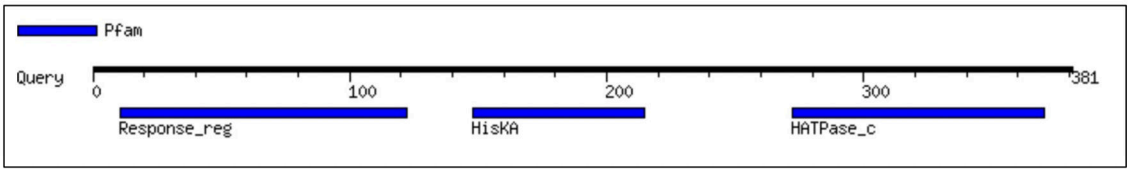


Figure S5. 3. Variable growth trajectories in microplate assays.

Growth curves for single technical replicates of top isolates from the **(A)** M0, **(B)** M1, and **(C)** M3 lineages. In each panel, matrix labels indicate the sampling time-points from which the isolates were derived (25, 75, 130 or 150 generations) and the sodium formate concentration in the medium (80mM or 160mM). Note that the WT references are identical in all three panels, and are provided in triplicate for ease of reference.

A



B

Pfam ID	Position (Independent E-value)	Description
PF00072	11-122(1.7e-15)	Response regulator receiver domain
PF02518	272-370(1.7e-12)	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
PF00512	148-215(5.5e-07)	His Kinase A (phospho-acceptor) domain

Figure S5. 4. Locus H16_A3118 encodes a histidine kinase/response regulator.

The MotifFinder bioinformatic tool, available through the KEGG database[279,280], was queried with the protein sequence encoded by the H16_A3118 CDS. **(A)** Schematic overview of protein motifs along the query sequence and **(B)** detailed description of the precise location and nature of the motifs.

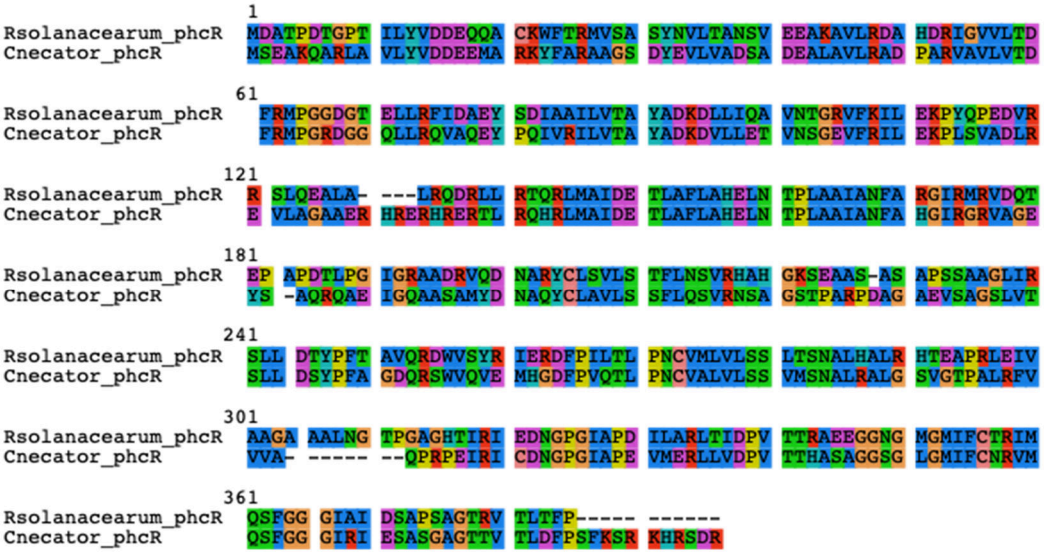


Figure S5. 5. Locus H16_A3118 encodes a PhcR homologue.

Multiple protein sequence alignment of PhcR from *C. necator* and *R. solanacearum*. Sequences were aligned with the Clustal Omega (ClustalW) tool[194]. Alignments were formatted in Seaview using the default colour scheme[281].

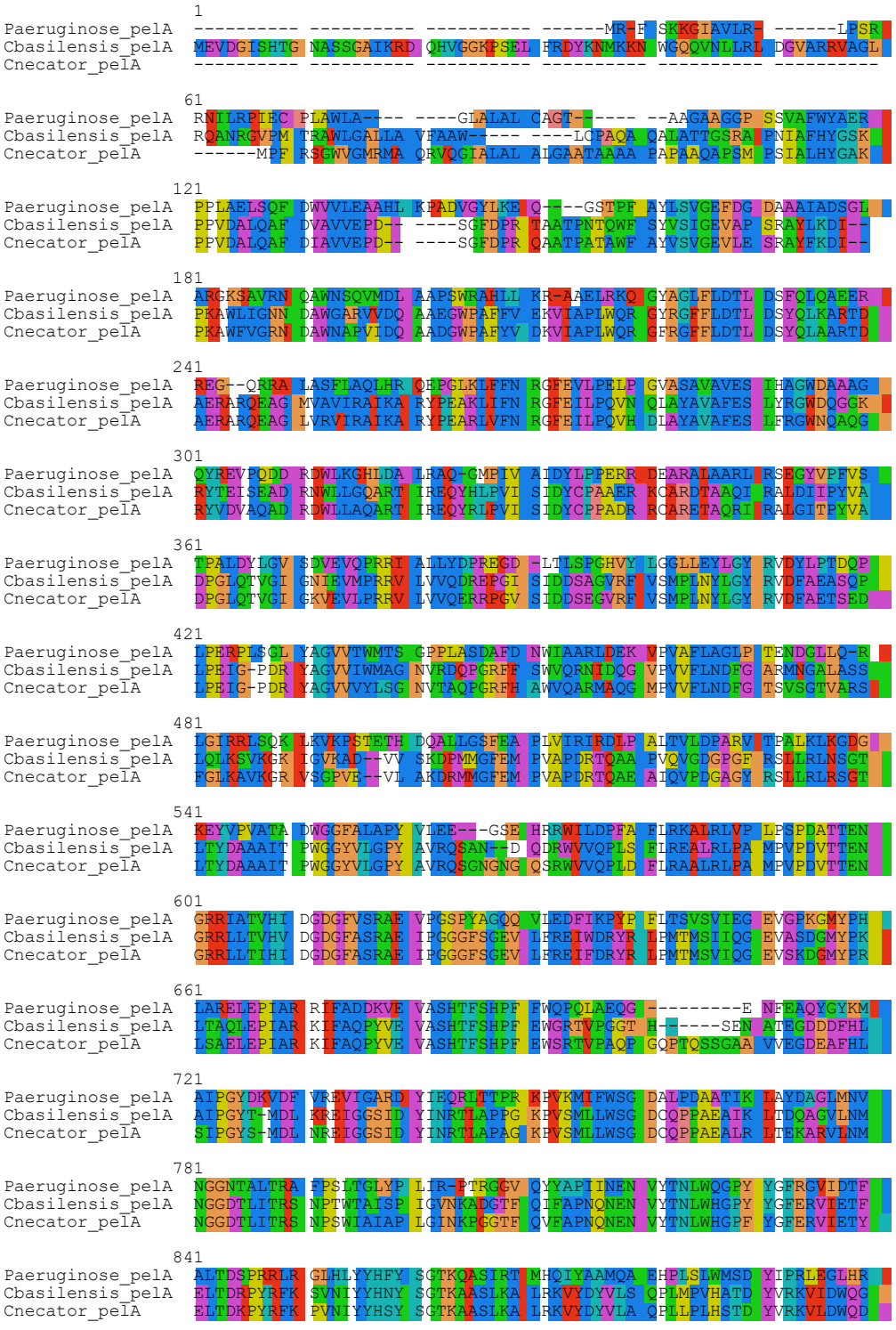


Figure S5. 6. Locus H16_B2204 encodes a PelA homologue.

Multiple protein sequence alignment of the PelA enzyme from *C. necator*, *C. basilensis* and *P. aeruginosa*. Sequences were aligned with the Clustal Omega (ClustalW) tool[194]. Alignments were formatted in Seaview using the default colour scheme[281].

6. Metabolic engineering of *Cupriavidus necator* for efficient conversion of acetate to ethanol

6.1. Aims and Objectives

In this Chapter, we aim to engineer *C. necator* as an efficient host chassis to produce liquid fuel (ethanol) from acetate, a two-carbon (C_2) feedstock which can be sourced renewably. To do so, we aim to conduct a detailed characterisation of both acetate and ethanol tolerance in this bacterium, and to investigate how the (regulated) expression of both heterologous and endogenous genes can be used to improve production parameters (titer, rate, and yield).

6.2. Statement of collaborative contributions

The work presented in this Chapter was conducted in close collaboration with Dr Fariza Ammam, who contributed to project ideation and jointly performed the ethanol production experiments. Experiments testing the ethanol tolerance of *C. necator*, as well as evaluating the performance of different inducible gene expression systems, were solely performed by Dr Ammam and are included in this Chapter for completion. Additionally, Dr Robert Habgood and Jochem Nielsen helped in designing and cloning the plasmids described in this study.

6.3. Introduction

Industrial biotechnology applies microorganisms to the bio-based production of pharmaceuticals, biofuels and other high-value chemicals, aiming to provide a more sustainable alternative to petrochemical-based synthesis[2]. Traditionally, industrial biotechnology has focused largely on harnessing plant biomass as feedstock, which provides an excellent carbon source for conventional heterotrophic bioproduction host organisms such as *Escherichia coli* and *Saccharomyces cerevisiae*.

Whilst more sustainable than the petrochemical alternative, the use of plant biomass as feedstock for bioproduction presents several challenges and limitations. Natural photosynthetic systems operate with low efficiency, leading to a low sunlight-to-product energy transfer efficiency in the resulting biomanufacturing systems[9]. Additionally, the cultivation of crops to be used as industrial biotechnology feedstock is environmentally costly, straining the availability of arable land and scarce resources such as freshwater and minerals[7,282]. Large monocultures of bioenergy crops are also known to have a detrimental effect on biodiversity[283]. Considering this, there has been extensive interest in finding alternative, sustainable and energy-efficient strategies to generate feedstock for industrial biotechnology. This is particularly urgent when considering the sustainable production of liquid fuels, with industry leaders having emphasised the potential benefits of placing microorganisms and synthetic biology at the centre of the imminent energy transition[284].

As outlined in Chapters 1 and 5, electrochemically-generated one-carbon (C_1) compounds such as formate are promising feedstocks, yet there are currently several limitations to their application in microbial growth and bio-production. Two-carbon (C_2) compounds, particularly acetate, have also attracted growing interest as renewable feedstocks for bio-based production[285,286]. Acetate has a higher energy density than formate, is less toxic to bacterial cells, and does not require specialised autotrophic metabolism, making it accessible to a wider range of microbial hosts. It is therefore a more tractable carbon source for immediate application in sustainable bioproduction.

Whilst feasible, the electrochemical production of acetate from CO_2 is less efficient than electrochemical formate production[104]. As an alternative, acetate can be sourced in abundance from cheap renewable sources, such as anaerobic digestion of organic waste[287,288], C_1 gas fermentation[289,290], or the pre-treatment of lignocellulosic biomass[291,292]. As with their C_1 counterparts, a key consideration in sustainable bio-based production from renewable C_2 feedstocks lies in harnessing microbial host organisms that can metabolise these compounds most efficiently to produce target molecules. The facultative chemolithoautotroph *Cupriavidus necator* can use a range of carbon substrates for growth, including the short volatile fatty acids acetate, propionate and butyrate[293]. Under unbalanced nutrient conditions, *C. necator* produces large quantities of polyhydroxyalkanoates, which can be accumulated to reach up to 80% of its dry cell mass[294]. The possibility of re-routing this unique carbon storage pathway to other valuable metabolites such as biofuels has been extensively explored, and *C. necator* has been successfully applied to the production of a wide range of industrially-relevant products[295–298]. Additionally, the biomass of *C. necator*

can be used as a sustainable source of single-cell protein for alternative food and feed[299–301], having been granted Qualified Presumption of Safety (QPS) status in the European Union[302].

Owing to its versatile and malleable metabolism and demonstrable potential for bioproduction, as well as an expanding availability of tools for its genetic engineering[133], *C. necator* is an ideal platform strain for converting renewable carbon substrates such as acetate into industrially-relevant biochemical products. Here, we focus on engineering a *C. necator* strain for the efficient bioconversion of acetate into the liquid fuel ethanol. Bio-based ethanol production has been a cornerstone of industrial biotechnology for several decades, having been introduced into transport fuel supply chains as early as the 1970s[303]. More recently, the possibility of "up-cycling" ethanol through its conversion to more specialised liquid fuels has been explored. For instance, the ethanol-to-jet-fuel production pathway has attracted growing interest from academics, governments, and industry alike[304,305]. Thus, finding more environmentally and economically sustainable feedstocks and host organisms for ethanol bioproduction is important to guarantee its long-term viability as a fossil fuel substitute.

Previous work has demonstrated that it is possible to engineer *C. necator* strains for ethanol production from acetate as the sole carbon source[306]. In this study, we build on this recent development to deliver ethanologenic *C. necator* strains with improved production titer, rate and yield. Having characterised the bacterium's tolerance to acetate and ethanol in the growth medium, we implement an inducible, two-enzyme biosynthetic pathway comprised of a heterologous alcohol dehydrogenase (AdhE)

from *E. coli* and an endogenous acetyl-CoA synthase. By combining this pathway design with a more aerotolerant variant of the AdhE enzyme, we constructed an ethanologenic strain of *C. necator* achieving a maximum ethanol titer of >1 g/l, corresponding to ~15% of the theoretical yield.

6.4. Materials and Methods

Bacterial Strains and Culture Conditions

All strains used in the study are detailed in Table S6.1. *Escherichia coli* strains were grown at 37 °C in LB (Luria Broth, Fisher). When appropriate, kanamycin (50 µg/ml) was added to the E. coli culture medium. *C. necator* H16 strains were grown at 30 °C in a rich medium (2.75% Tryptic Soy Broth) for precultures, or in a minimal medium supplemented with acetate. The composition of minimal medium used for *C. necator* cultivations is 1.45 g/l NaH_2PO_4, Na_2HPO_4, 1.0 g/l NH_4Cl, 0.5 g/l $MgSO_4$ x 7 H_2O, 0.01 g/l $CaCl_2$ x 2 H_2O, 0.005g/l $MnCl_2$ x 4 H_2O, 0.005 g/l $NaVO_3$ x H_2O, 0.5% (v/v) of SL-6 Trace element solution, and 0.5% (v/v) of vitamin solution containing the following vitamins : riboflavin 0. 1 g/l , thiamine-HCl x 2 H_2O 0.5 g/l , nicotinic acid 0.5 g/l pyridoxine-HCl 0.5 g/l Ca-pantothenate 0.5 g/l , biotin 0.1 g/l , folic acid 0.2 g/l. The final pH of the medium was adjusted to 6.8. The following antibiotics were added to the culture medium where appropriate: gentamycin (10 µg/ml), kanamycin (100 µg/ml).

