Understanding Microbial Coexistence Via Targeted Gene Knockouts With An Improved CRISPR/Cas9 Assisted Genome Editing Method

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Hongyi Liu

May 2023

ABSTRACT

This thesis is part of a larger project that aims to understand the fundamental principles governing the coexistence of strains in microbial consortia. The main objective of this thesis was to engineer a library of *Escherichia coli* strains, each of which could metabolize a prescribed set of carbon sources. To accomplish this, a scarless genetic knockout system that utilizes the CRISPR/Cas9 and the λ red recombination system was employed to obtain targeted gene knockouts. We encountered several difficulties building the library, which led us to improve the scarless genetic knockout system by introducing a fast screening method for false positive transformants, which is the main result of the thesis. Then, we used this improved scarless genetic knockout system to create scarless knockouts of genes involved in the metabolizing maltose (*malG, malF, malE, malK, lamB, malM*) and galactose (*galT, galK, galM*), and non-scarless knockouts of genes involved in metabolizing fructose (*fruB, fruK,fruA*), mannose (*manX, manY, manZ*) and ribose (*rbsD, rbsA, rbsC, rbsB, rbsK)*. In parallel to engineering these knockout strains, we utilized strains from the *E. coli* Keio knockout collection to investigate how catabolite repression might affect the study of coexistence criteria in our engineered consortia, and performed a preliminary experiment to test mathematical theories of microbial coexistence.

BIOGRAPHICAL SKETCH

Hongyi Liu is a Master student in Assistant Professor Andrea Giometto's lab. She received her bachelor's degree major in Water Science and Engineering at Hua Qiao University.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my parents for their unwavering support throughout my academic journey at Cornell University. I am deeply grateful to Professor Andrea Giometto for his considerate and patient guidance during my master's program. His mentorship and support were invaluable in shaping my research and academic endeavors. I would like to express my gratitude to my committee chair member Professor Buz Barstow for his guidance on my thesis. I would also like to extend my appreciation to my lab members and for their generous assistance and for creating a collaborative and supportive research environment. Finally, I would like to thank Yang Jia for providing support during the ups and downs of the research process.

TABLE OF CONTENTS

CHAPTER 1 LIBRARY CONSTRUCT

CHAPTER 2 METABOLISM AND COEXISTENCE

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Definition
E. coli	*Escherichia coli*
ABC	ATP-Binding Cassette Transporters
CA-NHEJ	CRISPR-Cas9 Assisted Non-Homologous End-Joining
CAGO	CRISPR/Cas9-Assisted gRNA-Free One-Step
CarPE	Carbon Phenotype Explore
CCR	Catabolite Repression
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DSB	Double-Strand Break
HPLC	High-Performance Liquid Chromatography
IPTG	Isopropyl ß-D-1-Thiogalactopyranoside
LB	Lysogeny Broth
mRFP1	Red Fluorescent Protein Gene
OD	Optical Density
PCR	Polymerase-Chain-Reaction
PTS	Phosphoenolpyruvate: Carbohydrate Phosphotransferase System
sgRNA	Single-Stranded Guide RNA
sfGFP	Superfolder Green Fluorescent Protein Gene
SOC	Super Optimal broth with Catabolite repression
TALENs	Transcription Activator-Like Effector Nucleases
Tm	Melting Temperature
WWTPs	Water And Wastewater Treatment Plants
ZFN	Zinc-Finger Nucleases

LIST OF SYMBOLS

Symbol	Definition
Δ*fruBKA::Neo/KanR-N20PAM*	Genes *fruBKA* deleted, replaced with Neo/Kan-N20PAM
Δ*fruBKA*	Genes *fruBKA* deleted
λ -RED	Refers to Lambda Red recombination system

PREFACE

CHAPTER 1 LIBRARY CONSTRUCT

The main objective of this chapter was to construct a library of *Escherichia coli* strains with specific genes knocked out, rendering them unable to metabolize a prescribed set of sugars. We successfully knocked out genes involved in the catabolism of maltose and galactose. However, a large fraction of false positive transformants posed challenges during the knockout process. Thus, we improved a recently published scarless genetic knockout system that adopted Cas9 and the λ Red recombinase to perform scarless deletions and insertions in *E. coli.*

1. Introduction

The birth of synthetic biology can be traced back to approximately 70 years ago[1]. Development of molecular cloning and Polymerase-Chain-Reaction (PCR) has enabled genetic manipulation research, leading to a rapid turnover with the invention of new techniques. Starting from the early 2000s, a lot of efforts in synthetic biology were dedicated to the development and characterization of genetic circuits[2,3]. An example is the construction of the Repressilator[4], where a plasmid was developed with three repressor systems that repress each other and associated fluorescence proteins. When the plasmid was transformed into *Escherichia coli*, a temporally periodic change in fluorescence was observed. Since then, several other circuits[5,6] have been developed and our understanding of how to design and modulate them has greatly increased.

Nowadays, synthetic biology is gradually shifting its focus from genetic circuits towards microbial consortia[7]. Microbial communities can perform complex functions that individual strains cannot, they are ubiquitous in nature and have great importance in various disciplines[8,9] These features drive researchers to investigate the fundamental mechanism behind microbial interactions within consortia. An example in the field of environmental engineering would be

microbes in wastewater treatment. Understanding how they perform when treating wastewater and how to control its composition are very important for the management and development of Water and Wastewater Treatment Plants (WWTPs). Specifically, engineering microbial consortia holds promise towards designing better wastewater treatment procedures. For example, a higher and more persistent wastewater treatment efficiency could be achieved by modifying certain characteristics of microbes, such as their physicochemical absorption and biotransformation abilities[10]. Similarly, understanding how to engineer microbial consortia may provide insights into modulating the gut microbiome composition, and introducing strains that perform desired functions into it. Other applications of interest include optimizing the production of probiotics, and improving the yield of metabolic pathways[11].

In this study, we are interested in understanding the basic principles that govern the coexistence of strains within microbial consortia. To develop such an understanding, our goal was to engineer a community of *Escherichia coli* strains that metabolize different sets of carbon sources, which will be used by the Giometto group in the future to investigate how metabolism and catabolic repression affect coexistence at the community level. Previous research in this field has used either mathematical modeling or experiments with natural communities to investigate the basic principle of microbial coexistence. For example, Posfai et al.[12]analyzed a mathematical resource competition model with trade-offs to explore the coexistence of microbes. Friedman et al.[13]have applied an experimental strategy in which they isolated strains from the wild, co-cultured strains in the laboratory, and tried to understand which strains characteristics correlated with the observed coexistence. On the engineering side, Xia et al.[14] investigated how the simultaneous utilization of lignocellulosic hydrolysates affected the coexistence of engineered *E. coli* strains. However, there is still a large gap between mathematical models or theories and a

comprehensive understanding of the mechanism behind it. The predictions of these theories have not been tested experimentally and one of the reasons for that is that the understanding of metabolic interaction microorganisms in the natural environment is very incomplete[15]. This leads to the formidable challenge of establishing a comprehensive blueprint of these theories in a non-model setting. Rather than attempting such a daunting investigation, we take the approach of engineering a microbial microcosm made of strains for which we know the exact metabolic interactions.

Can we predict microbial coexistence, based on precise knowledge of the resources that different strains can metabolize? To answer this question, we planned to engineer a library of *E. coli* strains, each of which can only metabolize a prescribed set of carbon sources, which is a subset of the carbon sources that wild type *E. coli* can metabolize. We began building this library by knocking out genes involved in the metabolization of specific sugars. We chose *E. coli* as they are genetically modified with ease and their utility as an industrial workhorse for producing different chemicals. To prepare this library, we attempted to use a "scarless" genetic knockout technique that relies on the CRISPR/Cas9 system and the λ Red recombination system[16]. In the process of applying this technique, we encountered a variety of difficulties. These include a large number of false positive transformants, with proportions ranging from 0-3% to 70% depending on the construct, which hindered our ability to perform all the desired knockouts. To overcome these challenges, the primary result of this thesis was to improve the scarless genetic knockout system by decreasing the occurrence of false positive transformants. To achieve this, genes encoding for different fluorescent proteins were introduced in different plasmids used in the scarless genetic knockout system, allowing for the detection of false positive transformants via a fluorescence microscope. After improving the genetic knockout technique, we successfully

created scarless knockouts of genes involved in the metabolism and uptake of maltose (*malG, malF, malE, malK, lamB, malM*), galactose (*galT, galK, galM*), and non-scarless knockouts of genes involved in the metabolism and uptake of fructose (*fruB, fruK,fruA*), mannose(*manX, manY, manZ*) and ribose (*rbsD, rbsA, rbsC, rbsB, rbsK)*. For these latter genes, the recombination process to remove the selection marker has not yet been completed.

1.1 Genetic Engineering

Genome editing is a powerful tool. The emergence of versatile genome editing tools has allowed the sequence-specific manipulation of gene editing of various cell types. It has revolutionized research on various disciplines, such as synthetic biology, disease modeling[17], gene therapy[18] and agricultural sciences[19]. The three mainly used genome editing platforms nowadays are Zinc-Finger Nucleases (ZFN), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated protein (Cas) systems (CRISPR/Cas system). ZFNs are fusions between a non-sequence-specific cleavage domain to a site-specific DNA-binding domain. The DNA-binding domain is loaded on the zinc finger and can be generated to target a wide range of possible DNA sequence[20]. However, off-target mutations hinder the universal usage of ZFNs for genome editing, it is challenging to create zinc-finger domains that can effectively recognize all DNA triplets. Following the work on ZFNs, another type of engineered nuclease, TALENs, has emerged. This system has multiple effector domains that are accessible to support the fusion of TALE repeats for different genomic modification purposes[21]. Compared to ZFNs, no selection or directed evolution is necessary to engineer TALE arrays, and it has been reported to show improved specificity and reduced toxicity. However, large scale of repetitive sequences in TALENs makes the design and efficient delivery of them into cells challenging.