Plasmid Construction

All molecular cloning was undertaken following standard methods in *E. coli* NEB® 5-alpha[307]. Cloning reagents were sourced from New England Biolabs (NEB) and used according to the manufacturer's specifications. All cloning steps were verified by Sanger sequencing. A metabolic pathway for ethanol production comprised of an aldehyde-alcohol dehydrogenase from *Escherichia coli* MG1655 (*adhE*; KEGG ID b1241) and the native acetyl-coenzyme A synthetase from *C. necator* H16 (*acoE*; KEGG ID H16_A2525) was utilised in this study. All plasmid vectors used in this study are listed in Table S6.2. The pSEVA231 vector[177,178] was used as the backbone for all *C. necator* expression plasmids. The *araC*-P_{araBAD} (arabinose-inducible), *xylS*-P_{PM} (toluic acid-inducible), and *hpdR*-P_{hpdH} (3-hydroxypropionate-inducible) transcriptional systems were obtained from pLO11 plasmid DNA[308], pCK227 plasmid DNA[309], and *Pseudomonas putida* UWC1 genomic DNA, respectively. All DNA fragments were amplified using Q5 High-fidelity polymerase and were assembled via HiFi DNA assembly (NEB), following the manufacturer's instructions. Oligonucleotides used in plasmid assembly are outlined in Table S6.3. Oligonucleotides used for assembly of plasmids pRH412, pRH422 and pRH423 contained spacers that introduced an ApaI cut site between the promoter region and the ethanol biosynthesis genes. Spacers also encoded the RBS for *adhE* (as present in the pHM07 plasmid) and a strong RBS for *acoE* (BBa_B0034). Plasmid pRH412K was constructed via site-directed mutagenesis, using vector pRH412 as a template to introduce the E568K mutation within the *adhE* coding DNA sequence. Plasmids were delivered into *C. necator* via conjugation from the donor strain *E. coli* S17 as described by Chen *et al*.[310], or via the electroporation protocol described in Chapter 2 once it became available.

Batch culture ethanol production experiments

Single colonies of all *C. necator* strains were precultured in rich medium. Cells from saturated precultures were washed twice and resuspended in minimal medium before being used to inoculate shake flask cultures to an initial OD of 0.075. Cells were grown in 50 ml minimal medium with 10 g/l acetate as the sole carbon source in 250 ml conical flasks, incubated at 30 °C with continuous shaking at 150 rpm for 72 h. Growth was monitored by measuring the optical density (OD) at 600 nm using the Shimadzu series UV-1800 UV/VIS spectrophotometer. All experiments were performed in triplicate. The plasmid-encoded biosynthetic pathway was induced by adding 0.1% (w/v) L-arabinose upon reaching an OD = 1-1.5 (exponential growth phase).

Determination of acetate and ethanol tolerance

For acetate tolerance determination, cells were precultured in rich medium, washed, and used for inoculation of mineral medium with acetate as sole carbon source, added in concentrations ranging from 0.001 to 12 g/l. Cultures were inoculated at an initial density of OD = 0.05. Growth was monitored by measuring OD of the cultures over time within a Chi.Bio reactor[183] operating in batch culture mode. For ethanol tolerance determination, cells were pre-cultured as above and grown in mineral medium with 10 g/l acetate, supplemented with ethanol in concentrations ranging from 0 to 10 g/l. Cells were cultured in 50 ml of medium, in 250 ml shake flasks. Growth was monitored by measuring the OD using a UV/VIS spectrophotometer, as described above.

Analytical methods

Samples for ethanol and acetate quantification were collected during bioproduction experiments (1 ml of the shake-flask batch culture was sampled at each timepoint). Cells were pelleted by centrifugation, and the supernatant was retained for chemical analysis. Acetate concentrations were determined by GC-FID (Shimadzu GC-2010) equipped with a flame ionisation detector (GC-FID) and a 30 m x 0.25 mm x 0.25 µm fused-silica capillary ZB-FFAP column (Phenomenex). The temperatures of the detector and injector were 350 °C and 250 °C, respectively. The oven temperatures were set at 100 °C for 2 minutes and then increased at a rate of 16 °C/min until reaching 180 °C. Ethanol concentrations were also quantified by GC-FID (Shimadzu GC-2010), equipped with a 30 m x 0.25 mm x 0.5 µm 30 SH-Stabilwax column (Shimadzu). The temperature of the injection port and detector was set at 250 °C, and the oven temperature was set at 45 °C for 5 minutes then increased at a rate of 12 °C/min until reaching 150 °C. The carrier gas was helium for both ethanol and acetate detection.

6.5. Results

6.5.1. Acetotrophic growth kinetics of *C. necator*

C. necator was chosen as a host organism for this study owing to its natural ability to use simple C_1 and C_2 compounds as carbon and energy sources. However, high concentrations of short-chain organic acids such as acetic acid (acetate) are toxic to bacterial cells. As they enter the cytoplasm, proton dissociation causes acidification, thereby depleting the proton motive force across the cell membrane[311]. We assessed

substrate toxicity and growth kinetics in aerobic culture to determine which acetate concentrations would be suitable for the ethanol bioproduction process. Namely, we measured the growth of *C. necator* H16 on different acetate concentrations in the range of 0.001-12 g/l (Figure 6.1).

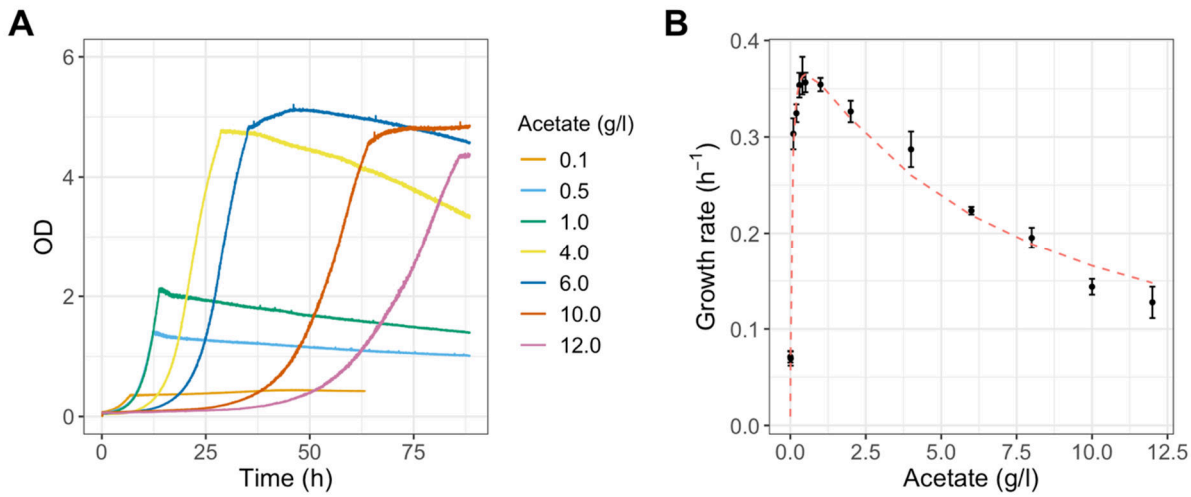


Figure 6. 1. Acetate tolerance in *C. necator* H16.

(A) Representative growth curves of *C. necator* cultures grown with acetate (0.1-12 g/l) as sole carbon source. Growth rate, lag phase duration and carrying capacity are a function of the initial substrate concentration. **(B)** Observed growth rate against acetate concentration for all tested acetate concentrations (0.001-12 g/l). Individual experimental data points are shown in black, with a Haldane equation fit shown as a dotted orange line.

The effect of substrate concentration on the growth rate for toxic substrates such as acetate is described by the Haldane equation:

$$\mu = \frac{\mu_{max} S}{K_s + S + \frac{S^2}{K_I}}$$

Where μ is the observed growth rate, μ_{max} is the maximum growth rate of this organism on the substrate in question, S is the initial substrate concentration, K_s is the saturation constant and K_I is the inhibition constant.

The relationship between *C. necator* growth rate and acetate concentration as described by the Haldane model is shown in Figure 6.1 B. Values for the experimentally-determined values for μ_{max}, K_s and K_I are shown in Table 6.1. The experimentally derived values were found to correlate well with previous reports in the literature[293,312–314]. We determined concentrations of acetate between 4 and 10 g/l to be acceptable for the cultivation of *C. necator* for ethanol production in this study, as they allow the bacterial population to reach a high cell density (OD ≥ 4.5) within 72h of inoculation. Given that acetate is an inexpensive substrate that can be derived from waste streams, maximising its starting concentration in the production process is desirable, as this will simultaneously maximise waste degradation and ethanol titer.

Table 6. 1. Kinetic parameters describing acetotrophic growth in *C. necator*.

Parameter	Value
μ_{max}	0.426 ± 0.010 h^{-1}
K_s	0.045 ± 0.001 g/l
K_I	6.383 ± 0.464 g/l

6.5.2. *C. necator* is tolerant to ethanol in the growth medium

Product toxicity is an important consideration when devising a bioproduction strategy. Ideally, host organisms used for a bioproduction process should be tolerant to the desired product as well as the feeding substrate. We evaluated the tolerance of *C. necator* to ethanol by adding different concentrations of this alcohol in the range of 0-10 g/l to cultures growing in mineral medium with 10 g/l acetate as sole carbon source (Figure 6.2 A). We observed that ethanol concentrations ≥4 g/l cause a significant growth delay relative to control cultures grown in acetate-only minimal medium. Specifically, high ethanol concentrations prolonged the lag phase, reaching up to ~60

hours for cultures grown in the presence of 10 g/l ethanol (Figure 6.2B). Ethanol concentrations of ≥4 g/l also cause a significant reduction in the bacterium's growth rate (Figure 6.2 C).

The platform strain is generally tolerant to ethanol concentrations in the growth medium of up to 4g/l. Considering the kinetic profile of acetate growth described in the previous section, we propose that the maximum acetate concentration suitable for the batch bioproduction process is 10 g/l. The theoretical ethanol yield from this starting substrate concentration is 7.8g/l, meaning that the strain would be broadly tolerant to up to ~50% of the theoretical yield from this feed.

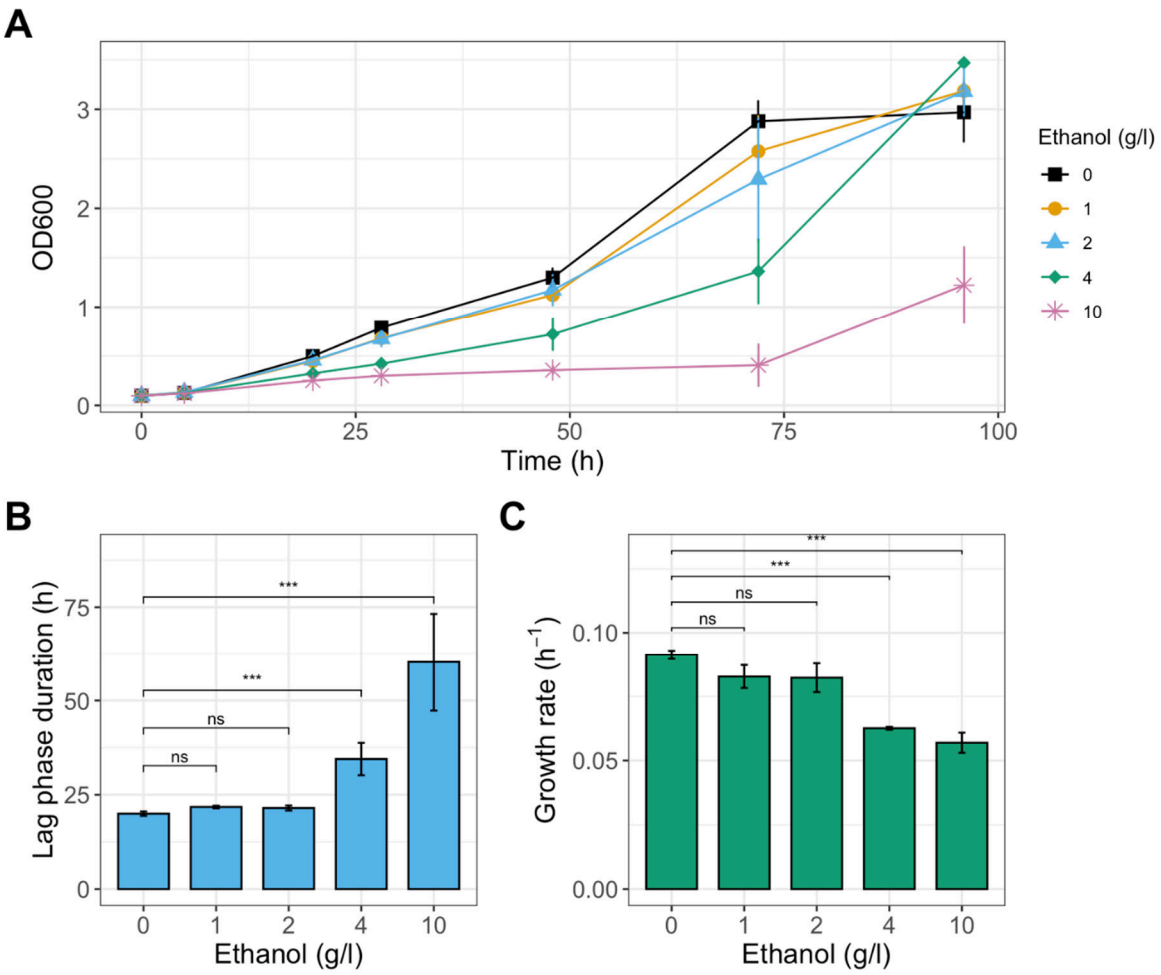


Figure 6. 2. Effect of ethanol on the acetotrophic growth of *C. necator*.

(A) Growth curves of *C. necator* cultured in the presence of 0-10 g/l ethanol. **(B)** Lag phase duration as a function of ethanol concentration. The lag phase duration was quantified as the time necessary for the bacterial culture to reach and OD = 0.5. **(C)** Growth rate of the bacterial population as a function of ethanol concentration. ns = not significant, *** = p<0.01, two-sample t-test.

6.5.3. Expression of the endogenous *acoE* gene boosts ethanol production titer, rate and yield

We used *C. necator* RHM5 as a platform strain for metabolic engineering of ethanol production. In this strain, carbon flux is diverted away from the endogenous poly-hydroxyalkanoate (PHA) production pathway via the deletion of the *phaCAB* operon on chromosome 1. The RHM5 strain does not naturally secrete organic compounds such as organic acids or alcohols. Previous studies have shown that *C. necator* RHM5 can be converted to an ethanologenic strain by constitutive heterologous expression of an alcohol dehydrogenase (AdhE) enzyme from *E. coli* under micro-anaerobic conditions[306]. Building on this, we combined plasmid-based, inducible *adhE* expression with the expression of an endogenous *acoE* gene. We hypothesised that expression of this gene would increase the cellular pool of acetyl-CoA, thereby boosting flux towards ethanol synthesis via the heterologous AdhE enzyme (Figure 6.3).

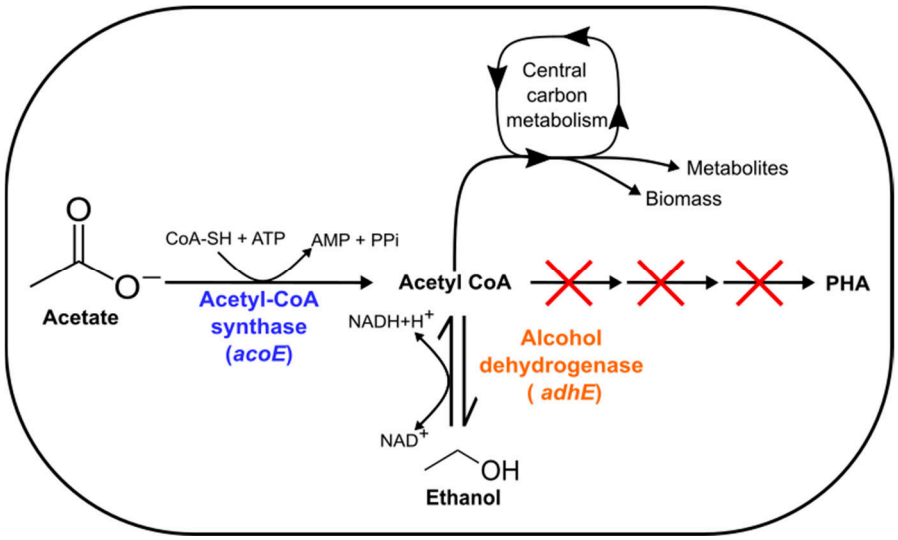


Figure 6. 3. A two-enzyme pathway for ethanol synthesis from acetate.