The CRISPR/Cas system is a unique prokaryotic immune system that is adaptive and inheritable. It is currently the known system with these characteristics and it has therefore been intensively adopted for both fundamental studies and biotechnological applications. The discovery of this system can be traced back to the year 1987, when an unusual structure of repetitive DNA sequences from *E. coli* iap gene was discovered[22], but its function remained unclear at that time. It was not until the emergence of genome and plasmid sequencing that researchers realized these similar repeated clusters were derived from invading genomes[23]. A few years later, the CRISPR/Cas system was proposed to function as an inheritable, adaptative immune system[24]. What made this system revolutionize the genome editing is its ability to recognize and cleave invasive genetic elements, generating double stranded breaks (DSBs) at any genomic locus that contains protospacer adjacent motif (PAM) sequences[25]. The CRISPR system has six types and at least twenty-nine subtypes, but the most frequently used subtype is CRISPR/Cas9 system, which is distinguished from others due to its flexibility and user-friendly platform. The CRISPR/Cas9 system comprises two components, a single-stranded guide RNA (sgRNA) and a Cas9 endonuclease (Fig. 1). It generally has three steps[26] in the editing process: recognition, cleavage and repair. The sgRNA recognizes the target sequence in the gene of interest and programs Cas9 to bind to the target protospacer, Cas9 then precisely cleaves the DNA to generate a double-strand break (DSB) at the PAM sequence, then the DSB is repaired. However, there exist PAM-free regions located at 20 or more base pairs away from the nearest NGG motif that cannot be edited by the CRISPR/Cas9 system. Additionally, the editing efficiency and accuracy can be deficient in single or small areas near the 5' end protospacer. Many approaches have been explored to increase the editing efficiency, such as chemical modification at specific guide sequences sites[27] and improving the Cas9 target recognition fidelity[28].

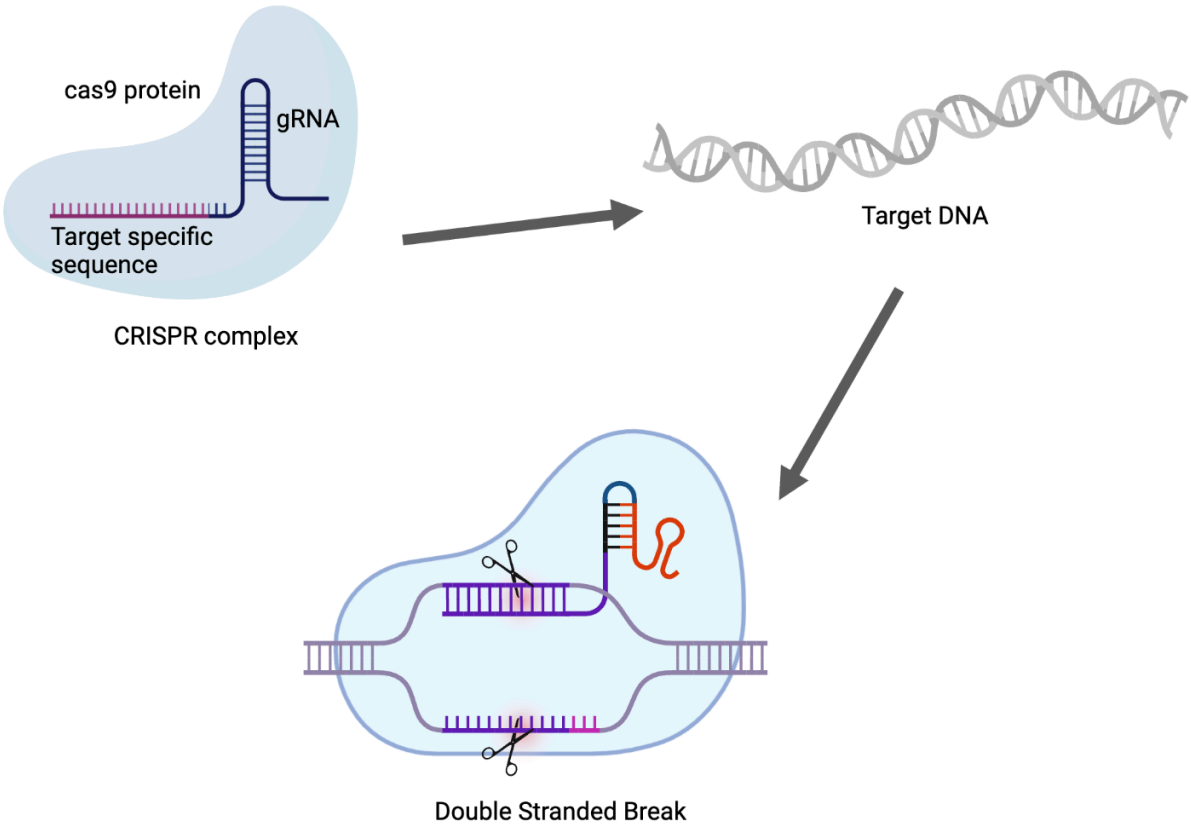


Figure 1 Illustration of double-strand break created by the CRISPR complex.

Despite a promising application in eukaryotic genome editing, the CRISPR/Cas9 genome editing system has not been as widely adopted for bacterial genome editing. This is because the DSB is repaired through non-homologous end joining or homologous recombination in eukaryotes, which rarely happens in bacteria[29]. To overcome this, current studies have developed methods to achieve CRISPR/Cas9-based genome-editing in bacteria. For example, via the use of heterologous recombinases and DNA repair mechanisms. For example, Su et al.[25] developed a one-step inactivation of chromosomal gene(s) by CRISPR-Cas9 assisted non-homologous end-joining (CA-NHEJ). In this protocol, DSBs can be successfully re-joined in the presence of heterologously expressed NHEJ systems with only the construction of a specific sgRNA and single step electroporation. This method was proved to be efficient in larger chromosomal

fragments deletions as well. However, multiple rounds of gRNA design are needed for different gene deletions.

Zhao et al.[16] developed a simple CRISPR/Cas9- assisted gRNA-free one-step (CAGO) genome editing technique that reduces the off-target frequency. This technique can achieve controlled genome editing and efficient homologous recombination without the design and construction of gRNAs. In their work, they designed a linear fragment (Fig. 4, Fig. 6) that contains different regions of homology to the chromosome and a universal N20PAM sequence. After insertion into the chromosome, the N20PAM sequence is recognized by a unique gRNA expressed by the pCAGO plasmid (Fig. 3). Then, addition of L-arabinose induces Cas9, expressed from the helper plasmid pCAGO, to generate a DSB at the universal N20PAM sequence. Finally, addition of isopropyl ß-D-1-thiogalactopyranoside (IPTG) induces the λ Red system, expressed from the helper plasmid pCAGO, to produce an intra-chromosomal recombination to achieve scarless editing. Details of the helper plasmid pCAGO were included in the Chapter 1, Section 2.1.1.

In this work, we adopted the pCAGO method to achieve gene knockouts. However, in this process, we encountered a high proportion of false positive transformants colonies. The estimated range was between 0-3 % to 70%, as determined by colony PCR. This hindered us from creating the entire library of knockout *E. coli* strains that we desired. Given the high frequency of false positive transformants, we decided to invest some of the thesis effort towards improving the pCAGO methodology. To improve construction efficiency of the linear fragment required for the pCAGO gene knockout methodology, rather than producing a linear DNA fragment via Golden Gate assembly, we decided to build a circular DNA fragment that included a backbone and the DNA sequence required for the pCAGO method. The addition of backbone

could enable the fragment to replicate as a plasmid in *E. coli.* The entire fragment was assembled with Golden Gate Assembly. In addition, to facilitate the detection of false positive transformants, we introduced the superfolder green fluorescent protein gene (sfGFP) in the backbone used for Golden Gate assembly and the red fluorescent protein gene (mRFP1) on the plasmid from which we amplified the resistance marker. In this way, we could screen the colonies before conducting the colony PCR, which allowed us to avoid picking false positive transformants. We have validated the improved screening methodology using colony PCR, with good agreement between the screening based on fluorescence imaging and the colony PCR measurements.

1.2 Genes involved in the transport of carbon sources

Sugars are a valuable source of carbon for *E. coli*, they can be used as carbon skeletons[30], and they are major carbon and energy sources for the metabolism of living organisms. In living organisms or cells, thousands of transport and metabolic reactions of various carbon sources are interconnected to facilitate the breakdown of organic compounds, generate energy[31] and provide the building blocks for biogenesis. Therefore, understanding sugar metabolization by microbes is of prime interest in systems biology. In particular, transport of sugars into the cytoplasm, which is the very first step of sugar catabolism in most living organisms[32], plays a vital role for cells to utilize these compounds effectively. There are specific enzymes responsible for the transport of sugars, which are transcribed and translated from different cognate genes. For example (Fig. 2), the mannose-specific PTS enzyme II is encoded by the genes *manX, manY* and *manZ*. These genes are important for the uptake of mannose, and without them, D-mannopyranose cannot be transported into the cytosol, therefore the following reactions could not occur and the cell could not grow on mannose as the sole carbon source.

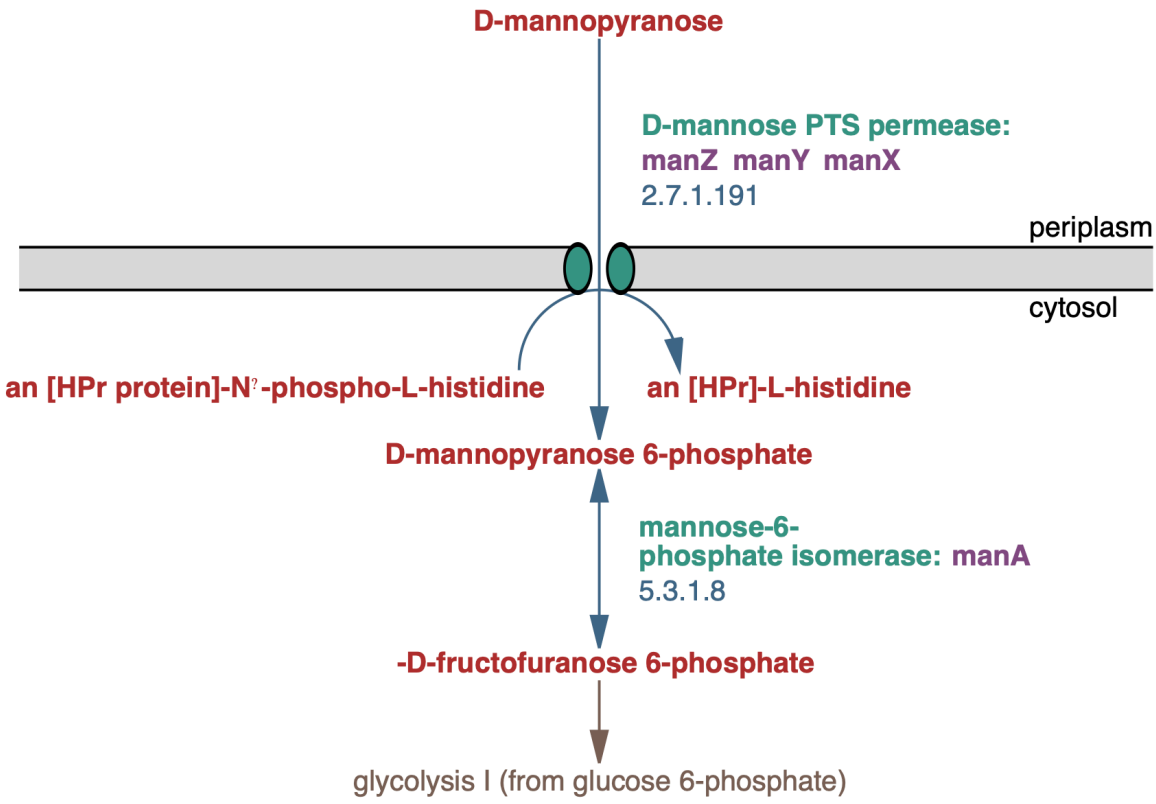


Figure 2 D-Mannose transport process. (From Ecocyc).