Acetate is converted to acetyl-CoA by the enzyme acetyl-CoA synthase (encoded by the *acoE* gene). Acetyl-CoA is then converted to ethanol by the enzyme alcohol dehydrogenase (*adhE*). Acetyl-CoA also feeds into central carbon metabolism for the production of biomass and metabolites. In the *C. necator* RHM5 platform strain, acetyl-CoA cannot be directed toward the synthesis of polyhydroxyalkanoates (PHA), as this biosynthetic pathway has been deleted from the genome.

Three distinct genetic constructs were designed to test this hypothesis, combining the inducible expression of the *adhE* gene with either constitutive or inducible expression of *acoE* (Figure 6.4A). The *araC*-P_{BAD} transcriptional regulation system was used for arabinose-inducible regulation of *adhE* and *acoE* expression. Previous studies have successfully characterised and applied this transcriptional regulation system in *C. necator*[133,221]. Moreover, the bacterium's inability to degrade the inducer L-arabinose[137] allows exposure to this compound to be prolonged throughout the cultivation time, yielding an improved heterologous expression profile over other inducible transcriptional regulation systems that were tested for the ethanol bioproduction process (Figure S6.1).

Production experiments were performed in batch culture, with cells grown in minimal medium supplemented with 10 g/l acetate as sole carbon source. Following induction

of the plasmid-encoded biosynthetic pathway, ethanol and acetate concentrations were determined by sampling the culture at regular intervals. A kinetic profile of ethanol accumulation over cultivation time is shown in Figure 6.4B. We observed that both constitutive and inducible expression of the *acoE* gene significantly boosted the ethanol production titer and rate relative to the *adhE*-only construct (Figure 6.4C-D), as well as the acetate consumption rate (Figure 6.4E). No significant difference in the acetate consumption rate was observed between strains expressing *acoE* constitutively or inducibly.

The best-performing ethanologenic strain was obtained by combining inducible *adhE* and *acoE* expression, reaching a maximum titer of 0.9 g/l ethanol, yield of 11.54%, and ethanol production rate of 2.39 mg/h (Table 6.2). No significant differences in growth rate or biomass yield were observed for constructs expressing constitutive or inducible *acoE* relative to the *adhE*-only and RHM5 control strains (Figure S6.2).

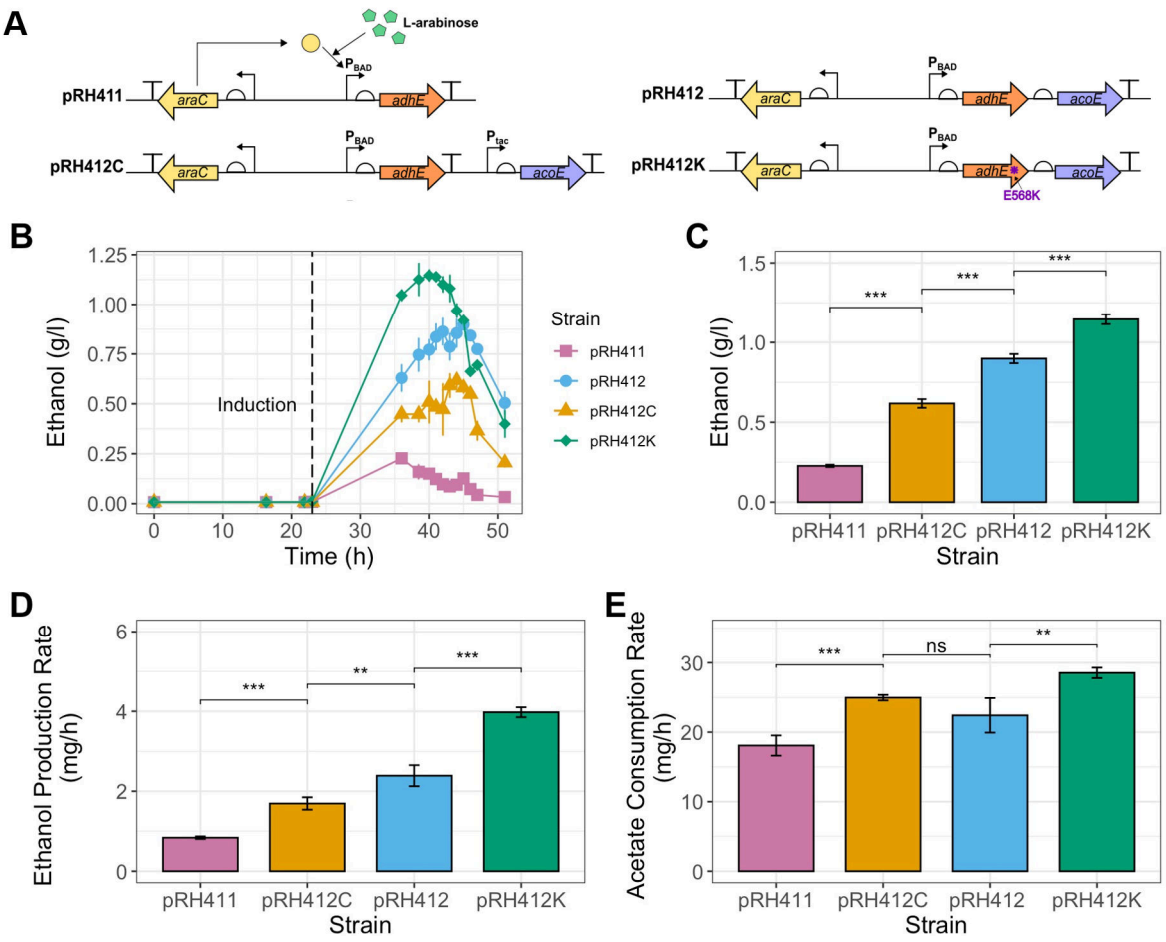


Figure 6. 4. Construction of ethanologenic strains via a two-enzyme pathway.

(A) Plasmid vectors used for ethanol production. The *adhE* gene (WT or E568K mutant) is expressed in isolation or in combination with *acoE*. For all constructs, the type of regulation applied to each gene (constitutive or L-arabinose-inducible) is indicated. **(B)** Ethanol production kinetics. The time at which expression of the biosynthetic pathway enzymes was induced is indicated by a vertical dashed line. **(C)** Maximum ethanol titer reached. **(D)** Maximum ethanol production rates. **(E)** Maximum acetate consumption rates. ns = not significant, ** = p<0.05, *** = p<0.01, two-sample t-tests.

6.5.4. AdhE variants with increased aerobic activity improve ethanol production

We hypothesised that introducing mutations known to boost aerobic AdhE activity in the strain harbouring the best-performing plasmid construct (pRH412) could lead to improved ethanol production kinetics. Several variants of the *E.* coli AdhE enzyme with increased aerobic activity have been described in previous studies[315,316]. We investigated whether one of these mutations (E568K) could boost ethanol production in the engineered strain.

The E568K mutation was introduced onto the *adhE* CDS on plasmid pRH412 via site-directed mutagenesis (Figure 6.4A). The ethanologenic *C. necator* strain harbouring the AdhE-E568K mutant displayed a significant improvement in the maximum ethanol titer (+2.8%), maximum ethanol production rate (+66.5%), and yield (+27.7%) of ethanol production relative to strains harbouring the WT enzyme sequence (Figure 6.4B-D and Table 6.2). The acetate consumption rate in this strain was also significantly boosted (Figure 6.4E).

Table 6. 2. Kinetic parameters of ethanol production in the engineered *C. necator* strains.

Plasmid	Max. ethanol titer (g/l)	Ethanol yield (%theoretical yield)	Max. ethanol production rate (mg/h)	Max. acetate consumption rate (mg/h)	Growth rate (h^{-1})	Biomass yield $Y_{x/s}$ (OD/g acetate)	Product yield $Y_{p/s}$ (mg ethanol /g acetate)
pRH411	0.23 ± 0.01	2.95	0.84 ± 0.03	18.08 ± 1.46	0.11 ± 0.01	9.89 ± 0.29	22.70 ± 0.73
pRH412C	0.62 ± 0.03	7.95	1.69 ± 0.16	24.97 ± 0.40	0.14 ± 0.01	8.79 ± 0.01	61.74 ± 2.72
pRH412	0.90 ± 0.03	11.54	2.39 ± 0.26	22.42 ± 2.49	0.15 ± 0.06	9.07 ± 0.59	89.86 ± 2.87
pRH412K	1.15 ± 0.03	14.74	3.98 ± 0.12	28.54 ± 0.75	0.13 ± 0.03	8.86 ± 0.87	114.58 ± 3.07

6.6. Discussion and Conclusions

We have provided a thorough characterisation of acetate and ethanol tolerance in *C. necator*, together with the first successful implementation of a metabolic engineering strategy for production of ethanol from acetate at the gram per liter scale.

Previous studies propose an optimal acetate concentration for cultivation of *C. necator* in the range of 5-6 g/l[313]. Similarly, our investigation of acetotrophic growth kinetics showed that, out of the acetate concentrations that were tested, supplying the starting substrate at ~6 g/l leads to maximum cell density in the shortest time. However, we sought to maximise the starting concentration of this C_2 substrate to boost ethanol production titer in the batch culture bioprocess. The main drawback of this strategy was a significantly prolonged lag phase. This was partially overcome during production experiments by inoculating the batch cultures at a higher optical density than was tested for the acetotrophic growth kinetics experiments.

We hypothesise that increasing the bacterium's acetate tolerance may prove beneficial for ethanol production, as it may enable the bacterial population to reach higher cell densities, and a higher ethanol titer, thereby simultaneously maximising waste substrate degradation and accumulation of valuable product. However, this would require strains of *C. necator* that have a higher tolerance to acetate, which could be attained by adaptive evolution, similar to the efforts described in Chapter 5. Alternatively, the ethanologenic strains obtained in this work could be applied to a fed-batch production process, where more acetate is consumed without the need for a high initial acetate concentration.

The investigation of ethanol tolerance described in this study shows that the bacterium's tolerance to the product is moderate in relation to its theoretical yield. However, it is still far behind the ethanol tolerance of conventional ethanologenic strains such as *S. cerevisiae* and *Zymomonas mobilis*[317–319]. At present, the maximum ethanol titer that can be attained during production in *C. necator* is below the bacterium's tolerance limit. Thus, its relatively low tolerance to this product is not an immediate limitation. If needed, ethanol tolerance could be improved through additional strain engineering and adaptation, as has been done for other bacterial strains and products[320].

By implementing a simple two-enzyme biosynthetic pathway for ethanol production, we engineered a proficient ethanologenic strain, reaching a maximum titer of 1.15g/l from consumption of 10 g/l of acetate, corresponding to 14.7% of the theoretical yield. This ethanologenic strain is significantly improved over a previously reported engineered strain[306], which reached a maximum titer of 0.35 g/l acetate from consumption of 12.51g/l of acetate, corresponding to 3.6% of the theoretical yield.

We demonstrated that expressing a plasmid-borne copy of the endogenous *acoE* gene constitutively or inducibly boosts ethanol production titer, rate and yield relative to strains that only express the heterologous *adhE* gene. We hypothesise that this is due to an increase in the cellular pool of acetyl-CoA. Interestingly, while the expression of *acoE* significantly increased the strain's acetate consumption rate relative to the *adhE*-only strains, it did not lead to significant improvements in growth rate or biomass yield. This would suggest that, whilst the accumulation of acetyl-CoA appears to be

rate-limiting for ethanol production, it may not be rate-limiting for growth. Moreover, we found that increasing the activity of the AdhE enzyme by introducing a mutation known to increase oxygen tolerance also boosts ethanol production. This suggests that there is ample room for improvement in the biosynthetic process, for example by screening other AdhE mutants, or variants of this enzyme from other proficient ethanologenic organisms.

An outstanding limitation of this work is that, during the latter stages of the batch production process, ethanol is consumed by *C. necator*. The bacterium can oxidise this product via endogenous and heterologous bi-directional AdhE enzymes, which provides a source of carbon and energy. This challenge could be addressed through bioprocess engineering. By implementing a fed batch or continuous culture process, the consumption of ethanol as a carbon and energy source upon acetate depletion could be minimised. Additionally, a genome-scale model for metabolic modelling of *C. necator* has recently become available[321]. Model-guided metabolic engineering strategies could therefore be implemented to boost carbon flux towards ethanol production, and away from ethanol consumption.

6.7. Chapter 6 Appendix – Supplementary Information

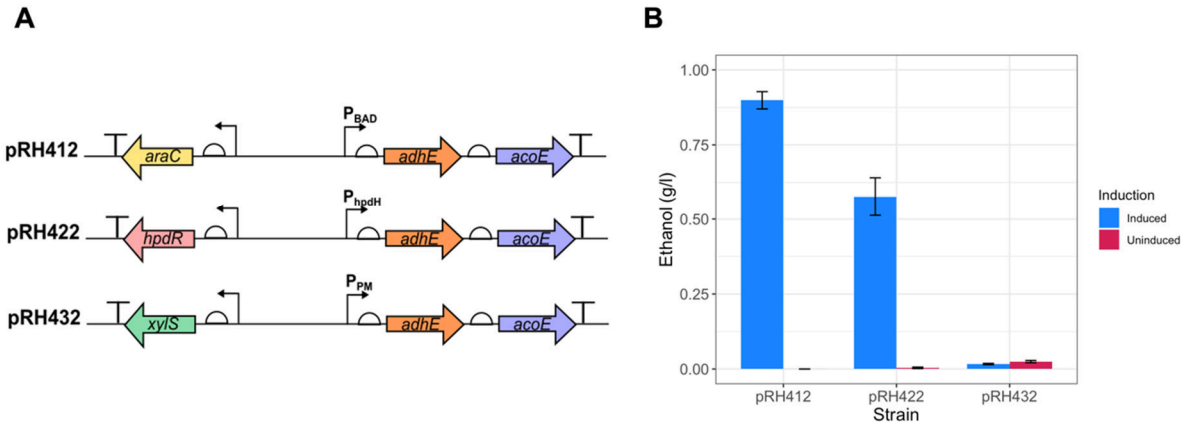


Figure S6. 1. Testing different inducible systems for ethanol production.

(A) Inducible expression of *adhE* and *acoE* was trialled with three distinct transcriptional regulators (*araC*, *hpdR*, *xylS*) and their cognate promoters. **(B)** Maximum ethanol titer obtained by ethanologenic strains harbouring constructs shown in (A), with and without the addition of the appropriate inducer.

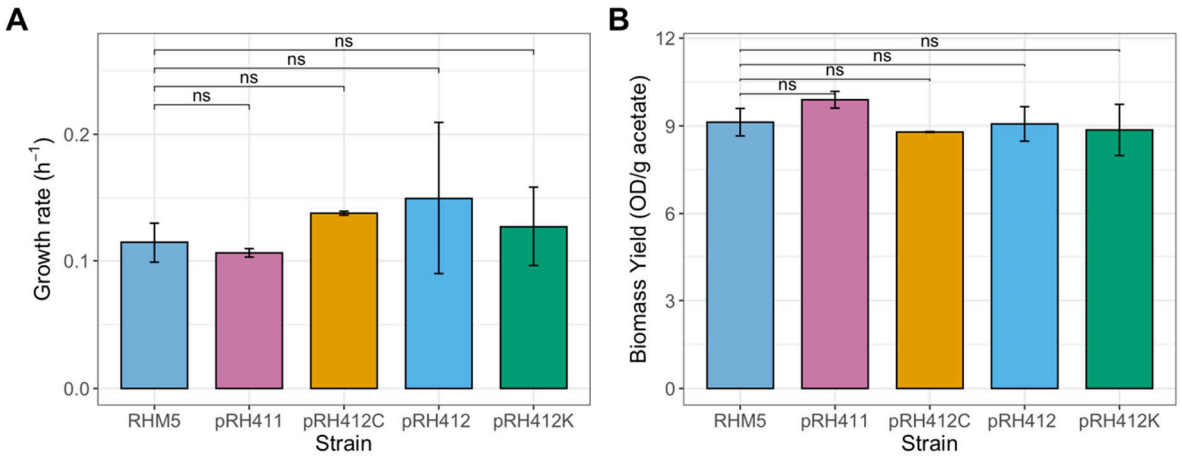


Figure S6. 2. Kinetic growth parameters in bioproduction strains.

(A) Growth rate and **(B)** biomass yield of platform strain RHM5 and ethanologenic strains. ns = not significant, two-sample t-test.

Table S6. 1. Bacterial strains used in this study.

Strain	Purpose	Genotype
E. coli MG1655	Source of adhE gene sequence	F- lambda- *ilvG*- *rfb*-50 *rph*-1
E. coli DH5α	Plasmid DNA storage and isolation	F– φ80*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *recA*1 *endA*1 *hsdR*17(rK⁻, mK⁺) *pho*A *sup*E44 λ– *thi*-1 *gyrA*96 *relA*1
C. necator H16	Source of *acoE* gene sequence	Wild-type *C. necator* strain
C. necator RHM5	Production host strain	Δ*phaCAB*

Table S6. 2. Plasmids used in this study.

The sequence of plasmids used in bioproduction experiments (pRH411, pRH412, pRH412C, and pRH412K) can be accessed via hyperlink, by clicking on the individual plasmid names.

Plasmid	Description	Source
pSEVA231	pBBR1, KanR, oriT,	[40]
pLO11	Source of arabinose inducible promoter system	[41]
pCK227	Source of toluic acid-inducible promoter system	[42]
pRH411	pBBR1, KanR, oriT, *araC*, ParaBAD, MG1655 *adhE*	This study
pRH412	pBBR1, KanR, oriT, *araC*, ParaBAD, MG1655 *adhE*, H16 *acoE*	This study
pRH422	pBBR1, KanR, oriT, *xylS*, PPM, MG1655 *adhE*, H16 *acoE*	This study
pRH432	pBBR1, KanR, oriT, PpHpdR, PPpHpdH, MG1655 *adhE*, H16 *acoE*	This study
pRH442	pBBR1, KanR, oriT, Plac, MG1655 *adhE*, H16 *acoE*	This study
pRH452	pBBR1, KanR, oriT, Ptac, MG1655 *adhE*, H16 *acoE*	This study
pRH412C	pBBR1, KanR, oriT, araC, ParaBAD, MG1655 *adhE*, Ptac, H16 *acoE*	This study
pRH412K	pBBR1, KanR, oriT, araC, ParaBAD, MG1655 *adhE*-E568K, H16 *acoE*	This study

Table S6. 3. Oligonucleotides used for plasmid assembly.

Gene-specific binding sites are identified by capital letters.

Oligo name	Oligo sequence	Purpose
231BB_For	TCTAGAGGATCCCCGGGTAC	Amplification of pSEVA231 backbone
231BB_Rev	CCTGCAGGCATGCAAGC	
41XaraC_For	aagcttgcatgcctgcaggTTATGACAACTTGACGGC	Amplification of *araC*-ParaBAD (pRH41X series)
41XaraC_Rev	gtcctcctttgggggggcccATGGAGAAACAGTAGAGAG	
42XPM_For	aagcttgcatgcctgcaggTCAAGCCACTTCCTTTTTG	Amplification of *xylS*-Ppm (pRH42X series)
42XPM_Rev	gtcctcctttgggggggcccTTGCATAAAGCCTAAGGG	
43XPHPD_For	aagcttgcatgcctgcaggCTACTCGGCTAGCAACTCGC	Amplification of *hpdR*-PhpdH (pRH43X series)
43XPHPD_Rev	gtcctcctttgggggggcccGTGCAACCTCGCGCCTGT	
4XXadhE_For	gggcccccccaaaggaggacaaccATGGCTGTTACTAATG TC	Amplification of *adhE* from *E. coli* MG1655
4X1adhE_Rev	gtacccggggatcctctagaTTAAGCGGATTTTTTCGC	
4X2adhE_Rev	tctttaatttaggatccccTTAAGCGGATTTTTTCGC	
acs_For	gggggatcctaaattaaagaggagaaaacagctATGTCCGCC ATCGAATCGGTG	Amplification of *acoE* from *C. necator* H16
acs_Rev	gtacccggggatcctctagataatTCACTGCGCCTGCTTGA GCTGCT	
Ptac_412C_Fw	ccgaaggtgagccagtgtgattgacaattaatcatcggctcg	Amplification of Ptac fragment (pRH412C)
Ptac_412C_Rv	ACCGATTCGATGGCGGACATggttgtcctcctttgggg	
412C_BB_Fw	ATGTCCGCCATCGAATCG	Amplification of *araC*-ParaBAD-*adhE-acoE* backbone (pRH412C)
412C_BB_Rv	tcacactggctcaccttc	
E568K_Fw	CTGGCGCTGCGCTTT	Site-directed mutagenesis to introduce the E568K mutation in *adhE* (pRH412K)
E568K_Rv	TTCGAAGTGAGTTTCCGGATG	

7. General discussion

7.1. Summary of research contributions, limitations, and future directions

As discussed in Chapter 1 (Section 1.3.1), there is an urgent need to shift towards more sustainable feedstocks in industrial biomanufacturing. C_1 and C_2 compounds, which can be derived from waste carbon streams, are particularly attractive substrates to achieve sustainability goals. To harness these feedstocks for industrial bioprocesses, it is necessary to build microbial cell factories that are capable of their efficient assimilation and bioconversion, i.e., that possess specialised metabolic capabilities. Whilst it is feasible to install the required metabolic characteristics into traditional industrial host strains like *E. coli* or *S. cerevisiae* (see Section 1.4.4), using non-model microbes that can naturally process renewable feedstocks as platform strains is an increasingly acknowledged alternative.

Naturally occurring bacterial species have evolved over billions of years on Earth, leading to an astonishing present-day variety of genomes and metabolisms, which constitutes a vast source of potential for biotechnology. However, leveraging this diversity presents numerous challenges. The challenges are rooted in the fact that, for any given bacterium, the same genotypic and phenotypic traits that guarantee their success in an environmental context often make them maladapted to life on a laboratory bench, or in an industrial bioreactor. This is further compounded by the

limited availability of molecular tools and biological knowledge of relevance to non-model bacteria. Though the past ~50 years have witnessed momentous developments in molecular biology, synthetic biology, systems biology, and metabolic engineering (see Section 1.3.2), most advances cannot be readily transferred from model strains (in which they are almost invariably developed) to industrially-relevant non-model bacteria. Considering this, domesticating non-model bacteria for biotechnological applications requires extensive (re-)development of advanced tools and methods, as well as the implementation of bioengineering strategies that can both uncover and improve their desirable metabolic traits towards sustainable bioproduction goals. This is the context within which the research presented in this thesis was conducted.

Specifically, we focused our efforts on the development of the facultative chemolithoautotrophic bacterium *Cupriavidus necator* H16, which holds vast potential for sustainable biomanufacturing and has been the subject of promising scientific advancements in recent years (see Section 1.4.3). As stated in Section 1.1, the key objectives of our research were to (i) develop tailored tools and methods for advanced genetic manipulation, (ii) engineer strains with improved growth on formate (C_1 substrate), and (ii) demonstrate ethanol production from acetate (C_2 substrate). Below, the fulfilment of these aims across the five research chapters (Chapters 2-6) is assessed, identifying avenues for future research.

7.1.1. Improved transformation of *C. necator*

In Part 1 of the thesis, we reported the development of tailored tools and methods for advanced genetic manipulation of *C. necator*, starting with efficient DNA delivery (Chapter 2). Namely. we described the optimisation of a novel electroporation protocol which can be used to transform plasmid DNA into wildtype *C. necator* with high efficiency. Using this method, we could reproducibly deliver transformation efficiencies in the range of 10^8 CFU/µg DNA. This is a ~10-fold improvement over previously reported values. The protocol is also quicker than other options available in the literature. By using saturated cultures for competent cell preparation, the hands-on time for the protocol is reduced, as cells are cultured overnight. Following this, the preparation of competent cells requires at most one hour, and electroporation can be conducted in as little as ~ 10 minutes if omitting recovery time (we show that this can still deliver high transformation efficiency in the range of ~10^6 CFU/ µg DNA in Section 2.4.3). Importantly, we implement a new buffer composition that includes a cryoprotectant (glycerol). This means that competent cells can be easily stored in a -80 ºC freezer and retrieved when needed, such that fresh cell aliquots do not have to be prepared prior to each transformation.

When the protocol was originally developed, the highest transformation efficiency reported in the literature was ~4x10^5 CFU/µg DNA using plasmid pBHR1[179]. There was no record in the literature of plasmids from the SEVA suite being efficiently transformed in *C. necator*, which is a key result from our analysis. In 2021, Azubuike *et al.* reported the development of the pCAT suite of plasmids, which were optimised for use in *C. necator*[182]. This study was contemporaneous to the DPhil project, having

been published once the bulk of our protocol optimisation had already been finalised. In their paper, Azubuike *et al.* demonstrated that pCAT vectors can be transformed with efficiencies in the range of ~10^5-10^7. The authors proposed that, at least for some plasmids, the previously reported low transformation efficiency was likely due to a suboptimal kanamycin-resistance cassette being used. Once this finding was published, we extended our study to test the performance of the optimised electroporation protocol with these engineered vectors. We reported superior transformation efficiency values for all tested pCAT plasmids, and demonstrate that plasmids from pCAT and SEVA suites can be used interchangeably, as they are transformed with similar efficiencies (Section 2.4.7). This is an important consideration. Though it is useful to have a dedicated suite of pCAT vectors which can specifically be used in *C. necator*, many available molecular tools and methods rely on the use of SEVA vectors (i.e. are derived from SEVA backbone plasmids or components thereof). This is the case, for instance, for the SIBR-Cas plasmids[209], which were instrumental to the development of a genome editing tool (Chapter 3, discussed below). In the future, it would be interesting to characterise the performance of this protocol with a wider range of plasmids, providing a more direct comparison with previous literature reports. Plasmids with a range of properties (sizes and sequences) could also be tested to determine the utility of the protocol for different applications. Additionally, by undertaking one-factor-at-a-time optimisation, we limited the combinatorial exploration of the parameter space. For any given protocol parameter, the optimal value we identified was likely (or at least partially) influenced by the choice of other values in the protocol. It may be possible to attain further improvements in transformation efficiency, for instance through a design of experiments approach to multi-factorial optimisation. Despite these limitations, this

method and its characterisation will be useful for the expanding community of researchers working with *C. necator*, and we expect that further optimisation will be conducted through a distributed learning by doing approach.

7.1.2. Advanced tools for genome editing and the regulation of gene expression

The electroporation method optimised in Chapter 2 greatly accelerates the prototyping of synthetic DNA constructs, a key development which supported much of the research for this thesis, particularly the advances reported in Chapters 3 and 4. Further to the dearth of reliable and efficient methods for plasmid DNA delivery, another factor which previously made it difficult to engineer *C. necator* was the limited availability of genome editing tools. In Chapter 3, we addressed this limitation by tailoring the SIBR-Cas plasmids and protocols for use in this non-model bacterium. Our motivation for focusing on this genome editing tool over its many alternatives was that the SIBR-Cas plasmids were designed to be portable, and their functionality had been previously demonstrated in wildtype bacterial strains other than *E. coli*. The method uses an inducible self-splicing intron (SIBR intron) to tightly control the expression of a CRISPR-Cas12a system, which can be used for counterselection of the wild-type strain following homologous recombination. The key advantage of this method is that, by precisely regulating the expression of a Cas nuclease, it enables distinct temporal separation of homologous recombination and counterselection. This is an important consideration for many wildtype organisms that have inefficient homologous recombination mechanisms.

During our investigation, however, we discovered that the SIBR intron could not tightly regulate the expression of a Cas12a nuclease in *C. necator* (Section 3.5.3), though the regulatory module was functional with Cas9. Through bioinformatic and experimental analysis, we identified alternative translation initiation from within the SIBR intron sequence as the likely cause for this defective regulation. These results highlight the difficulties in tailoring molecular tools for use in genomic contexts that are different to the one in which the tools were originally developed.

We successfully circumvented this limitation by developing a new iteration of the SIBR system, SIBR2.0, where the SIBR intron is transferred from its canonical location (immediately following the start codon) to the inner CDS of the Cas12a expression cassette. SIBR2.0 enabled tightly inducible Cas12a expression and was subsequently applied to mediate genome editing (specifically a gene knockout at the *acoC* locus). Though further optimisation and characterisation of the editing protocol would be beneficial, the method we describe enables high editing efficiencies (~70%) and can be executed in a single working day (Section 3.5.5).

Taken together, the results outlined in Chapter 3 deliver three significant contributions. Firstly, it is now possible to generate *C. necator* knockout strains rapidly and reliably, which will support strain engineering in this organism. Many advanced metabolic engineering strategies, such as growth-coupling and modular engineering approaches (see Section 1.4.1), require knocking out several genomic loci in other to rewire cellular metabolism towards bioproduction goals. Using the SIBR-Cas toolkit, these designs can be implemented on a reduced timescale, expediting the DBTL optimisation pipeline.

Secondly, we demonstrate that the SIBR and SIBR2.0 systems can be used to tightly control the expression of two distinct Cas nuclease, Cas9 and Cas12a, though the former has not yet been used to mediate genome editing (this work is currently in progress). These results highlight the modular nature of the SIBR genetic control system. The SIBR-Cas method could easily be modified to operate with other Cas nucleases and implement an arsenal of CRISPR-Cas-based tools in *C. necator* (such as those described in Section 1.3.2). These tools can not only be used to implement rational metabolic designs, but also to gain fundamental biological insights which are relevant to strain engineering. For instance, CRISPRi and CRISPRa tools can be used to conduct genome-wide screens and reveal genotype-phenotype relationships underlying particular metabolic traits (see ref. [322] for a recent example).

Finally, we highlight that the SIBR2.0 system for translation-level control of gene expression could be applied to implement more complex regulatory architectures in *C. necator* (such as genetic logic gates) for any gene(s) of interest. In Chapter 4, we explore this possibility by combining the Jungle Express repressor system with the SIBR2.0 intron, to implement a genetic AND logic gate which regulates gene expression at both the transcription and translation level.

7.1.3. Improved assimilation and bioconversion of renewable feedstocks

In Part 2 of the thesis, we turned our attention to engineering *C. necator* as a catalyst for the valorisation of C_1 (formate) and C_2 (acetate) feedstocks. The identification of

evolved isolates with improved growth on formate was a key finding reported in Chapter 5. Though their phenotypic and genotypic characterisation was limited (see Section 5.5), upon further analyses these strains could be applied to deliver improved C_1 bioprocesses. It should be noted that, since these strains were not subject to genetic manipulation but were instead isolated by random mutagenesis and continuous cultivation, they face fewer regulatory hurdles for their implementation in industrial bioprocesses. For instance, these *C. necator* isolates could be applied to produce naturally-accumulated PHAs from formate. The biomass of these *C. necator* strains is also valuable, as it could be used as a source of single cell protein, enabling the production of food and feed from C_1 compounds.