Such carbon source metabolism has been extensively studied for decades, but how the regulation of genes involved in the metabolism of different sugars affect the expression of others remains incompletely understood. Nevertheless, the advent of genome editing and genome-scale sequencing have enabled the collections of mutant strains and the analysis of gene functions, facilitating a more thorough exploration of carbon metabolism. Tong et al.[33] utilized the *E. coli* Keio knockout collection to explore gene dispensability in *E. coli*. In this collection[34], *E. coli* strains with single genes knocked out were generated. Tong et al.[33] tested and summarized the kinetic growth measurements of these strains under thirty different carbon source conditions. While this study yielded a foundation to understand *E. coli* metabolism and catabolic repression, it is insufficient to comprehensively understand the coexistence in mixed carbon source

environments, as the metabolic regulation is still unclear. To overcome this limitation, we aim to build a library of strains that have a determined incapability of metabolizing certain types of carbon sources. In the pursuit of this goal, for each sugar we investigated, we decided to delete genes involved in the transport of the sugar, as well as the first enzyme in the corresponding catabolic pathway. This approach will enable us to confirm that these engineered strains are incapable of both intaking and metabolizing specific carbon sources.

The library will be further applied in the investigation of metabolic network and coexistence of *E. coli* community, which are further explained in Chapter 2.

2. Methods

2.1 *E. coli* Strains library Construct

2.1.1 Plasmid construct

The functional region of the plasmid pCAGO contains three main sections: the cas9 region, the gRNA region and the λ Red system region (Fig. 3). The cas9 region is capable of creating a double- stranded break (DSB) at the inserted universal N20PAM sequence (in this specific case, the sequence is 5'-TAGTCCATCGAACCGAAGTA-3'), to which it is guided to by the sgRNA expressed from the pCAGO plasmid. The cas9 gene is controlled by the arabinose-inducible PBAD promoter, L-arabinose is added as 2g/L final concentration in this work. The λ Red system contains the proteins Exo, Beta, and Gam. All three components are expressed to allow both for the insertion of the editing cassette and for recombination used to excise the selection marker from the genome. Exo degrades linear dsDNA for a better performance. Beta promotes the annealing of complementary ssDNA and helps achieve recombination. Gam prevents the inserted linear DNA from being digested. The λ Red system is controlled by the IPTG inducible promoter Ptrc, and a final concentration of 0.1 mM IPTG is used in this experiment to induce it. The pCAGO plasmid contains a temperature-sensitive replicon repA101ts (not shown in figure), which was used for the easy curing of the pCAGO plasmid from the edited cells.

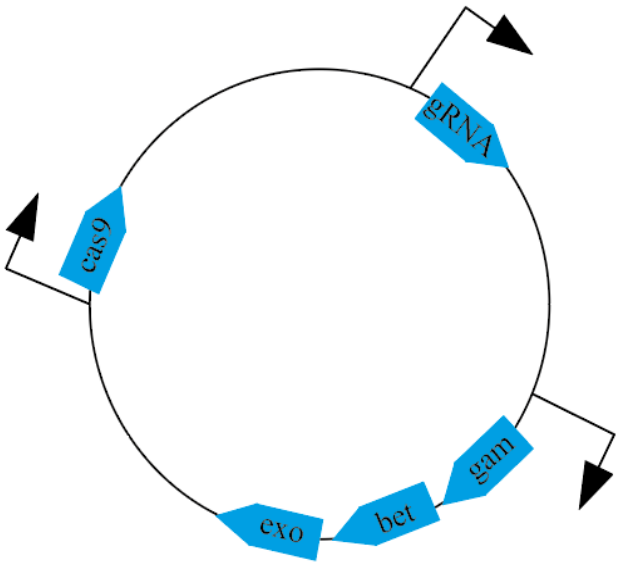


Figure 3 pCAGO function region. Three major sections are illustrated in the graph: the cas9 region, the gRNA region and the λ Red system region.

2.1.2 Genes targeted for deletion

Genes were selected based on their ability to encode proteins that control the sugar transport system as indicated by the bioinformatics database Ecocyc[35]. Transport systems include both phosphoenolpyruvate: carbohydrate phosphotransferase system (PTS) and ATP-binding cassette (ABC) transporters. A table (Table 1) of target genes is given below.

Table 1 Selected genes

Carbon Source	Genes
Maltose	*malG, malF, malE, malK, LamB, malM*
Mannose	*manX, manY,manZ*
Galactose	*galT, galK, galM*
Fructose	*fruB, fruK, fruA*
Sorbitol	*srlB, srlD, gutM*
Ribose	*rbsD, rbsA, rbsC, rbsB, rbsK*

2.1.3 Construction of the editing cassette

The linear DNA cassette used to knockout genes of interest consists of four different parts, a left homologous arm (L), a selection marker (Kanamycin resistance gene, Neo/KanR) with a CRISPR/Cas9 recognition region (N20PAM), a L_short arm, and a right homologous arm (R) (Fig. 4). The L and R homologous arms, each around 600 base pairs long, were amplified from the genomic DNA of *E. coli* DH5α. They were designed to be homologous to the upstream region of genes targeted for deletion. The selection marker was amplified from the plasmid pBbAW4k-loxP-TT-loxP-mRFP1[36](plasmid#134405 from the plasmid repository Addgene), and the N20PAM sequence was added as an overhang in one of the primers used for amplification; this fragment was the same for different Golden Gate Assembly designs. The L_short arm, approximately 250 base pairs in length, was designed to have around 200 bp homologous to the L part. This L_short arm allowed for intracellular recombination after insertion, following a DSB at the N20PAM region.

Figure 4 The general construct of the linear DNA cassette. Four parts are included, a left homologous arm (L), a selection marker (Kanamycin resistance gene, Neo/KanR) with a CRISPR/Cas9 recognition region (N20PAM), a L_short arm, and a right homologous arm (R). L and R homologous arms are approximately 600 bp each, the L_short arm is approximately 250bp. Each of these parts is designed with different 4-nt linkers where the restriction enzyme will cut from. Then the overhangs anneal in Goldengate Assembly.

2.1.4 Modularized construction of the editing cassette

In the original work in which pCAGO was developed[16], the optimized protocol stipulated assembling the four fragments (L, Neo/KanR and N20PAM, L_short, and R) via Golden Gate assembly. However, in our hands, multiple fragments of various lengths were created in the Golden Gate assembly (Fig. 11), resulting in low recombination efficiencies for different

constructs after electroporation. Therefore, we investigated two alternative approaches to improve this method.

2.1.4.1 Golden Gate Assembly of the linear DNA cassette

Golden Gate Assembly allows for efficient and seamless assembly of multiple inserts into a vector backbone[37]. It employs Type IIS restriction enzymes (e.g., BsaI) and the T4 DNA ligase[38]. The restriction enzyme cleaves DNA outside the recognition sequence of each part ('XXXX' and 'YYYY' in Fig. 5), leaving 4-nt linkers between the insertion fragments as overhangs. These linkers were designed to be complementary to the overhang region from the part that precedes or follows each fragment, enabling simultaneous and seamless assembly of pieces. We adopted a consistent set of linkers for all of our constructs, following protocols from the Barrick lab[39].

Figure 5 Golden Gate Assembly Design based on protocols from the Barrick Lab. The highlighted restriction enzyme (BsaI), recognizes and cleaves DNA outside of its recognition sequences. BsaI cut along the dashed line and formed overhangs, these overhangs then anneal to sequences complementary to their sequence, allowing for Golden Gate Assembly.

2.1.4.2 Assembly via PCR amplification

An alternative assembly method to construct the linear DNA cassettes required for gene knockout is directly via PCR amplification, which does not involve Golden Gate Assembly. We tested this method to knock out genes involved in the utilization of mannose and galactose. To achieve this, we used PCR amplification to assemble the fragment containing left homologous arm (L), Neo/KanR and N20PAM, L_short and a right homologous arm (R). However, the

length and construct of each part were different from the design that involves Golden Gate assembly. Instead of assembling four parts via Golden Gate assembly, we performed a one-step PCR to amplify the marker (Neo/KanR) and N20PAM region with oligos that had homology to the region we wanted to delete (Fig. 6). The left homologous arm (L) of 65bp was embedded within the forward primer, while the right homologous arm (R, 40bp) and L_short (40 bp) arm were embedded within the reverse primer. The L_short sequence was taken from the region upstream of the left homologous arm.

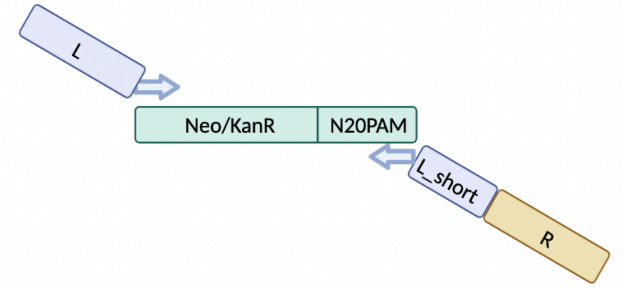
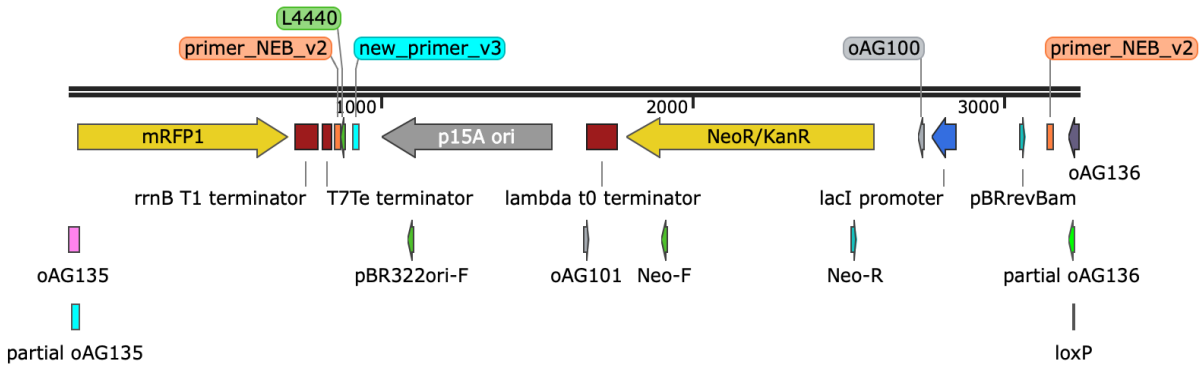


Figure 6 The annealing process of linear DNA cassette via PCR amplification. Four parts are included, a left homologous arm (L), a selection marker (Kanamycin resistance gene, Neo/KanR) with a CRISPR/Cas9 recognition region (N20PAM), a L_short arm, and a right homologous arm (R). The left homologous arm (L) of 65bp was embedded within the forward primer, while the right homologous arm (R, 40bp) and L_short (40 bp) arm were embedded within the reverse primer. They are the overhangs of the forward and reverse primers when amplifying the marker region, respectively.

To enhance screening efficiency, we modified the plasmid (plasmid #134405) from which we amplified the resistance marker (Neo/KanR and N20PAM) to make it constitutively express a red fluorescent protein gene(mRFP1). The plasmid #134405 contained mRFP1, which was not expressed because a terminator was located between the promoter and the start codon (the plasmid was originally designed as a reporter for a light-inducible Cre-lox system). To remove the terminator, two primers oAG135 and oAG136 (Table 2, Fig. 7a) were employed to amplify the plasmid #134405. The PCR product (genes layout shown in Fig. 7b) was analyzed in gel electrophoresis, treated with DpnI (1ul of DpnI mixed with 50ul of product and incubated at 37°C for 1h) to digest the methylated original plasmid which were not produced via PCR. Then

the product was purified with column-based purification and transformed into wildtype DH5α. Two transformants exhibiting fluorescence under fluorescence microscope were selected and plasmid extracted. Till this point, the deletion of the terminator loxP upstream of the mRFP1 region was achieved, achieving the continuous mRPF1 expression in the modified plasmid. The Neo/KanR and N20PAM region, which was used PCR amplification template, was subsequently amplified from the product with oligos oAG100 and oAG101 (Table 2). It is important to note that mRFP1 only exists on the template and is not included in the Neo/KanR and N20PAM fragment, and the template was expected to be digested by DpnI after amplification, as it is not methylated after PCR. Therefore, this approach enabled the screening of false positives transformants that were transformed with circular plasmids which showed red fluorescence under the fluorescence microscope.

a)



b)

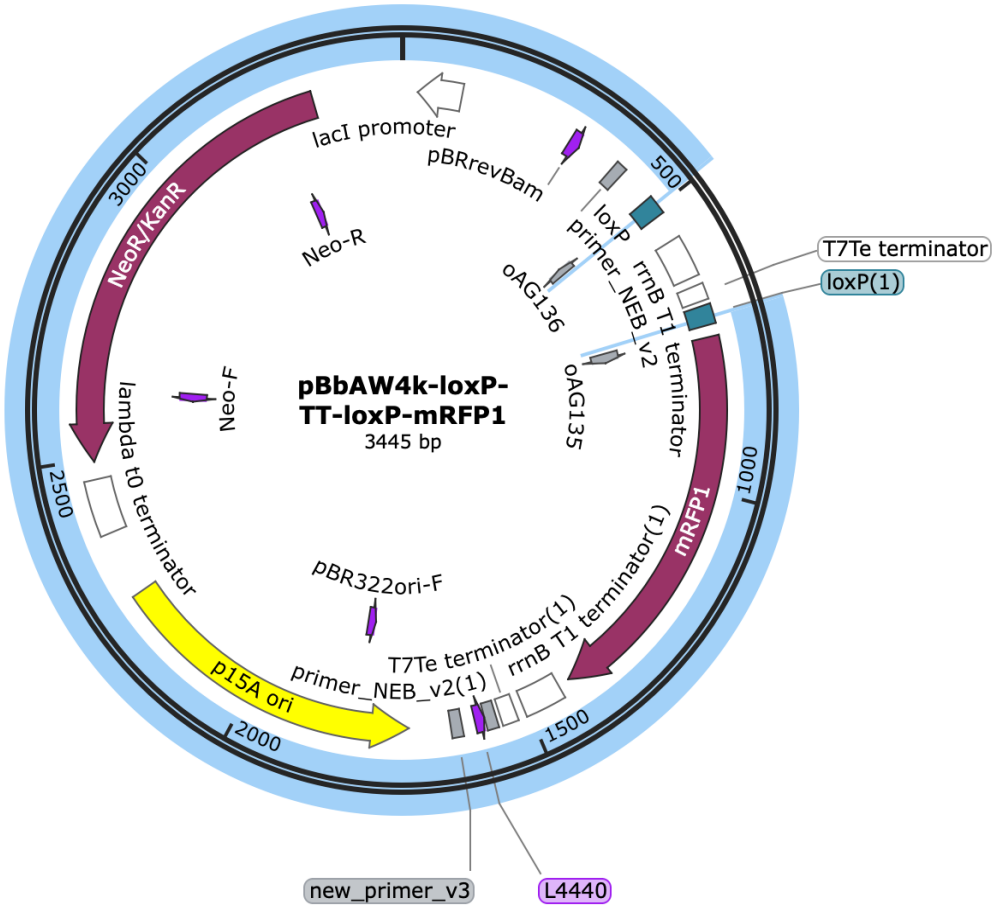


Figure 7 a) Map of part of the plasmid #134405 region after amplification removed, loxP was removed. Pictures are from SnapGene. b) Map of the plasmid #134405. Two primers (oAG135, oAG136) were used to amplify the region highlighted in blue to remove the terminator loxP.

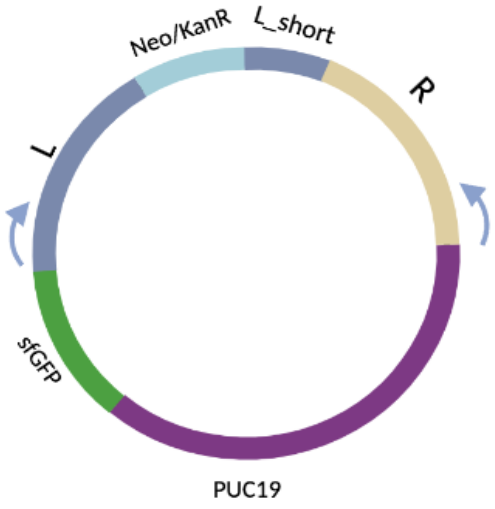
Table 2 Oligos Sequences

Oligos name	Sequence (5'- 3')
oAG135	GAAATAACTTTTTAAGAAGGAGATATACATATGG
oAG136	CCTTCTTAAAAAGTTATTTCGCCGTGTC
oAG100	TAGCATCACGGTCTCAATCCTCACTCATTAGGCACC
oAG101	AAAGGTCTCACAGCCCTTACTTCGGTTCGATGGACTAGTGCTTGGATTCTC

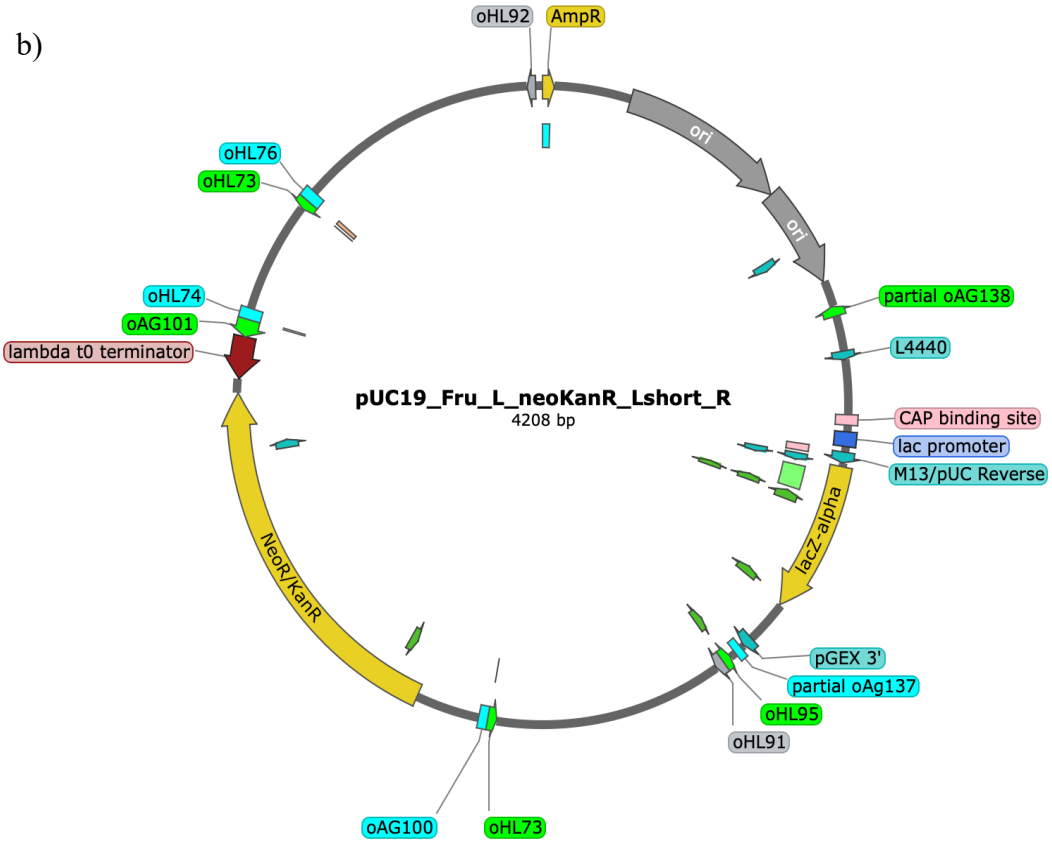
2.1.4.3 Assembly of the linear DNA cassette via Golden Gate cloning, using a plasmid backbone