Further, the insights gained from this ALE experiment can be used to inform more targeted metabolic engineering. For instance, we observed that top-performing isolates shared a common mutation in the *rpoD* gene, which encodes an RNA polymerase sigma subunit. Though we did not conduct transcriptomic analysis for these strains, we hypothesise that this mutation may have led to global changes in gene expression patterns, some of which may be particularly advantageous for formatotrophic growth. In future work, transcriptomic characterisation of these isolates could shed light on the effects of this mutation and their relationship to the improved formatotrophic growth kinetics, revealing key genomic targets for up- or down-regulation.

In Chapter 6, we explored the valorisation of C_2 substrates (specifically acetate). Relative to C_1 compounds, substrates such as acetate are more palatable to production host strains such as *C. necator*, making them attractive feedstocks for more

immediate application in industrial bioprocesses (Section 6.3). In our study, we provide a thorough characterisation of acetate tolerance in *C. necator*, and demonstrate how a two-enzyme biosynthetic pathway can mediate the conversion of acetate into ethanol. We optimised the architecture of this pathway, combining the expression of endogenous and heterologous genes, and modulating their regulation. The final product was an engineered strain that could produce ethanol at the g/l scale.

Though we achieved significant improvements in bioproduction parameters over previous reports, the production of ethanol from acetate is not likely to be a valuable biosynthetic process. Ethanol is currently produced commercially from plant biomass (see Section 1.3.1). Additionally, it can be produced autotrophically by acetogenic hosts through syngas fermentation[323]. The production strains we report are therefore unlikely to attain industrial-scale commercial success. Nevertheless, the insights gained from this work can be extrapolated to other C_2 bioproduction processes, and are therefore broadly useful to further the valorisation of C_2 substrates. For instance, we demonstrated how the expression of an endogenous gene (*acoE*) can boost carbon flux towards the accumulation of ethanol. This could be extrapolated for the bioconversion of acetate into other (more valuable) products, both in *C. necator* and other acetotrophic hosts.

7.2. Concluding remarks

In recent years, great strides have been made towards engineering non-model microbes as platform strains to mediate the bioconversion of renewable substrates into value-added products. This represents a promising avenue to transition away from

the unsustainable consumption of both fossil resources and plant biomass as feedstocks for (bio)manufacturing. The wealth of tools, methods, and biological insights documented in this thesis are expected to contribute significantly towards this shift, specifically by advancing the use of *C. necator* as a domesticated platform strain for sustainable bioprocesses.

Bibliography

1. Aftalion, F. *A History of the International Chemical Industry*. (Chemical Heritage Foundation, 2001).
2. Clomburg, J. M., Crumbley, A. M. & Gonzalez, R. Industrial biomanufacturing: The future of chemical production. *Science* **355**, aag0804 (2017).
3. Ruchala, J., Kurylenko, O. O., Dmytruk, K. V. & Sibirny, A. A. Construction of advanced producers of first- and second-generation ethanol in Saccharomyces cerevisiae and selected species of non-conventional yeasts (Scheffersomyces stipitis, Ogataea polymorpha). *J Ind Microbiol Biotechnol* **47**, 109–132 (2020).
4. Dien, B. S., Cotta, M. A. & Jeffries, T. W. Bacteria engineered for fuel ethanol production: Current status. *Applied Microbiology and Biotechnology* **63**, 258–266 (2003).
5. Alternative Fuels Data Center: Ethanol. https://afdc.energy.gov/fuels/ethanol.html.
6. Otero, J. M., Panagiotou, G. & Olsson, L. Fueling industrial biotechnology growth with bioethanol. *Advances in Biochemical Engineering/Biotechnology* **108**, 1–40 (2007).
7. Naik, S. N., Goud, V. V., Rout, P. K. & Dalai, A. K. Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews* **14**, 578–597 (2010).
8. Zhu, X.-G., Long, S. P. & Ort, D. R. What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Current Opinion in Biotechnology* **19**, 153–159 (2008).
9. Claassens, N. J., Sousa, D. Z., dos Santos, V. A. P. M., de Vos, W. M. & van der Oost, J. Harnessing the power of microbial autotrophy. *Nature Reviews Microbiology* **14**, 692–706 (2016).
10. Cotton, C. A. R. *et al.* Photosynthetic constraints on fuel from microbes. *Frontiers in Bioengineering and Biotechnology* **3**, (2015).
11. Blankenship, R. E. *et al.* Comparing Photosynthetic and Photovoltaic Efficiencies and Recognizing the Potential for Improvement. *Science* **332**, 805–809 (2011).
12. Wendisch, V. F. *et al.* The flexible feedstock concept in Industrial Biotechnology: Metabolic engineering of Escherichia coli, Corynebacterium glutamicum, Pseudomonas, Bacillus and yeast strains for access to alternative carbon sources. *Journal of Biotechnology* **234**, 139–157 (2016).
13. Scown, C. D. & Keasling, J. D. Sustainable manufacturing with synthetic biology. *Nat Biotechnol* **40**, 304–307 (2022).
14. Satanowski, A. & Bar-Even, A. A one-carbon path for fixing CO2. *EMBO reports* **21**, e50273 (2020).
15. Wittmann, C. & Liao, J. C. *Industrial Biotechnology: Microorganisms*. (John Wiley & Sons, 2017).
16. United States. & States, U. *Impacts of applied genetics : micro-organisms, plants, and animals*. 1–360 (Congress of the U.S., Office of Technology Assessment, 1981). doi:10.5962/bhl.title.118897.
17. Quianzon, C. C. & Cheikh, I. History of insulin. *J Community Hosp Intern Med Perspect* **2**, 10.3402/jchimp.v2i2.18701 (2012).
18. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat Rev Microbiol* **12**, 381–390 (2014).

19. Meng, F. & Ellis, T. The second decade of synthetic biology: 2010–2020. *Nat Commun* **11**, 5174 (2020).
20. Garner, K. L. Principles of synthetic biology. *Essays Biochem* **65**, 791–811 (2021).
21. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nat Rev Genet* **11**, 367–379 (2010).
22. Müller, K. M. & Arndt, K. M. Standardization in synthetic biology. *Methods Mol Biol* **813**, 23–43 (2012).
23. Galdzicki, M. *et al.* The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat Biotechnol* **32**, 545–550 (2014).
24. Baig, H. *et al.* Synthetic biology open language visual (SBOL visual) version 3.0. *J Integr Bioinform* **18**, 20210013 (2021).
25. Engler, C. & Marillonnet, S. Golden Gate cloning. *Methods Mol Biol* **1116**, 119–131 (2014).
26. Werner, S., Engler, C., Weber, E., Gruetzner, R. & Marillonnet, S. Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioeng Bugs* **3**, 38–43 (2012).
27. Iverson, S. V., Haddock, T. L., Beal, J. & Densmore, D. M. CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synth. Biol.* **5**, 99–103 (2016).
28. Taylor, G. M., Mordaka, P. M. & Heap, J. T. Start-Stop Assembly: a functionally scarless DNA assembly system optimized for metabolic engineering. *Nucleic Acids Res* **47**, e17 (2019).
29. Kamens, J. The Addgene repository: an international nonprofit plasmid and data resource. *Nucleic Acids Res* **43**, D1152-1157 (2015).
30. Silva-Rocha, R. & de Lorenzo, V. Mining logic gates in prokaryotic transcriptional regulation networks. *FEBS Letters* **582**, 1237–1244 (2008).
31. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249–253 (2012).
32. Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519 (2008).
33. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339–342 (2000).
34. Mishra, D. *et al.* An engineered protein-phosphorylation toggle network with implications for endogenous network discovery. *Science* **373**, eaav0780 (2021).
35. Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463**, 326–330 (2010).
36. Kwok, R. Five hard truths for synthetic biology. *Nature* **463**, 288–290 (2010).
37. Nicholson, D. J. Is the cell really a machine? *Journal of Theoretical Biology* **477**, 108–126 (2019).
38. Lou, C., Stanton, B., Chen, Y.-J., Munsky, B. & Voigt, C. A. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat Biotechnol* **30**, 1137–1142 (2012).
39. Carr, S. B., Beal, J. & Densmore, D. M. Reducing DNA context dependence in bacterial promoters. *PLOS ONE* **12**, e0176013 (2017).
40. Davidsohn, N. *et al.* Accurate Predictions of Genetic Circuit Behavior from Part Characterization and Modular Composition. *ACS Synth. Biol.* **4**, 673–681 (2015).

41. Morey, K. J. *et al.* Crosstalk between endogenous and synthetic components--synthetic signaling meets endogenous components. *Biotechnol J* **7**, 846–855 (2012).
42. Zhang, Y., Ding, W., Wang, Z., Zhao, H. & Shi, S. Development of Host-Orthogonal Genetic Systems for Synthetic Biology. *Advanced Biology* **5**, 2000252 (2021).
43. Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F. & Voigt, C. A. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res* **40**, 8773–8781 (2012).
44. Shis, D. L. & Bennett, M. R. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proc Natl Acad Sci U S A* **110**, 5028–5033 (2013).
45. Boo, A., Ellis, T. & Stan, G.-B. Host-aware synthetic biology. *Current Opinion in Systems Biology* **14**, 66–72 (2019).
46. Keating, K. W. & Young, E. M. Systematic Part Transfer by Extending a Modular Toolkit to Diverse Bacteria. *ACS Synth. Biol.* **12**, 2061–2072 (2023).
47. Riley, L. A. & Guss, A. M. Approaches to genetic tool development for rapid domestication of non-model microorganisms. *Biotechnology for Biofuels* **14**, 30 (2021).
48. Kitano, S., Lin, C., Foo, J. L. & Chang, M. W. Synthetic biology: Learning the way toward high-precision biological design. *PLOS Biology* **21**, e3002116 (2023).
49. Boyle, P. M. & Silver, P. A. Parts plus pipes: Synthetic biology approaches to metabolic engineering. *Metabolic Engineering* **14**, 223–232 (2012).
50. García-Granados, R., Lerma-Escalera, J. A. & Morones-Ramírez, J. R. Metabolic Engineering and Synthetic Biology: Synergies, Future, and Challenges. *Frontiers in Bioengineering and Biotechnology* **7**, (2019).
51. Jeschek, M., Gerngross, D. & Panke, S. Combinatorial pathway optimization for streamlined metabolic engineering. *Current Opinion in Biotechnology* **47**, 142–151 (2017).
52. Naseri, G. & Koffas, M. A. G. Application of combinatorial optimization strategies in synthetic biology. *Nature Communications* **11**, 2446 (2020).
53. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**, 946–950 (2009).
54. Hossain, A. *et al.* Automated design of thousands of nonrepetitive parts for engineering stable genetic systems. *Nat Biotechnol* **38**, 1466–1475 (2020).
55. LaFleur, T. L., Hossain, A. & Salis, H. M. Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria. *Nat Commun* **13**, 5159 (2022).
56. Espah Borujeni, A., Channarasappa, A. S. & Salis, H. M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research* **42**, 2646–2659 (2014).
57. Bordbar, A., Monk, J. M., King, Z. A. & Palsson, B. O. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet* **15**, 107–120 (2014).
58. Simeonidis, E. & Price, N. D. Genome-scale modeling for metabolic engineering. *Journal of Industrial Microbiology and Biotechnology* **42**, 327–338 (2015).
59. Burgard, A. P., Pharkya, P. & Maranas, C. D. OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. *Biotechnology and Bioengineering* **84**, 647–657 (2003).

60. Ranganathan, S., Suthers, P. F. & Maranas, C. D. OptForce: An Optimization Procedure for Identifying All Genetic Manipulations Leading to Targeted Overproductions. *PLoS Computational Biology* **6**, e1000744 (2010).
61. Pharkya, P. & Maranas, C. D. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metabolic Engineering* **8**, 1–13 (2006).
62. Pharkya, P., Burgard, A. P. & Maranas, C. D. OptStrain: A computational framework for redesign of microbial production systems. *Genome Research* **14**, 2367–2376 (2004).
63. van der Oost, J. & Patinios, C. The genome editing revolution. *Trends in Biotechnology* **41**, 396–409 (2023).
64. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* **169**, 5429–5433 (1987).
65. Mojica, F. J., Juez, G. & Rodríguez-Valera, F. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol Microbiol* **9**, 613–621 (1993).
66. van Belkum, A., Scherer, S., van Alphen, L. & Verbrugh, H. Short-Sequence DNA Repeats in Prokaryotic Genomes. *Microbiology and Molecular Biology Reviews* **62**, 275–293 (1998).
67. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**, 174–182 (2005).
68. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading)* **151**, 2551–2561 (2005).
69. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315**, 1709–1712 (2007).
70. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nat Rev Microbiol* **13**, 722–736 (2015).
71. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
72. Gleditzsch, D. *et al.* PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol* **16**, 504–517 (2018).
73. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
74. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR–Cas technologies and applications. *Nat Rev Mol Cell Biol* **20**, 490–507 (2019).
75. Su, T. *et al.* A CRISPR-Cas9 Assisted Non-Homologous End-Joining Strategy for One-step Engineering of Bacterial Genome. *Sci Rep* **6**, 37895 (2016).
76. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* **31**, 233–239 (2013).
77. Chung, M.-E. *et al.* Enhanced integration of large DNA into E. coli chromosome by CRISPR/Cas9. *Biotechnol Bioeng* **114**, 172–183 (2017).
78. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).

79. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an RNA-guided DNA binding protein Supplementary Materials. *Nucleic Acids Research* **41**, 7429–7437 (2013).
80. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
81. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
82. Tong, Y., Jørgensen, T. S., Whitford, C. M., Weber, T. & Lee, S. Y. A versatile genetic engineering toolkit for E. coli based on CRISPR-prime editing. *Nat Commun* **12**, 5206 (2021).
83. Volke, D. C., Orsi, E. & Nikel, P. I. Emergent CRISPR–Cas-based technologies for engineering non-model bacteria. *Current Opinion in Microbiology* **75**, 102353 (2023).
84. Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat Methods* **12**, 415–418 (2015).
85. Renda, B. A., Hammerling, M. J. & Barrick, J. E. Engineering reduced evolutionary potential for synthetic biology. *Mol. BioSyst.* **10**, 1668–1678 (2014).
86. Vapnek, D., Alton, N. K., Bassett, C. L. & Kushner, S. R. Amplification in Escherichia coli of enzymes involved in genetic recombination: construction of hybrid ColE1 plasmids carrying the structural gene for exonuclease I. *Proc Natl Acad Sci U S A* **73**, 3492–3496 (1976).
87. Pósfai, G. *et al.* Emergent properties of reduced-genome Escherichia coli. *Science* **312**, 1044–1046 (2006).
88. Sleight, S. C., Bartley, B. A., Lieviant, J. A. & Sauro, H. M. Designing and engineering evolutionary robust genetic circuits. *J Biol Eng* **4**, 12 (2010).
89. Dragosits, M. & Mattanovich, D. Adaptive laboratory evolution - principles and applications for biotechnology. *Microbial Cell Factories* **12**, 64 (2013).
90. Portnoy, V. A., Bezdan, D. & Zengler, K. Adaptive laboratory evolution—harnessing the power of biology for metabolic engineering. *Current Opinion in Biotechnology* **22**, 590–594 (2011).
91. Sandberg, T. E., Salazar, M. J., Weng, L. L., Palsson, B. O. & Feist, A. M. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metabolic Engineering* **56**, 1–16 (2019).
92. Calvey, C. H. *et al.* Improving growth of Cupriavidus necator H16 on formate using adaptive laboratory evolution-informed engineering. *Metabolic Engineering* **75**, 78–90 (2023).
93. Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angew Chem Int Ed Engl* **57**, 4143–4148 (2018).
94. Haseltine, E. L. & Arnold, F. H. Synthetic gene circuits: design with directed evolution. *Annu Rev Biophys Biomol Struct* **36**, 1–19 (2007).
95. Bloom, J. D. & Arnold, F. H. In the light of directed evolution: Pathways of adaptive protein evolution. *Proceedings of the National Academy of Sciences* **106**, 9995–10000 (2009).
96. Latif, H., Zeidan, A. A., Nielsen, A. T. & Zengler, K. Trash to treasure: production of biofuels and commodity chemicals via syngas fermenting microorganisms. *Current Opinion in Biotechnology* **27**, 79–87 (2014).
97. Liew, F. *et al.* Gas Fermentation—A Flexible Platform for Commercial Scale Production of Low-Carbon-Fuels and Chemicals from Waste and Renewable Feedstocks. *Frontiers in Microbiology* **7**, (2016).