To ensure the creation of a single DNA cassette from Golden Gate cloning, rather than several different sizes of incomplete assemblies, a circular fragment is desirable. To achieve this, we decided to include the PUC19 vector backbone in our Golden Gate assemblies. The backbone was obtained by amplifying PUC19 via PCR using oligos with overhangs containing the BsaI recognition site and the correct recognition sequence for assembly. Then, Golden Gate Assembly was conducted to assemble the circular fragment (Fig. 8a). An example of PUC19 backbone created for the deletion of genes involved in Fructose is illustrated in figure 8b. This product was then transformed into wildtype DH5α (without pCAGO) and plated on LB plate with Kanamycin under 37°C. After culturing for one day, three single colonies from each plate were picked and incubated overnight for plasmid extraction. The extraction product was then amplified with oligos homologous to the beginning and end of the linear fragment (oHL91and oHL92 in this example) that we wanted to use to perform the gene knockout. Following digestion with DpnI, the amplification products were column extracted and verified by long-read sequencing. At this point, the targeted linear fragment was created.

a)



b)


oHL92
AmpR
ori
ori
partial oAG138
L4440
CAP binding site
lac promoter
M13/pUC Reverse
lacZ-alpha
pGEX 3'
partial oAg137
oHL95
oHL91
oAG100
oHL73
NeoR/KanR
lambda t0 terminator
oAG101
oHL74
oHL73
oHL76
pUC19_Fru_L_neoKanR_Lshort_R
4208 bp

c)

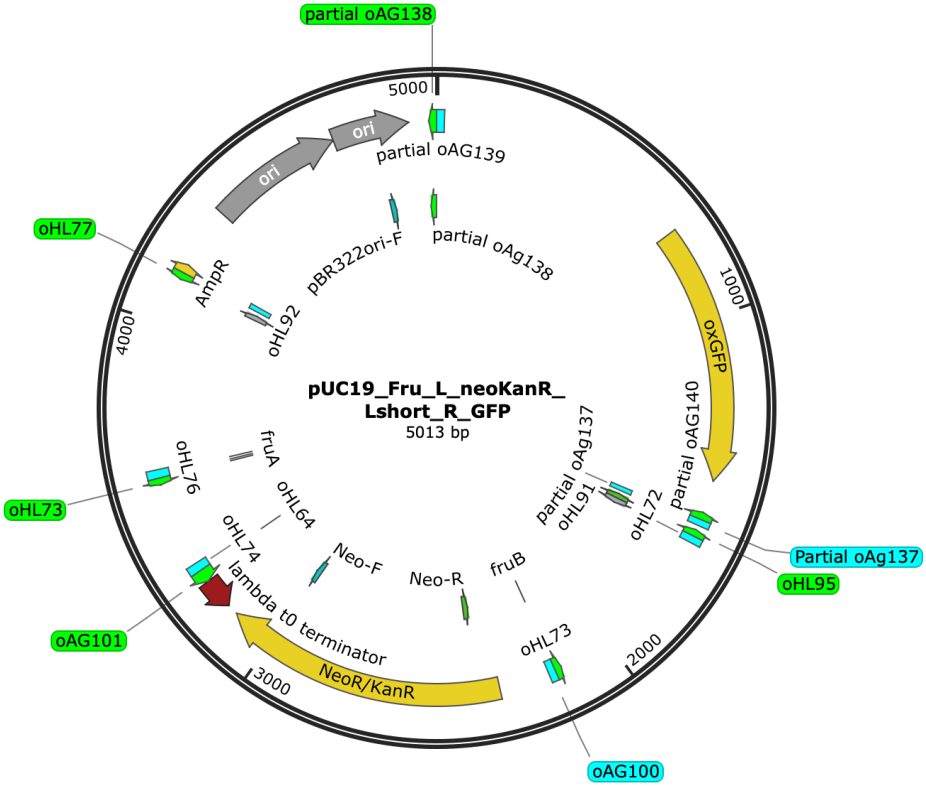

partial oAG138
5000
ori
ori
partial oAG139
oHL77
AmpR
pBR322ori-F
partial oAg138
oHL92
4000
1000
oXGFP
pUC19_Fru_L_neoKanR_
Lshort_R_GFP
5013 bp
partial oAG137
oHL91
partial oAG140
oHL72
Partial oAg137
oHL95
oHL73
oHL76
fruA
oHL64
oHL74
lambda t0 terminator
oAG101
Neo-F
Neo-R
fruB
oHL73
NeoR/KanR
3000
2000
oAG100

Figure 8 a) Schematic of the created circular fragment. Besides the L, Neo/KanR and N20PAM, L_short and R region, a PUC19 backbone and a superfolder green fluorescent protein gene(sfGFP) were mounted on the linear fragment as the backbone for Golden Gate Assembly. The blue arrows represent the oligos amplifying location of the circular fragment to create a single linear fragment. b) Example of a circular fragment design for the deletion of genes involved in Fructose metabolization, oligos oHL91and oHL92 were used to create the linear fragment. c) Example of a circular fragment design for the deletion of genes involved in Fructose metabolization after integrating sfGFP. Graph b), c) are from SnapGene.

To achieve better screening efficiency, the superfolder green fluorescent protein gene (sfGFP) was integrated into the backbone via Gibson assembly. A plasmid containing sfGFP was amplified with oligos oAG139 and oAG140, while the circular fragment created previously was amplified via PCR with oAG137 and oAG138. Then these products were mixed with Gibson mix at 50°C for 30min to produce a circular fragment with the fluorescent protein gene mounted on PUC19 in vivo, then the products were used directly for transformation or stored at -20°C.

Table 3 Oligos Sequences

Oligos name	Sequence (5'- 3')
oAG137	TTTGGTACCGAGAAACCGTCATCACCGAAACGC
oAG138	ACACTGGAAACACAGGGATAACGCAGGAAAGAACATG
oAG139	TTTCCTGCGTTATCCCTGTGTTTCCAGTGTTAGAATC
oAG140	TTTCGGTGATGACGGTTTCTCGGTACCAAATTCCAG

2.1.5 Colony PCR

In order to identify the transformants that inserted the editing cassette at the correct locus, two sets of primers were used in colony PCR (Fig. 9a, 9b). The forward primer is identical for both sets, which is complementary to the ~20bp sequences located upstream of the deleted region. The reverse primer for testing the correct insertion is homologous to the selection marker (Neo/KanR) and could be used to test the integration of all DNA cassettes designed to perform gene knockouts. Transformants with the correct editing cassette would show bands in gel electrophoresis at a length of approximately 1000bp. If no bands are observed, this may be due to

either unsuccessful transformants that did not insert the correct editing cassette, or colony PCR failure. To rule out the latter possibility, a control reverse primer is used running the same thermocycler process, it is homologous to parts of the deleted region, and should show bands in gel electrophoresis at a length of around 1000bp if the transformation failed.

To identify the marker–free strains after intra cellular recombination following cas9 induction, two primers are used in Colony PCR (Fig. 9c). The forward primer remains the same. The reverse primer is downstream of the right homology arm; the length of the amplified DNA fragment can be used to differentiate the recombination status.

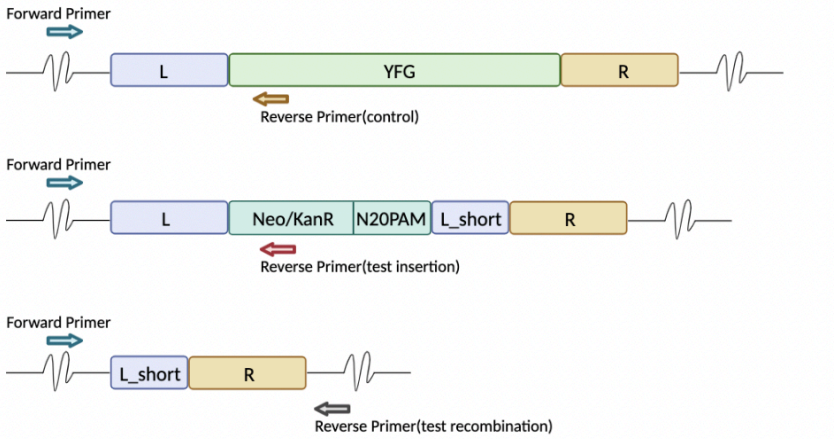


Figure 9 a) Regions of genes of interest in wildtype *E. coli* strains, your favorite genes (YFG) may represent any gene of interest. A forward primer is homologous to the upstream of the left homology region, a reverse primer is homologous to parts of the YFG. b) Regions of genes of interests after successful insertion of the linear fragment. A forward primer is homologous to the upstream of the left homology region, a reverse primer is homologous to parts of the Neo/KanR region. c) Regions of genes of interests after successful intracellular recombination of the linear fragment. A forward primer is homologous to the upstream of the left homology region, a reverse primer is homologous to downstream of the right homology region.

2.1.6 Genome editing procedure

The genome editing process is illustrated in Figure 10. Firstly, DH5a electrocompetent cells were transformed with the pCAGO plasmid. Then, the editing cassette DNA was transformed into electrocompetent cells by electroporation. Then, cells (~100 µL) were plated on an LB plate with ampicillin, Kanamycin and 1% glucose (to repress the expression of cas9), the remaining

solution was plated on a second LB plate with the same antibiotics and glucose concentration. The cells were cultured at 30°C for one and a half days, then eight colonies that did not show green or red fluorescence at the fluorescent microscope were chosen for further colony PCR. Two correct single colonies that passed double verification were picked and incubated with LB medium and ampicillin, IPTG, and L-arabinose at 30°C for more than 6h to express both Cas9 and the λ Red recombination system. Then 2uL of the medium was spread on LB agar plates with ampicillin. After another one and a half days of growth, eight colonies were picked to identify the correct transformants via colony PCR, then incubated in M9 medium with the corresponding sugar and plated on LB with kanamycin for additional verification. Successful deletion of genes was indicated by the absence of growth on these plates.

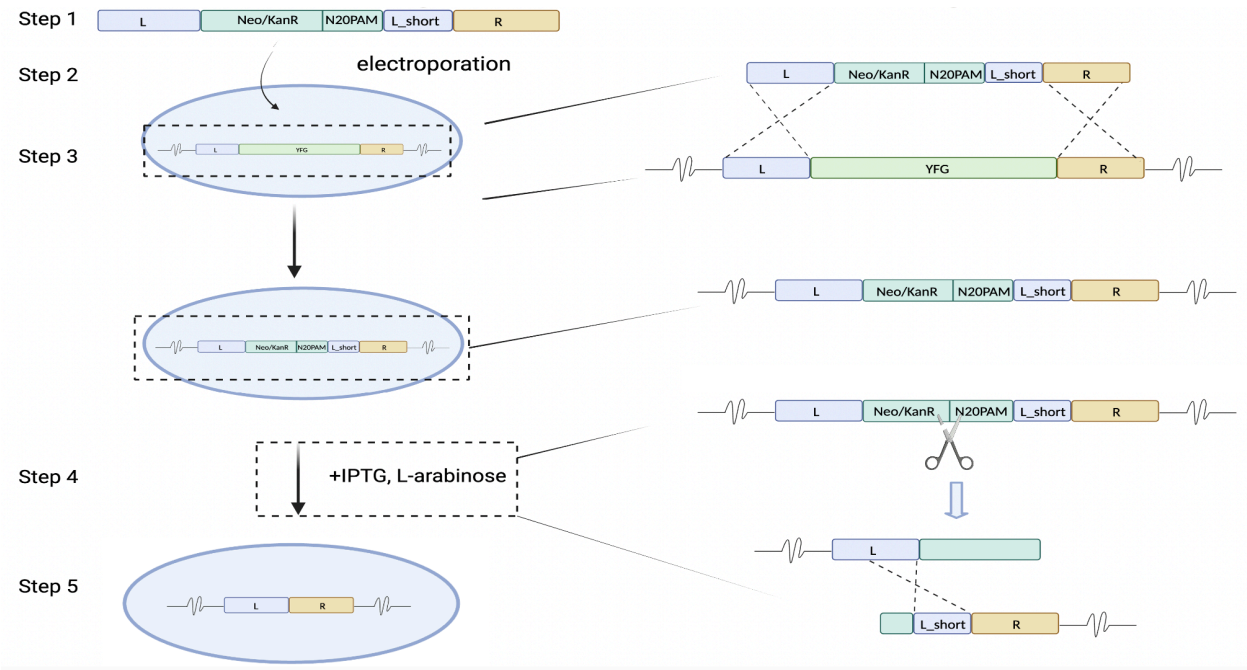


Figure 10 Overview of the genome editing procedure. Step 1, linear DNA cassette construct. Step 2 electroporation of the DNA cassette into the electrocomponent cells. Step 3, insertion of the linear fragment (intra-cellular homologous recombination). Step 4, induce IPTG and L-arabinose. Step 5, L and L_short homologous recombination.

2.2 Materials and Protocols

2.2.1 Strains and culture conditions.

E. coli DH5α was used as a cloning host, DH5α cells harboring the pCAGO plasmid were grown in lysogeny broth (LB) at 30 °C to maintain the plasmid. Kanamycin (x1000) and ampicillin (x1000) were added to the media where appropriate. 1% (w/v) glucose was added to the cultures to repress the expression of *cas9*, L-arabinose (2g/L final concentration) was used for *cas9* induction, and IPTG (0.1 mM final concentration) was used to induce the λ Red system.

2.2.2 Golden Gate Assembly Materials and Procedure

A total volume assembly reaction mixture was 15ul. Assembly parts were all equimolar with the Neo/KanR-N20PAM fragment. The mass of Neo/KanR and N20PAM part was determined so that the minimal added Deionized water volume was achieved. Reaction reagents were: 1.5uL of T4 DNA Ligase Buffer 10x, 0.15 uL of T4 ligase, 0.45 uL BasI-HFv2. The assembly reaction was performed in a thermocycler following New England Biolabs (NEB) recommendations, as described in the table below.

Table 4 Golden Gate assembly thermocycler procedure

Temperature	Time	Cycle
37 °C	5 min	30 cycles
16 °C	5 min	
60 °C	5 min	1 cycle
4 °C	indefinite	

2.2.3 Preparing Electrocompetent *E. coli* cells expressing λ Red (small batch, pCAGO plasmid)

The frozen glycerol stock of DH5α was streaked onto an LB plate with ampicillin (0.1 mg/mL) and grown overnight at 30°C. Then, a single colony was selected from the plate and inoculated in 3 mL LB with ampicillin in a glass tube for another overnight at 30°C. *E. coli* cells were then grown in 30 ml of LB medium with ampicillin at 30 °C until reaching an OD600 of 0.3 ~0.35, then IPTG (100μM) was added to induce the λ Red system, and the cells are washed with water and 10% ice-cold glycerol. Then the electrocompetent cells are divided into 50 uL aliquots and stored at -80°C, or transformed immediately.

2.2.4 Transforming *E. coli* by Electroporation

The prepared electrocompetent aliquots were first thawed on ice and mixed with ~100ng of the editing cassette DNA in a chilled 1-mm Gene Pulser cuvette (Bio-Rad). After electroporation at 2.5 kV, the cells were immediately resuspended in 1 ml of ice-cold Super Optimal broth with Catabolite repression (SOC) medium with 5ul $MgCl_2$ to recover the cells. The cells were incubated for 1 hour at 30 °C, and then ~100uL was spread on LB agar plates with ampicillin, kanamycin and 1% glucose if necessary for overnight incubation.

3. Results and Discussion

3.1 Constructing the linear fragment was problematic therefore impeding the genes knockouts, adding fluorescent genes helped screening false positives transformants

In our hands, the methodology suggested by Zhao et al.[16] for the assembly of L, Neo/KanR and N20PAM, L short and R as a single linear fragment was problematic. Running the assembled fragments via DNA gel electrophoresis revealed the successful creation of the correct linear fragment (with lengths typically around 3000 bp) after Golden Gate assembly. However, unidentified bands were also observed, possibly due to partial assembly. This circumstance persisted despite modifications to the reaction conditions, such as altering melting temperature (Tm) and polymerases (NEB OneTaq, NEB Q5 and APExBIO Phusion), adjusting the reactant concentrations and reaction time to align with manufacturer recommendations, and amplification of the target fragment.

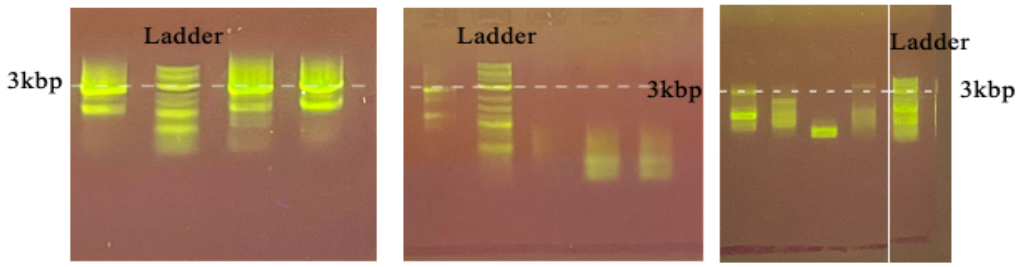


Figure 11 Illustration of gel electrophoresis results after amplifying different linear DNA cassettes designed for knockouts. Different figures represent various adjustment of the method. a) Phusion polymerase. From left to right: Man, Gal and Rib fragments b) Higher melting temperature. From left to right: Man, Gal, Rib, Fru fragments c) NEB specification of reactant concentrations. From left to right: Man, Gal, Rib, Fru fragments. Different columns represent different linear fragments constructs (Man= Amplified linear fragment designed to knockout *ManXYZ*, Gal = Amplified linear fragment designed to knockout *galETKM*, Rib = Amplified linear fragment designed to knockout *rbSDABKR*, Fru= Amplified linear fragment designed to knockout *fruBKA*. The dashed line represents the DNA length of 3000bp. The ladders represents the locations of different DNA lengths.

Given the circumstances, we attempted to use the product containing multiple bands for the subsequent transformation, provided that the fragment of the desired length existed. However, the transformation efficiencies were unsatisfactory for various linear gene knockout cassettes.

The transformants only exhibited macroscopic (colony) growth after nearly 60 hours. Furthermore, despite multiple attempts, transformants that correctly integrated the linear assembly could not be identified from the colony PCR(Fig. 12). In light of these challenges, our primary objective would be to ensure the successful creation of the single linear fragment for subsequent transformation. Therefore, we assessed two potential solutions.

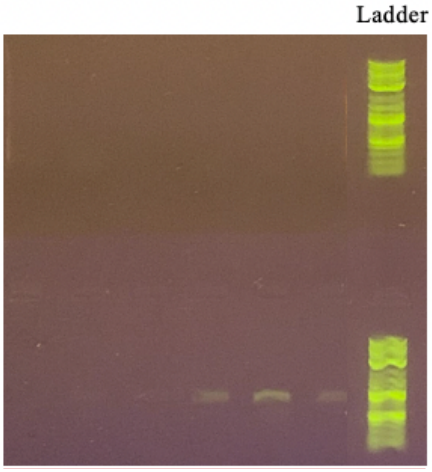


Figure 12 Gel electrophoresis results following colony PCR. Transformants were picked after transformation with the product having multiple bands. The upper region shows bands if transformation is successful, and the linear fragment has integrated into the knockout region. In this case no bands shown. The below region shows bands if transformation is unsuccessful, in this case five out of six showed bands, representing unsuccessful integration of the linear fragment.

Solution 1: PCR amplification to create linear fragments

We opted for a simpler approach to achieve linear fragment creation, using a PCR amplification with fewer steps and a shorter reaction time (Fig. 6). In this method, the left and right homology (L- 65bp, L short-40bp and R-65bp) were designed as the overhangs to amplify the Neo/KanR sequence. The resulting linear fragment was shorter, around 1000 bp in length. However, at least a 65bp length at each side of the homology was needed for efficient recombination, which required time-consuming production of long oligos. Consequently, this was not considered as the preferred approach.