98. Bengelsdorf, F. R. *et al.* Bacterial Anaerobic Synthesis Gas (Syngas) and CO2+H2 Fermentation. in *Advances in Applied Microbiology* (eds. Sariaslani, S. & Gadd, G. M.) vol. 103 143–221 (Academic Press, 2018).
99. Marcellin, E. *et al.* Low carbon fuels and commodity chemicals from waste gases – systematic approach to understand energy metabolism in a model acetogen. *Green Chem.* **18**, 3020–3028 (2016).
100. Cotton, C. A., Edlich-Muth, C. & Bar-Even, A. Reinforcing carbon fixation: CO2 reduction replacing and supporting carboxylation. *Curr Opin Biotechnol* **49**, 49–56 (2018).
101. Song, D. *et al.* Electrochemical CO2 reduction catalyzed by organic/inorganic hybrids. *eScience* **3**, 100097 (2023).
102. Nam, D.-H. *et al.* Molecular enhancement of heterogeneous CO2 reduction. *Nat. Mater.* **19**, 266–276 (2020).
103. Claassens, N. J., Sánchez-Andrea, I., Sousa, D. Z. & Bar-Even, A. Towards sustainable feedstocks: A guide to electron donors for microbial carbon fixation. *Current Opinion in Biotechnology* (2018) doi:10.1016/j.copbio.2018.01.019.
104. Claassens, N. J., Cotton, C. A. R., Kopljar, D. & Bar-Even, A. Making quantitative sense of electromicrobial production. *Nature Catalysis* **2**, 437–447 (2019).
105. Stöckl, M., Claassens, N., Lindner, S., Klemm, E. & Holtmann, D. Coupling electrochemical CO2 reduction to microbial product generation – identification of the gaps and opportunities. *Current Opinion in Biotechnology* **74**, 154–163 (2022).
106. Jiang, W. *et al.* Metabolic engineering strategies to enable microbial utilization of C1 feedstocks. *Nat Chem Biol* **17**, 845–855 (2021).
107. Bar-Even, A., Noor, E., Flamholz, A. & Milo, R. Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. *Biochimica et Biophysica Acta - Bioenergetics* **1827**, 1039–1047 (2013).
108. Bar-Even, A., Flamholz, A., Noor, E. & Milo, R. Thermodynamic constraints shape the structure of carbon fixation pathways. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1817**, 1646–1659 (2012).
109. Bar-Even, A., Noor, E. & Milo, R. A survey of carbon fixation pathways through a quantitative lens. *Journal of Experimental Botany* **63**, 2325–2342 (2012).
110. Wenk, S., Yishai, O., Lindner, S. N. & Bar-Even, A. An Engineering Approach for Rewiring Microbial Metabolism. in *Methods in Enzymology* (ed. Scrutton, N.) vol. 608 329–367 (Academic Press, 2018).
111. Claassens, N. J. *et al.* Replacing the Calvin cycle with the reductive glycine pathway in Cupriavidus necator. *Metabolic Engineering* **62**, 30–41 (2020).
112. Dronsella, B. *et al.* Engineered synthetic one-carbon fixation exceeds yield of the Calvin Cycle. 2022.10.19.512895 Preprint at https://doi.org/10.1101/2022.10.19.512895 (2022).
113. Kim, S. *et al.* Growth of E. coli on formate and methanol via the reductive glycine pathway. *Nat Chem Biol* **16**, 538–545 (2020).
114. Turlin, J., Dronsella, B., De Maria, A., Lindner, S. N. & Nikel, P. I. Integrated rational and evolutionary engineering of genome-reduced Pseudomonas putida strains promotes synthetic formate assimilation. *Metabolic Engineering* **74**, 191–205 (2022).
115. Bruinsma, L., Wenk, S., Claassens, N. J. & Martins dos Santos, V. A. P. Paving the way for synthetic C1 - Metabolism in Pseudomonas putida through the reductive glycine pathway. *Metabolic Engineering* **76**, 215–224 (2023).

116. Toya, Y. *et al.* Optogenetic reprogramming of carbon metabolism using light-powering microbial proton pump systems. *Metabolic Engineering* **72**, 227–236 (2022).
117. Davison, P. A. *et al.* Engineering a Rhodopsin-Based Photo-Electrosynthetic System in Bacteria for CO 2 Fixation. *Cite This: ACS Synth. Biol* **2022**, 3805–3816 (2022).
118. Sakimoto, K. K., Wong, A. B. & Yang, P. Self-photosensitization of nonphotosynthetic bacteria for solar-to-chemical production. *Science* **351**, 74–77 (2016).
119. Guo, J. *et al.* Light-driven fine chemical production in yeast biohybrids. *Science (New York, N.Y.)* **362**, 813–816 (2018).
120. Yilmaz, S., Nyerges, A., van der Oost, J., Church, G. M. & Claassens, N. J. Towards next-generation cell factories by rational genome-scale engineering. *Nat Catal* **5**, 751–765 (2022).
121. Choi, K. R. *et al.* Systems Metabolic Engineering Strategies: Integrating Systems and Synthetic Biology with Metabolic Engineering. *Trends Biotechnol* **37**, 817–837 (2019).
122. Orsi, E., Claassens, N. J., Nikel, P. I. & Lindner, S. N. Growth-coupled selection of synthetic modules to accelerate cell factory development. *Nat Commun* **12**, 5295 (2021).
123. Nielsen, J. R., Weusthuis, R. A. & Huang, W. E. Growth-coupled enzyme engineering through manipulation of redox cofactor regeneration. *Biotechnology Advances* **63**, 108102 (2023).
124. Rasala, B. A. & Mayfield, S. P. Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. *Photosynth Res* **123**, 227–239 (2015).
125. Muthukrishnan, L. Bio-engineering of microalgae: Challenges and future prospects toward industrial and environmental applications. *Journal of Basic Microbiology* **62**, 310–329 (2022).
126. Satta, A., Esquirol, L. & Ebert, B. E. Current Metabolic Engineering Strategies for Photosynthetic Bioproduction in Cyanobacteria. *Microorganisms* **11**, 455 (2023).
127. Schuchmann, K. & Müller, V. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat Rev Microbiol* **12**, 809–821 (2014).
128. Humphreys, C. M. & Minton, N. P. Advances in metabolic engineering in the microbial production of fuels and chemicals from C1 gas. *Current Opinion in Biotechnology* **50**, 174–181 (2018).
129. Köpke, M. *et al.* Clostridium ljungdahlii represents a microbial production platform based on syngas. *Proceedings of the National Academy of Sciences* **107**, 13087–13092 (2010).
130. Liew, F. E. *et al.* Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. *Nat Biotechnol* **40**, 335–344 (2022).
131. Karim, A. S. *et al.* In vitro prototyping and rapid optimization of biosynthetic enzymes for cell design. *Nat Chem Biol* **16**, 912–919 (2020).
132. Haas, T., Krause, R., Weber, R., Demler, M. & Schmid, G. Technical photosynthesis involving CO2 electrolysis and fermentation. *Nat Catal* **1**, 32–39 (2018).
133. Pan, H., Wang, J., Wu, H., Li, Z. & Lian, J. Synthetic biology toolkit for engineering Cupriviadus necator H16 as a platform for CO2 valorization. *Biotechnology for Biofuels* **14**, 212 (2021).

134. Schlegel, H. G., Kaltwasser, H. & Gottschalk, G. Ein Submersverfahren zur Kultur wasserstoffoxydierender Bakterien: Wachstumsphysiologische Untersuchungen. *Archiv für Mikrobiologie* **38**, 209–222 (1961).
135. Vaneechoutte, M., Kämpfer, P., De Baere, T., Falsen, E. & Verschraegen, G. Wautersia gen. nov., a novel genus accomodating the phylogenetic lineage including Ralstonia eutropha and related species, and proposal of Ralstonia [Pseudomonas] syzygii (Roberts et al. 1990) comb. nov. *International Journal of Systematic and Evolutionary Microbiology* **54**, 317–327 (2004).
136. Davis, D. H., Doudoroff, M., Stanier, R. Y. & Mandel, M. Proposal to reject the genus Hydrogenomonas: Taxonomic implications. *International Journal of Systematic Bacteriology* **19**, 375–390 (1969).
137. Pohlmann, A. *et al.* Genome sequence of the bioplastic-producing "Knallgas" bacterium Ralstonia eutropha H16. *Nature Biotechnology* **24**, 1257–1262 (2006).
138. Römermann, D. & Friedrich, B. Denitrification by Alcaligenes eutrophus is plasmid dependent. *Journal of bacteriology* **162**, 852–4 (1985).
139. Cramm, R. Genomic View of Energy Metabolism in <i>Ralstonia eutropha</i> H16. *Journal of Molecular Microbiology and Biotechnology* **16**, 38–52 (2009).
140. Lee, S. Y. & Park, S. J. *Fermentative production of short-chain-length PHAs. in Biotechnology of Biopolymers: From Synthesis to Patents, vol. 1*. (Wiley–VCH, Weinheim, Germany, 2005).
141. Schlegel, H. G. & Lafferty, R. Growth of 'Knallgas' Bacteria (Hydrogenomonas) using Direct Electrolysis of the Culture Medium. *Nature 1965 205:4968* **205**, 308–309 (1965).
142. Bowien, B. & Schlegel, H. G. Physiology and Biochemistry of Aerobic Hydrogen-Oxidizing Bacteria. *Annual Review of Microbiology* **35**, 405–452 (1981).
143. Volodina, E., Raberg, M. & Steinbüchel, A. Engineering the heterotrophic carbon sources utilization range of Ralstonia eutropha H16 for applications in biotechnology. *Critical Reviews in Biotechnology* **36**, 978–991 (2016).
144. Bowien, B. & Kusian, B. Genetics and control of CO2 assimilation in the chemoautotroph Ralstonia eutropha. *Archives of Microbiology* **178**, 85–93 (2002).
145. Windhöver, U. & Bowien, B. Identification of cfxR, an activator gene of autotrophic CO2 fixation in Alcaligenes eutrophus. *Molecular Microbiology* **5**, 2695–2705 (1991).
146. Schwartz, E. *et al.* Complete nucleotide sequence of pHG1: A Ralstonia eutropha H16 megaplasmid encoding key enzymes of H2-based lithoautotrophy and anaerobiosis. *Journal of Molecular Biology* **332**, 369–383 (2003).
147. Friedrich, C. G., Bowien, B. & Friedrich, B. Formate and Oxalate Metabolism in Alcaligenes eutrophus. *Microbiology* **115**, 185–192 (1979).
148. Bhatia, S. K. *et al.* Production of (3-hydroxybutyrate-co-3-hydroxyhexanoate) copolymer from coffee waste oil using engineered Ralstonia eutropha. *Bioprocess and Biosystems Engineering* **41**, 229–235 (2018).
149. Budde, C. F., Riedel, S. L., Willis, L. B., Rha, C. K. & Sinskey, A. J. Production of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) from plant oil by engineered Ralstonia eutropha strains. *Applied and Environmental Microbiology* **77**, 2847–2854 (2011).
150. Fukui, T., Mukoyama, M., Orita, I. & Nakamura, S. Enhancement of glycerol utilization ability of Ralstonia eutropha H16 for production of polyhydroxyalkanoates. *Applied Microbiology and Biotechnology* **98**, 7559–7568 (2014).

151. Nangle, S. N. *et al.* Valorization of CO2 through lithoautotrophic production of sustainable chemicals in Cupriavidus necator. *Metabolic Engineering* **62**, 207–220 (2020).
152. Löwe, H., Beentjes, M., Pflüger-Grau, K. & Kremling, A. Trehalose production by Cupriavidus necator from CO2 and hydrogen gas. *Bioresource Technology* **319**, 124169 (2021).
153. Collas, F. *et al.* Engineering the biological conversion of formate into crotonate in Cupriavidus necator. *Metabolic Engineering* **79**, 49–65 (2023).
154. Li, H. *et al.* Integrated Electromicrobial Conversion of CO2 to Higher Alcohols. *Science* **335**, 1596–1596 (2012).
155. Liu, C., Colón, B. C., Ziesack, M., Silver, P. A. & Nocera, D. G. Water splitting–biosynthetic system with CO 2 reduction efficiencies exceeding photosynthesis. *Science* **352**, 1210–1213 (2016).
156. Mï¿½ller, J. *et al.* Engineering of Ralstonia eutropha H16 for Autotrophic and Heterotrophic Production of Methyl Ketones. *Applied and Environmental Microbiology* **79**, 4433–4439 (2013).
157. Krieg, T., Sydow, A., Faust, S., Huth, I. & Holtmann, D. CO2 to Terpenes: Autotrophic and Electroautotrophic α-Humulene Production with Cupriavidus necator. *Angewandte Chemie International Edition* **57**, 1879–1882 (2018).
158. Crépin, L., Lombard, E. & Guillouet, S. E. Metabolic engineering of Cupriavidus necator for heterotrophic and autotrophic alka(e)ne production. *Metabolic Engineering* **37**, 92–101 (2016).
159. Chen, X., Cao, Y., Li, F., Tian, Y. & Song, H. Enzyme-Assisted Microbial Electrosynthesis of Poly(3-hydroxybutyrate) via CO2 Bioreduction by Engineered Ralstonia eutropha. *ACS Catal.* **8**, 4429–4437 (2018).
160. Lim, J., Choi, S. Y., Lee, J. W., Lee, S. Y. & Lee, H. Biohybrid CO2 electrolysis for the direct synthesis of polyesters from CO2. *Proceedings of the National Academy of Sciences* **120**, e2221438120 (2023).
161. Torella, J. P. *et al.* Efficient solar-to-fuels production from a hybrid microbial–water-splitting catalyst system. *Proceedings of the National Academy of Sciences* **112**, 2337–2342 (2015).
162. Zhu, X.-G., Long, S. P. & Ort, D. R. Improving photosynthetic efficiency for greater yield. *Annu Rev Plant Biol* **61**, 235–261 (2010).
163. Li, Z. *et al.* Engineering the Calvin–Benson–Bassham cycle and hydrogen utilization pathway of Ralstonia eutropha for improved autotrophic growth and polyhydroxybutyrate production. *Microbial Cell Factories* **19**, (2020).
164. Sánchez-Andrea, I. *et al.* The reductive glycine pathway allows autotrophic growth of Desulfovibrio desulfuricans. *Nature Communications* **11**, (2020).
165. Tabita, F. R. & Small, C. L. Expression and assembly of active cyanobacterial ribulose-1,5-bisphosphate carboxylase/oxygenase in Escherichia coli containing stoichiometric amounts of large and small subunits. *Proceedings of the National Academy of Sciences* **82**, 6100–6103 (1985).
166. Antonovsky, N. *et al.* Sugar Synthesis from CO2 in Escherichia coli. *Cell* **166**, 115–125 (2016).
167. Gleizer, S. *et al.* Conversion of Escherichia coli to Generate All Biomass Carbon from CO2. *Cell* **179**, 1255-1263.e12 (2019).
168. Bang, J. & Lee, S. Y. Assimilation of formic acid and CO2 by engineered Escherichia coli equipped with reconstructed one-carbon assimilation pathways. *Proc Natl Acad Sci U S A* **115**, E9271–E9279 (2018).