In our trial, we designed DNA cassettes to delete genes involved in mannose and galactose uptake and catabolism, specifically, *manXYZ*, and *galTKM*. Our preliminary results indicate a higher success rate of transformation with the mannose fragment, compared to the galactose one. At the time of writing, a verified insertion was identified for the mannose genes knockout. We were unable to delete genes involved in the uptake and catabolism of galactose genes using this method.

Solution 2:

Assembly of the linear DNA cassette via Golden Gate cloning, using a plasmid backbone.

The transformation efficiency with a linear DNA fragment is usually less than that of a circular fragment due to exonucleolytic degradation of the incoming DNA[40]. With that bear in mind, we leveraged the advantages of the circular fragment to ensure the successful creation of the linear DNA fragment required for gene knockouts using the pCAGO methodology. The product of a Golden Gate assembly including the L, Neo/KanR-N20PAM, Lshort, R fragments and a plasmid backbone, this plasmid were transformed in wildtype DH5a without pCAGO. After plasmid extraction and sequencing, the product was amplified with oligos homologous to the beginning and end of the L and R fragments, respectively. Then the linear fragment was created and ready for transformation.

Although this methodology significantly increased the number and growth rate of transformants, most transformants were false positives, suggesting that the plasmid assembled via Golden Gate, rather than the linear DNA fragment, successfully transformed into the cells. Although digestion with DpnI was used to remove the plasmid after amplification of the linear cassette, the higher efficiency of plasmid transformation may explain why most transformants were false positives,

they may have been transformed with residual, undigested plasmids. Such an explanation is in line with our observation that false positive transformants showed antibiotic resistance, but no corresponding bands in gel electrophoresis after colony PCR. To overcome this issue, we incorporated different fluorescent protein genes in different plasmids to enhance the screening efficiency. The mRFP1 gene was introduced in the plasmid from which Neo/KanR was amplified from, and the sfGFP gene was introduced in the backbone used to assemble circular fragments via Golden Gate. In this way, we could identify transformants that were transformed with either the original plasmid from which the resistance marker was amplified, or the Golden Gate assembly plasmid. Under the microscope, the percentage of transformants showing either red or green fluorescence showed great variability (Table 5), ranging from around 0-3% to 70%, which proves our assumption that even with the help of DpnI, circular DNA fragments could have high efficiency of transformation. In addition, we found that fluorescence screening perfectly aligned with the colony PCR results. In conclusion, we improved the pCAGO assembly method by including a fast fluorescence screening that allows us to quickly disregard false positive transformants that integrated circular, rather than linear DNA fragments.

Table 5 Average number of transformants that showed fluorescence from random picked sections in different transformants. (Percentage = number of colonies showed fluorescence / number of total colonies observed*100%), 0-3% of fluorescence transformants in bHL20 showed that the count number could not be presented in precision.

Strains	Percentage
bHL17 *(Δfru**BKA**::Neo/KanR-N20PAM*)	69.6%
bHL18 (Δ*rbSDABKR::Neo/KanR-N20PAM*)	45.1%
bHL19 (Δ*GalTKM::Neo/KanR-N20PAM*)	33.9%
bHL20 (Δ*yoaL-ManXYZ-yobD::Neo/KanR-N20PAM*)	0-3%

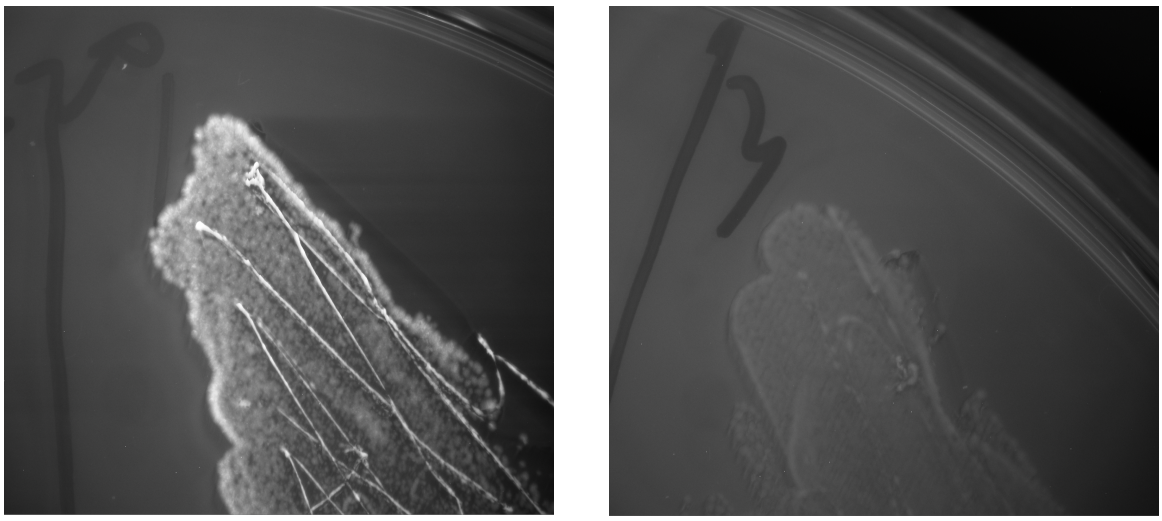


Figure 13 Spread of colonies showed fluorescence(left) and no fluorescence(right) in microscope. The two images correspond to different single colonies picked after transformation of an assembled linear fragment.

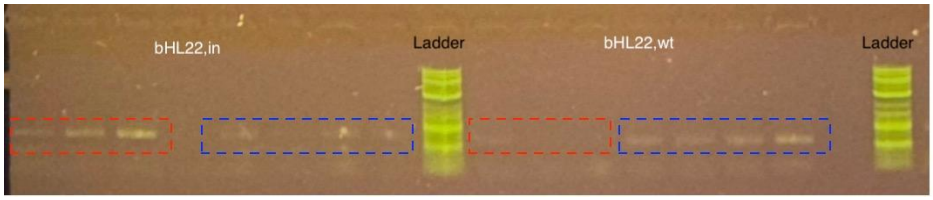


Figure 14 Gel electrophoresis following colony PCR of knockout transformants after identifying the fluorescence colonies under microscope. Strain name: bHL22(*ΔfruB, fruK, fruA (*Bands in red rectangle: colonies identified not showing fluorescence. Bands in blue rectangle: colonies identified showing fluorescence. In: Identifying the insertion of the linear fragment without a successful recombination to excise the Neo/KanR cassette. wt: control, Identifying the possibility of colony PCR failure .

3.2 Successful deletion of genes involved in maltose and galactose catabolism.

At the time of writing, we successfully created the knockout of genes involved in the metabolism of maltose (*malG, malF, malE, malK, lamB, malM*) and galactose (*galT, galK, galM*). The map below shows the corresponding loci before and after deletion. Gel electrophoresis following colony PCR of various transformants are shown below. These results show that the target genes were successfully deleted in the transformants, and no selection marker was left in them.

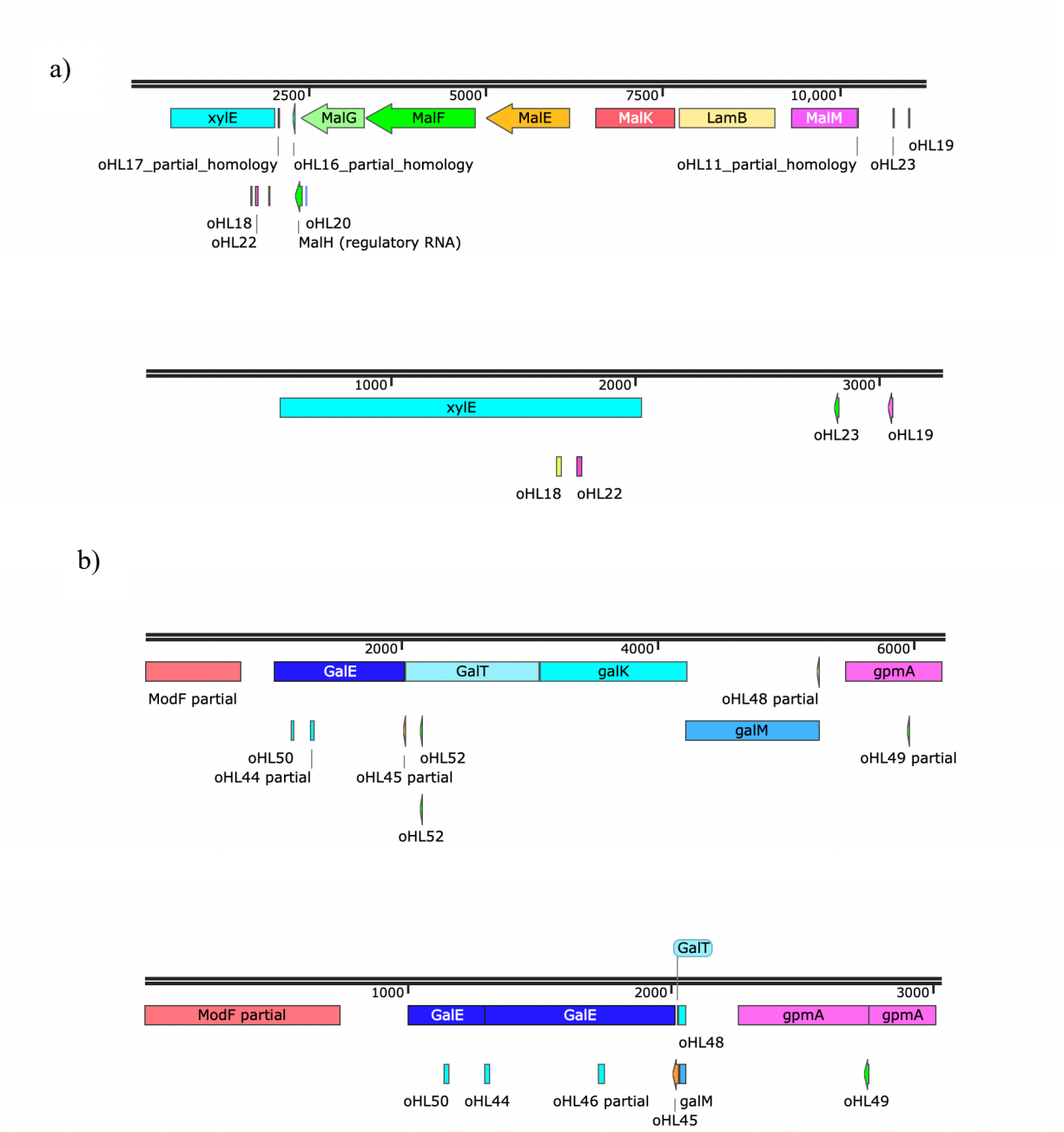


Figure 15 Genome region before(above) and after(below) the knockout and recombination to excise the selection marker. a) Genes involved in maltose catabolism targeted for knockout, b) Genes involved in galactose catabolism targeted for knockout.