169. Bang, J., Hwang, C. H., Ahn, J. H., Lee, J. A. & Lee, S. Y. Escherichia coli is engineered to grow on CO2 and formic acid. *Nature Microbiology* **5**, 1459–1463 (2020).
170. Kim, S. J., Yoon, J., Im, D. K., Kim, Y. H. & Oh, M. K. Adaptively evolved Escherichia coli for improved ability of formate utilization as a carbon source in sugar-free conditions. *Biotechnology for Biofuels* **12**, 1–12 (2019).
171. Tashiro, Y., Hirano, S., Matson, M. M., Atsumi, S. & Kondo, A. Electrical-biological hybrid system for CO2 reduction. *Metabolic Engineering* **47**, 211–218 (2018).
172. Satanowski, A. *et al.* Awakening a latent carbon fixation cycle in Escherichia coli. *Nat Commun* **11**, 5812 (2020).
173. Flamholz, A. I. *et al.* Functional reconstitution of a bacterial CO2 concentrating mechanism in Escherichia coli. *Elife* **9**, e59882 (2020).
174. Guadalupe-Medina, V. *et al.* Carbon dioxide fixation by Calvin-Cycle enzymes improves ethanol yield in yeast. *Biotechnol Biofuels* **6**, 125 (2013).
175. Gassler, T. *et al.* The industrial yeast Pichia pastoris is converted from a heterotroph into an autotroph capable of growth on CO2. *Nat Biotechnol* **38**, 210–216 (2020).
176. Zhan, C. *et al.* Reprogramming methanol utilization pathways to convert Saccharomyces cerevisiae to a synthetic methylotroph. *Nat Catal* 1–16 (2023) doi:10.1038/s41929-023-00957-w.
177. Silva-Rocha, R. *et al.* The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Research* **41**, (2013).
178. Martínez-García, E. *et al.* SEVA 4.0: an update of the Standard European Vector Architecture database for advanced analysis and programming of bacterial phenotypes. *Nucleic Acids Research* **51**, D1558–D1567 (2023).
179. Tee, K. L. *et al.* An Efficient Transformation Method for the Bioplastic-Producing “Knallgas” Bacterium Ralstonia eutropha H16. *Biotechnology Journal* **12**, (2017).
180. Park, H. C., Lim, K. J., Park, J. S., Lee, Y. H. & Huh, T. L. High frequency transformation of Alcaligenes eutrophus producing poly-β-hydroxybutyric acid by electroporation. *Biotechnology Techniques* **9**, 31–34 (1995).
181. Tee, K. L. *et al.* An Efficient Transformation Method for the Bioplastic-Producing “Knallgas” Bacterium Ralstonia eutropha H16. *Biotechnology Journal* **12**, 1700081 (2017).
182. Azubuike, C. C., Gatehouse, A. M. R. & Howard, T. P. pCAT vectors overcome inefficient electroporation of Cupriavidus necator H16. *New Biotechnology* **65**, 20–30 (2021).
183. Steel, H., Habgood, R., Kelly, C. & Papachristodoulou, A. In situ characterisation and manipulation of biological systems with Chi.Bio. *PLOS Biology* **18**, e3000794 (2020).
184. Huang, W. & Wilks, A. A rapid seamless method for gene knockout in Pseudomonas aeruginosa. *BMC Microbiology* **17**, 1–8 (2017).
185. Byagathvalli, G. *et al.* ElectroPen: An ultra-low–cost, electricity-free, portable electroporator. *PLOS Biology* **18**, e3000589 (2020).
186. Hall, B. G., Acar, H., Nandipati, A. & Barlow, M. Growth Rates Made Easy. *Molecular Biology and Evolution* **31**, 232–238 (2014).
187. Solaiman, D. K. Y., Swingle, B. M. & Ashby, R. D. A new shuttle vector for gene expression in biopolymer-producing Ralstonia eutropha. *Journal of Microbiological Methods* **82**, 120–123 (2010).

188. Grant, S. G., Jessee, J., Bloom, F. R. & Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proceedings of the National Academy of Sciences* **87**, 4645–4649 (1990).
189. Xiong, B. *et al.* Genome editing of Ralstonia eutropha using an electroporation-based CRISPR-Cas9 technique. *Biotechnology for Biofuels* **11**, (2018).
190. Murray, N. E. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* **64**, 412–434 (2000).
191. Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A. & Wilson, G. G. Type I restriction enzymes and their relatives. *Nucleic Acids Research* **42**, 20–44 (2014).
192. Sato, S., Fujiki, T. & Matsumoto, K. Construction of a stable plasmid vector for industrial production of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) by a recombinant Cupriavidusnecator H16 strain. *Journal of Bioscience and Bioengineering* **116**, 677–681 (2013).
193. Ehsaan, M., Baker, J., Kovács, K., Malys, N. & Minton, N. P. The pMTL70000 modular, plasmid vector series for strain engineering in Cupriavidus necator H16. *J Microbiol Methods* **189**, 106323 (2021).
194. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
195. Wang, J. Y. & Doudna, J. A. CRISPR technology: A decade of genome editing is only the beginning. *Science* **379**, eadd8643 (2023).
196. Alexeyev, M. F. The pKNOCK series of broad-host-range mobilizable suicide vectors for gene knockout and targeted DNA insertion into the chromosome of gram-negative bacteria. *Biotechniques* **26**, 824–826, 828 (1999).
197. Lenz, O., Schwartz, E., Dernedde, J., Eitinger, M. & Friedrich, B. The Alcaligenes eutrophus H16 hoxX gene participates in hydrogenase regulation. *Journal of Bacteriology* **176**, 4385–4393 (1994).
198. Lenz, O. & Friedrich, B. A novel multicomponent regulatory system mediates H2 sensing in Alcaligenes eutrophus. *Proceedings of the National Academy of Sciences* **95**, 12474–12479 (1998).
199. Quandt, J. & Hynes, M. F. Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. *Gene* **127**, 15–21 (1993).
200. Schäfer, A. *et al.* Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene* **145**, 69–73 (1994).
201. Lenz, O., Lauterbach, L. & Frielingsdorf, S. O2-tolerant [NiFe]-hydrogenases of Ralstonia eutropha H16: Physiology, molecular biology, purification, and biochemical analysis. *Methods Enzymol* **613**, 117–151 (2018).
202. Gruber, S. *et al.* Design of inducible expression vectors for improved protein production in Ralstonia eutropha H16 derived host strains. *Journal of Biotechnology* **235**, 92–99 (2016).
203. Srivastava, S., Urban, M. & Friedrich, B. Mutagenesis of Alcaligenes eutrophus by insertion of the drug-resistance transposon Tn5. *Arch Microbiol* **131**, 203–207 (1982).
204. Peoples, O. P. & Sinskey, A. J. Poly-beta-hydroxybutyrate (PHB) biosynthesis in Alcaligenes eutrophus H16. Identification and characterization of the PHB polymerase gene (phbC). *J Biol Chem* **264**, 15298–15303 (1989).
205. Barnard, G. C., Henderson, G. E., Srinivasan, S. & Gerngross, T. U. High level recombinant protein expression in Ralstonia eutropha using T7 RNA polymerase based amplification. *Protein Expression and Purification* **38**, 264–271 (2004).

206. Park, J. M., Jang, Y.-S., Kim, T. Y. & Lee, S. Y. Development of a gene knockout system for Ralstonia eutropha H16 based on the broad-host-range vector expressing a mobile group II intron. *FEMS Microbiology Letters* **309**, 193–200 (2010).
207. Xiong, B. *et al.* Genome editing of Ralstonia eutropha using an electroporation-based CRISPR-Cas9 technique. *Biotechnology for Biofuels* **11**, (2018).
208. Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T. & Kado, C. I. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *Journal of Bacteriology* **164**, 918–921 (1985).
209. Patinios, C. *et al.* Streamlined CRISPR genome engineering in wild-type bacteria using SIBR-Cas. *Nucleic Acids Research* **49**, 11392 (2021).
210. Kim, H. K. *et al.* In vivo high-throughput profiling of CRISPR–Cpf1 activity. *Nat Methods* **14**, 153–159 (2017).
211. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
212. Priefert, H. *et al.* Identification and molecular characterization of the Alcaligenes eutrophus H16 aco operon genes involved in acetoin catabolism. *J Bacteriol* **173**, 4056–4071 (1991).
213. Jugder, B.-E. *et al.* An analysis of the changes in soluble hydrogenase and global gene expression in Cupriavidus necator (Ralstonia eutropha) H16 grown in heterotrophic diauxic batch culture. *Microb Cell Fact* **14**, 42 (2015).
214. Windhorst, C. & Gescher, J. Efficient biochemical production of acetoin from carbon dioxide using Cupriavidus necator H16. *Biotechnology for Biofuels* **12**, 1–11 (2019).
215. Bommareddy, R. R. *et al.* A Sustainable Chemicals Manufacturing Paradigm Using CO2 and Renewable H2. *iScience* **23**, 101218 (2020).
216. Espah Borujeni, A. & Salis, H. M. Translation Initiation is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism. *J. Am. Chem. Soc.* **138**, 7016–7023 (2016).
217. Espah Borujeni, A. *et al.* Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic Acids Research* **45**, 5437–5448 (2017).
218. Swarts, D. C., van der Oost, J. & Jinek, M. Structural basis for guide RNA processing and seed-dependent DNA targeting and cleavage by CRISPR-Cas12a. *Mol Cell* **66**, 221-233.e4 (2017).
219. Boy, C., Lesage, J., Alfenore, S., Guillouet, S. E. & Gorret, N. Study of plasmid-based expression level heterogeneity under plasmid-curing like conditions in Cupriavidus necator. *Journal of Biotechnology* **345**, 17–29 (2022).
220. Fan, C. *et al.* Defensive Function of Transposable Elements in Bacteria. *ACS Synthetic Biology* **8**, 2141–2151 (2019).
221. Alagesan, S. *et al.* Functional genetic elements for controlling gene expression in Cupriavidus necator H16. *Applied and Environmental Microbiology* **84**, (2018).
222. Holtz, W. J. & Keasling, J. D. Engineering Static and Dynamic Control of Synthetic Pathways. *Cell* **140**, 19–23 (2010).
223. Ni, C., Dinh, C. V. & Prather, K. L. J. Dynamic Control of Metabolism. *Annu Rev Chem Biomol Eng* **12**, 519–541 (2021).
224. Ni, C., Fox, K. J. & Prather, K. L. J. Substrate-activated expression of a biosynthetic pathway in Escherichia coli. *Biotechnol J* **17**, e2000433 (2022).

225. Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A* **102**, 12678–12683 (2005).
226. Dinh, C. V. & Prather, K. L. Layered and multi-input autonomous dynamic control strategies for metabolic engineering. *Current Opinion in Biotechnology* **65**, 156–162 (2020).
227. Dahl, R. H. *et al.* Engineering dynamic pathway regulation using stress-response promoters. *Nat Biotechnol* **31**, 1039–1046 (2013).
228. DeAngelis, K. M. *et al.* Complete genome sequence of “Enterobacter lignolyticus” SCF1. *Stand in Genomic Sci* **5**, 69–85 (2011).
229. Ruegg, T. L. *et al.* An auto-inducible mechanism for ionic liquid resistance in microbial biofuel production. *Nat Commun* **5**, 3490 (2014).
230. Ruegg, T. L. *et al.* Jungle Express is a versatile repressor system for tight transcriptional control. *Nature Communications* **9**, 3617 (2018).
231. Murugan, K., Seetharam, A. S., Severin, A. J. & Sashital, D. G. CRISPR-Cas12a has widespread off-target and dsDNA-nicking effects. *J Biol Chem* **295**, 5538–5553 (2020).
232. Reisch, C. R. & Prather, K. L. J. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in Escherichia coli. *Scientific Reports* **5**, 1–12 (2015).
233. Stukenberg, D., Hoff, J., Faber, A. & Becker, A. NT-CRISPR, combining natural transformation and CRISPR-Cas9 counterselection for markerless and scarless genome editing in Vibrio natriegens. *Commun Biol* **5**, 1–13 (2022).
234. Yishai, O., Lindner, S. N., Gonzalez de la Cruz, J., Tenenboim, H. & Bar-Even, A. The formate bio-economy. *Curr Opin Chem Biol* **35**, 1–9 (2016).
235. Agarwal, A. S., Zhai, Y., Hill, D. & Sridhar, N. The Electrochemical Reduction of Carbon Dioxide to Formate/Formic Acid: Engineering and Economic Feasibility. *ChemSusChem* **4**, 1301–1310 (2011).
236. Jhong, H.-R. “Molly”, Ma, S. & Kenis, P. J. Electrochemical conversion of CO2 to useful chemicals: current status, remaining challenges, and future opportunities. *Current Opinion in Chemical Engineering* **2**, 191–199 (2013).
237. Philips, M. F., Gruter, G.-J. M., Koper, M. T. M. & Schouten, K. J. P. Optimizing the Electrochemical Reduction of CO2 to Formate: A State-of-the-Art Analysis. *ACS Sustainable Chem. Eng.* **8**, 15430–15444 (2020).
238. Al-Tamreh, S. A. *et al.* Electroreduction of Carbon Dioxide into Formate: A Comprehensive Review. *ChemElectroChem* **8**, 3207–3220 (2021).
239. Jahn, M. *et al.* Protein allocation and utilization in the versatile chemolithoautotroph Cupriavidus necator. *eLife* **10**, e69019 (2021).
240. Portnoy, V. A., Bezdan, D. & Zengler, K. Adaptive laboratory evolution—harnessing the power of biology for metabolic engineering. *Current Opinion in Biotechnology* **22**, 590–594 (2011).
241. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol* **1151**, 165–188 (2014).
242. Lee, D.-H., Feist, A. M., Barrett, C. L. & Palsson, B. Ø. Cumulative Number of Cell Divisions as a Meaningful Timescale for Adaptive Laboratory Evolution of Escherichia coli. *PLOS ONE* **6**, e26172 (2011).
243. Friedrich, C. G., Bowien, B. & Friedrich, B. Formate and oxalate metabolism in Alcaligenes eutrophus. *Journal of General Microbiology* **115**, 185–192 (1979).

244. Friedebold, J. & Bowien, B. Physiological and biochemical characterization of the soluble formate dehydrogenase, a molybdoenzyme from Alcaligenes eutrophus. *Journal of Bacteriology* **175**, 4719–4728 (1993).
245. Grunwald, S. *et al.* Kinetic and stoichiometric characterization of organoautotrophic growth of Ralstonia eutropha on formic acid in fed-batch and continuous cultures. *Microbial Biotechnology* **8**, 155–163 (2015).
246. Lee, S.-E., Li, Q. X. & Yu, J. Proteomic examination of Ralstonia eutropha in cellular responses to formic acid. *PROTEOMICS* **6**, 4259–4268 (2006).
247. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* **4**, 457–469 (2003).
248. Lachapelle, J., Reid, J. & Colegrave, N. Repeatability of adaptation in experimental populations of different sizes. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20143033 (2015).
249. Vogwill, T., Phillips, R. L., Gifford, D. R. & MacLean, R. C. Divergent evolution peaks under intermediate population bottlenecks during bacterial experimental evolution. *Proceedings of the Royal Society B: Biological Sciences* **283**, 20160749 (2016).
250. Wein, T. & Dagan, T. The Effect of Population Bottleneck Size and Selective Regime on Genetic Diversity and Evolvability in Bacteria. *Genome Biology and Evolution* **11**, 3283–3290 (2019).
251. Shibai, A. *et al.* Mutation accumulation under UV radiation in Escherichia coli. *Sci Rep* **7**, 14531 (2017).
252. Drlica, K. & Zhao, X. DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiol Mol Biol Rev* **61**, 377–392 (1997).
253. Sanders, C. C. Mechanisms responsible for cross-resistance and dichotomous resistance among the quinolones. *Clin Infect Dis* **32 Suppl 1**, S1-8 (2001).
254. Krishna, S., Maslov, S. & Sneppen, K. UV-Induced Mutagenesis in Escherichia coli SOS Response: A Quantitative Model. *PLoS Computational Biology* **3**, 0451–0462 (2007).
255. Schell, M. A. Control of Virulence and Pathogenicity Genes of Ralstonia Solanacearum by an Elaborate Sensory Network. *Annu Rev Phytopathol* **38**, 263–292 (2000).
256. Garg, R. P. *et al.* Evidence that Ralstonia eutropha (Alcaligenes eutrophus) contains a functional homologue of the Ralstonia solanacearum Phc cell density sensing system. *Molecular Microbiology* **38**, 359–367 (2000).
257. Ray, J. *et al.* Complete Genome Sequence of Cupriavidus basilensis 4G11, Isolated from the Oak Ridge Field Research Center Site. *Genome Announc* **3**, e00322-15 (2015).
258. Marmont, L. S. *et al.* PelA and PelB proteins form a modification and secretion complex essential for Pel polysaccharide-dependent biofilm formation in Pseudomonas aeruginosa. *J Biol Chem* **292**, 19411–19422 (2017).
259. Colvin, K. M. *et al.* PelA deacetylase activity is required for Pel polysaccharide synthesis in Pseudomonas aeruginosa. *J Bacteriol* **195**, 2329–2339 (2013).
260. LaCroix, R. A. *et al.* Use of Adaptive Laboratory Evolution To Discover Key Mutations Enabling Rapid Growth of Escherichia coli K-12 MG1655 on Glucose Minimal Medium. *Applied and Environmental Microbiology* **81**, 17–30 (2015).
261. Rugbjerg, P., Feist, A. M. & Sommer, M. O. A. Enhanced Metabolite Productivity of Escherichia coli Adapted to Glucose M9 Minimal Medium. *Frontiers in Bioengineering and Biotechnology* **6**, (2018).

262. Rajaraman, E. *et al.* Transcriptional analysis and adaptive evolution of Escherichia coli strains growing on acetate. *Appl Microbiol Biotechnol* **100**, 7777–7785 (2016).
263. Große, C., Grau, J., Große, I. & Nies, D. H. Importance of RpoD- and Non-RpoD-Dependent Expression of Horizontally Acquired Genes in Cupriavidus metallidurans. *Microbiol Spectr* **10**, e0012122 (2022).
264. Alper, H. & Stephanopoulos, G. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab Eng* **9**, 258–267 (2007).
265. Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1565–1568 (2006).
266. Tan, F. *et al.* Using global transcription machinery engineering (gTME) to improve ethanol tolerance of Zymomonas mobilis. *Microbial Cell Factories* **15**, 4 (2016).
267. Santos, C. N. S., Xiao, W. & Stephanopoulos, G. Rational, combinatorial, and genomic approaches for engineering L-tyrosine production in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13538–13543 (2012).
268. Guo, X.-W. *et al.* Improved xylose tolerance and 2,3-butanediol production of Klebsiella pneumoniae by directed evolution of rpoD and the mechanisms revealed by transcriptomics. *Biotechnology for Biofuels* **11**, 307 (2018).
269. Clough, S. J., Flavier, A. B., Schell, M. A. & Denny, T. P. Differential Expression of Virulence Genes and Motility in Ralstonia (Pseudomonas) solanacearum during Exponential Growth. *Applied and Environmental Microbiology* **63**, 844–850 (1997).
270. Liu, H., Kang, Y., Genin, S., Schell, M. A. & Denny, T. P. Twitching motility of Ralstonia solanacearum requires a type IV pilus system. *Microbiology* **147**, 3215–3229 (2001).
271. Kang, Y., Liu, H., Genin, S., Schell, M. A. & Denny, T. P. Ralstonia solanacearum requires type 4 pili to adhere to multiple surfaces and for natural transformation and virulence. *Molecular Microbiology* **46**, 427–437 (2002).
272. Dykhuizen, D. E. & Hartl, D. L. Selection in chemostats. *Microbiol Rev* **47**, 150–168 (1983).
273. Gresham, D. & Dunham, M. J. The enduring utility of continuous culturing in experimental evolution. *Genomics* **104**, 399–405 (2014).
274. de Crécy, E. *et al.* Development of a novel continuous culture device for experimental evolution of bacterial populations. *Appl Microbiol Biotechnol* **77**, 489–496 (2007).
275. Selveshwari, S., Lele, K. & Dey, S. Genomic signatures of UV resistance evolution in Escherichia coli depend on the growth phase during exposure. *J Evol Biol* **34**, 953–967 (2021).
276. Barrick, J. E. *et al.* Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* **461**, 1243–1247 (2009).
277. Lenski, R. E. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J* **11**, 2181–2194 (2017).
278. Pfeifer, E., Gätgens, C., Polen, T. & Frunzke, J. Adaptive laboratory evolution of Corynebacterium glutamicum towards higher growth rates on glucose minimal medium. *Sci Rep* **7**, 16780 (2017).
279. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30 (2000).

280. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* **51**, D587–D592 (2023).
281. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**, 221–224 (2010).
282. Jager, H. I. & Kreig, J. A. F. Designing landscapes for biomass production and wildlife. *Global Ecology and Conservation* **16**, (2018).
283. Fletcher, R. J. *et al.* Biodiversity conservation in the era of biofuels: risks and opportunities. *Frontiers in Ecology and the Environment* **9**, 161–168 (2011).
284. Shears, J. Is there a role for synthetic biology in addressing the transition to a new low-carbon energy system? *Microbial Biotechnology* **12**, 824–827 (2019).
285. Lim, H. G., Lee, J. H., Noh, M. H. & Jung, G. Y. Rediscovering Acetate Metabolism: Its Potential Sources and Utilization for Biobased Transformation into Value-Added Chemicals. *Journal of Agricultural and Food Chemistry* **66**, 3998–4006 (2018).
286. Bankefa, O. E. & Claassens, N. J. Hybrid chem-bio production from electricity and CO2 via two-carbon mediators. *Chem* **8**, 3162–3165 (2022).
287. Holm-Nielsen, J. B., Al Seadi, T. & Oleskowicz-Popiel, P. The future of anaerobic digestion and biogas utilization. *Bioresource Technology* **100**, 5478–5484 (2009).
288. Mao, C., Feng, Y., Wang, X. & Ren, G. Review on research achievements of biogas from anaerobic digestion. *Renewable and Sustainable Energy Reviews* **45**, 540–555 (2015).
289. Straub, M., Demler, M., Weuster-Botz, D. & Dürre, P. Selective enhancement of autotrophic acetate production with genetically modified Acetobacterium woodii. *Journal of Biotechnology* **178**, 67–72 (2014).
290. Baek, J. *et al.* Separation of Acetate Produced from C1 Gas Fermentation Using an Electrodialysis-Based Bioelectrochemical System. *Energies* **11**, 2770 (2018).
291. Wierckx, N., Koopman, F., Bandounas, L., De Winde, J. H. & Ruijssenaars, H. J. Isolation and characterization of Cupriavidus basilensis HMF14 for biological removal of inhibitors from lignocellulosic hydrolysatembt. *Microbial Biotechnology* **3**, 336–343 (2010).
292. Chen, H. *et al.* A review on the pretreatment of lignocellulose for high-value chemicals. *Fuel Processing Technology* **160**, 196–206 (2017).
293. Yang, Y. H. *et al.* Optimization of growth media components for polyhydroxyalkanoate (PHA) production from organic acids by Ralstonia eutropha. *Applied microbiology and biotechnology* **87**, 2037–2045 (2010).
294. Reinecke, F. & Steinbüchel, A. Ralstonia eutropha strain H16 as model organism for PHA metabolism and for biotechnological production of technically interesting biopolymers. *Journal of Molecular Microbiology and Biotechnology* **16**, 91–108 (2008).
295. Lu, J., Brigham, C. J., Li, S. & Sinskey, A. J. Chapter 12 - Ralstonia eutropha H16 as a Platform for the Production of Biofuels, Biodegradable Plastics, and Fine Chemicals from Diverse Carbon Resources. in *Biotechnology for Biofuel Production and Optimization* (eds. Eckert, C. A. & Trinh, C. T.) 325–351 (Elsevier, 2016). doi:10.1016/B978-0-444-63475-7.00012-1.
296. Chakravarty, J. & Brigham, C. J. Solvent production by engineered Ralstonia eutropha: channeling carbon to biofuel. *Applied Microbiology and Biotechnology* **102**, 5021–5031 (2018).

297. Raberg, M., Volodina, E., Lin, K. & Steinbüchel, A. Ralstonia eutropha H16 in progress: Applications beside PHAs and establishment as production platform by advanced genetic tools. *Critical Reviews in Biotechnology* **38**, 494–510 (2018).
298. Panich, J., Fong, B. & Singer, S. W. Metabolic Engineering of Cupriavidus necator H16 for Sustainable Biofuels from CO2. *Trends in Biotechnology* **39**, 412–424 (2021).
299. Kunasundari, B., Murugaiyah, V., Kaur, G., Maurer, F. H. J. & Sudesh, K. Revisiting the Single Cell Protein Application of Cupriavidus necator H16 and Recovering Bioplastic Granules Simultaneously. *PLoS One* **8**, e78528 (2013).
300. Sillman, J. *et al.* Bacterial protein for food and feed generated via renewable energy and direct air capture of CO2: Can it reduce land and water use? *Global Food Security* **22**, 25–32 (2019).
301. Chee, J. Y., Lakshmanan, M., Jeepery, I. F., Hairudin, N. H. M. & Sudesh, K. The Potential Application of Cupriavidus necator as Polyhydroxyalkanoates Producer and Single Cell Protein: A Review on Scientific, Cultural and Religious Perspectives. *Applied Food Biotechnology* **6**, 19–34 (2019).
302. Hazards (BIOHAZ), E. P. on B. *et al.* Update of the list of qualified presumption of safety (QPS) recommended microbiological agents intentionally added to food or feed as notified to EFSA 17: suitability of taxonomic units notified to EFSA until September 2022. *EFSA Journal* **21**, e07746 (2023).
303. Pessoa-Jr, A. *et al.* Perspectives on bioenergy and biotechnology in Brazil. in *Applied Biochemistry and Biotechnology - Part A Enzyme Engineering and Biotechnology* vol. 121 59–70 (Appl Biochem Biotechnol, 2005).
304. Geleynse, S. *et al.* The Alcohol-to-Jet Conversion Pathway for Drop-In Biofuels: Techno-Economic Evaluation. *ChemSusChem* **11**, 3728–3741 (2018).
305. Doliente, S. S. *et al.* Bio-aviation Fuel: A Comprehensive Review and Analysis of the Supply Chain Components. *Frontiers in Energy Research* **8**, 110 (2020).
306. Lee, H. M., Jeon, B. Y. & Oh, M. K. Microbial production of ethanol from acetate by engineered Ralstonia eutropha. *Biotechnology and Bioprocess Engineering 2016 21:3* **21**, 402–407 (2016).
307. Sambrook, J. & Russell, D. Molecular Cloning. A Laboratory Manual. *Biochemical Education* **11**, 82 (1983).
308. Schwarze, A., Kopczak, M. J., Rogner, M. & Lenz, O. Requirements for Construction of a Functional Hybrid Complex of Photosystem I and [NiFe]-Hydrogenase. *Applied and Environmental Microbiology* **76**, 2641 (2010).
309. Kelly, C. L. *et al.* Synthetic negative feedback circuits using engineered small RNAs. *Nucleic Acids Research* **46**, 9875 (2018).
310. Chen, J. S. *et al.* Production of fatty acids in Ralstonia eutropha H16 by engineering β-oxidation and carbon storage. *PeerJ* **3**, e1468 (2015).
311. Axe, D. D. & Bailey, J. E. Transport of lactate and acetate through the energized cytoplasmic membrane of Escherichia coli. *Biotechnology and bioengineering* **47**, 8–19 (1995).
312. Marudkla, J., Lee, W. C., Wannawilai, S., Chisti, Y. & Sirisansaneeyakul, S. Model of acetic acid-affected growth and poly(3-hydroxybutyrate) production by Cupriavidus necator DSM 545. *Journal of Biotechnology* **268**, 12–20 (2018).
313. Wang, J. & Yu, J. Kinetic analysis on formation of poly(3-hydroxybutyrate) from acetic acid by Ralstonia eutropha under chemically defined conditions. *Journal of Industrial Microbiology and Biotechnology* **26**, 121–126 (2001).

314. Grousseau, E. Potentialités de production de Poly-Hydroxy-Alcanoates (PHA) chez Cupriavidus necator sur substrats de type acides gras volatifs : études cinétiques et métaboliques. (Toulouse, INSA, 2012).
315. Membrillo-Hernández, J. *et al.* Evolution of the adhE gene product of Escherichia coli from a functional reductase to a dehydrogenase. Genetic and biochemical studies of the mutant proteins. *The Journal of biological chemistry* **275**, 33869–33875 (2000).
316. Dellomonaco, C., Rivera, C., Campbell, P. & Gonzalez, R. Engineered respiro-fermentative metabolism for the production of biofuels and biochemicals from fatty acid-rich feedstocks. *Applied and environmental microbiology* **76**, 5067–5078 (2010).
317. Ghareib, M., Youssef, K. A. & Khalil, A. A. Ethanol tolerance of Saccharomyces cerevisiae and its relationship to lipid content and composition. *Folia microbiologica* **33**, 447–452 (1988).
318. Rogers, P. L., Jeon, Y. J., Lee, K. J. & Lawford, H. G. Zymomonas mobilis for fuel ethanol and higher value products. *Advances in Biochemical Engineering/Biotechnology* **108**, 263–288 (2007).
319. Ma, M. & Liu, Z. L. Mechanisms of ethanol tolerance in Saccharomyces cerevisiae. *Applied microbiology and biotechnology* **87**, 829–845 (2010).
320. Mukhopadhyay, A. Tolerance engineering in bacteria for the production of advanced biofuels and chemicals. *Trends in Microbiology* **23**, 498–508 (2015).
321. Pearcy, N. I. *et al.* A genome-scale metabolic model of Cupriavidus necator H16 integrated with TraDIS and transcriptomic data reveals metabolic insights for biotechnological applications. *PLOS Computational Biology* **18**, e1010106 (2022).
322. Shin, J. *et al.* Genome-wide CRISPRi screen identifies enhanced autolithotrophic phenotypes in acetogenic bacterium Eubacterium limosum. *Proceedings of the National Academy of Sciences* **120**, e2216244120 (2023).
323. Bengelsdorf, F. R. *et al.* Bacterial Anaerobic Synthesis Gas (Syngas) and CO2+H2 Fermentation. *Adv Appl Microbiol* **103**, 143–221 (2018).