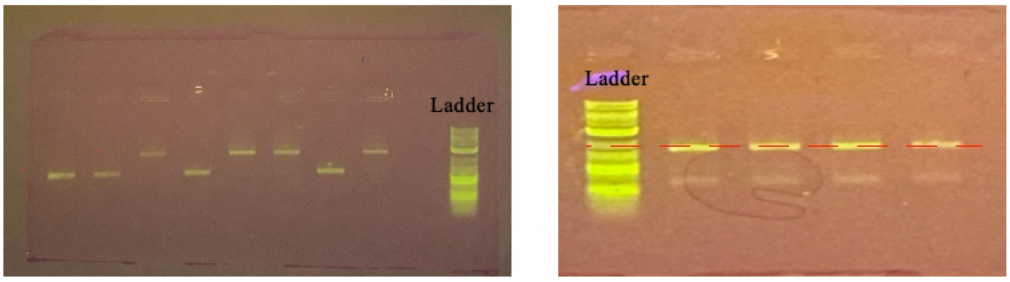


Figure 16 Gel electrophoresis following colony PCR of knockout transformants after induction with IPTG and L-arabinose to excise the resistance marker a) Deletion of *malG, malF, malE, malK, lamB, malM*, the longer bands(2734bp) represent the successful insertion of the linear fragment without a successful recombination to excise the Neo/KanR cassette, while the shorter bands(1739bp) represent the successful scarless knockout of target genes following recombination of the L and L short fragments. B) Deletion of *galT, galK, galM*. The band at 1674bp shows that the fragments L and L short recombined to excise the Neo/KanR cassette. In this picture, four different single colonies achieved the desired scarless knockout.

After identifying the correct transformants via colony PCR, we conducted a further verification. The strains identified above were grown in an LB plate with kanamycin, as well as M9 minimal medium with a cognate carbon source provided. The observed phenotype corresponded with the colony PCR results: The correct transformants did not grow on plates with kanamycin, nor in M9 minimal medium with only maltose or galactose provided, respectively.

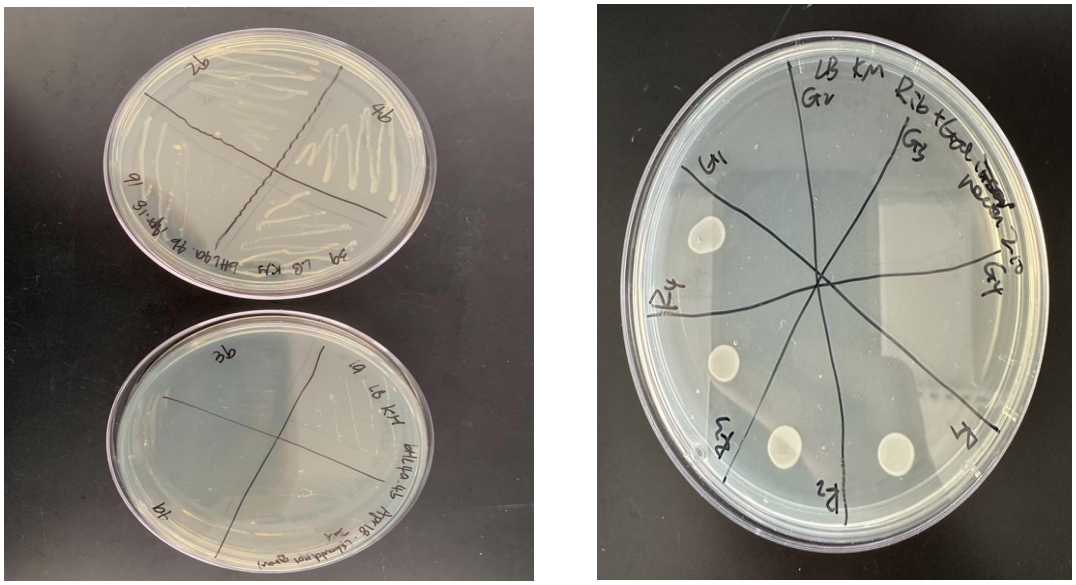


Figure 17 Colonies in LB plates with kanamycin. A) Engineered strains that have deletion of genes involved in Maltose intake. This corresponds with colony PCR results, where the correct ones shown in colony PCR did not grow and the incorrect ones did grow. B) Engineered strains that have deletion of genes involved in galactose intake. The correct ones shown in colony PCR (G1 – G4) did not grow whereas the incorrect ones(R1-R4) did grow.

To further investigate the growth rate of these transformants, the knockout strains were cultured in 96-well plates (Fig. 16). When maltose/galactose were provided as the single carbon source, strains did not grow consistent with expectations. In the presence of M9 minimal medium and glucose, impaired of growth were noticed the engineered strains. It took both bHL6(Δ*malGFEK,* Δ*lamB,* Δ*malM)* around 10 more hours and bHL21*(*Δ*galTKM)* around 15 more hours to reach mid-log phase compared with Wildtype *E. coli* in the presence of glucose. In general, it took the engineered strains around 30 hours to start increasing OD, and the highest OD reached by bHL21 was 0.25, compared to an OD of 0.3 of wide type strains.

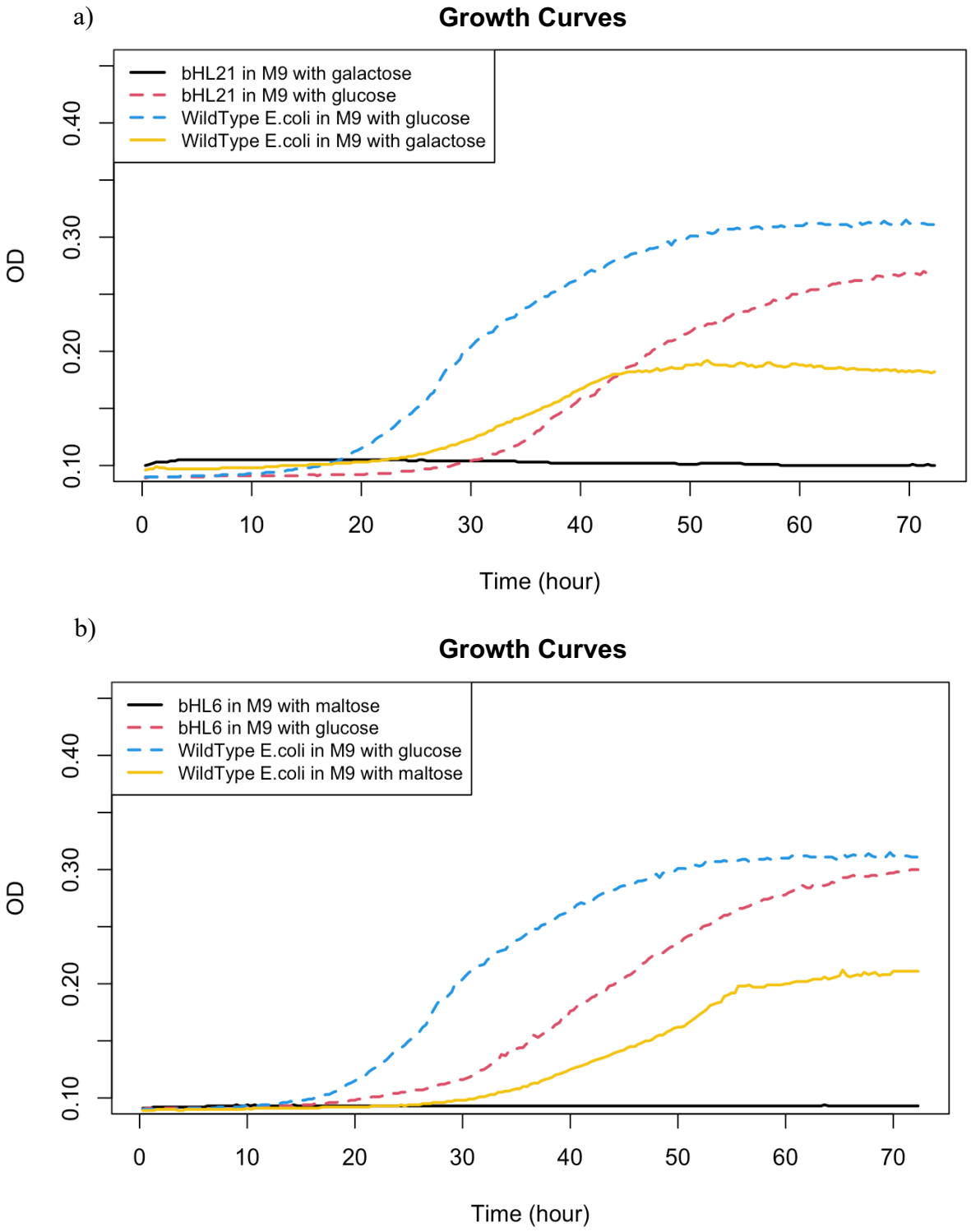


Figure 18 Growth curves of strains with successful deletion of genes involved in a) maltose metabolism. b) galactose metabolism. They were cultured in 96-well plates in the presence of glucose and corresponding carbon source. Wildtype *E. coli* were used as a control.

3.3 Non-scarless knockouts of genes involved in the metabolism of fructose, mannose and ribose were successful, yet removal of the selection marker has not been completed.

At the time of writing, the insertion of the linear fragment and non-scarless knockouts of genes involved in the metabolism and uptake for fructose, mannose and ribose were successful (Fig. 19). However, despite inducing IPTG and L-arabinose, the second recombination of L and L_short to remove the selection marker failed several times. This may indicate a weak recombination efficiency using the pCAGO method.

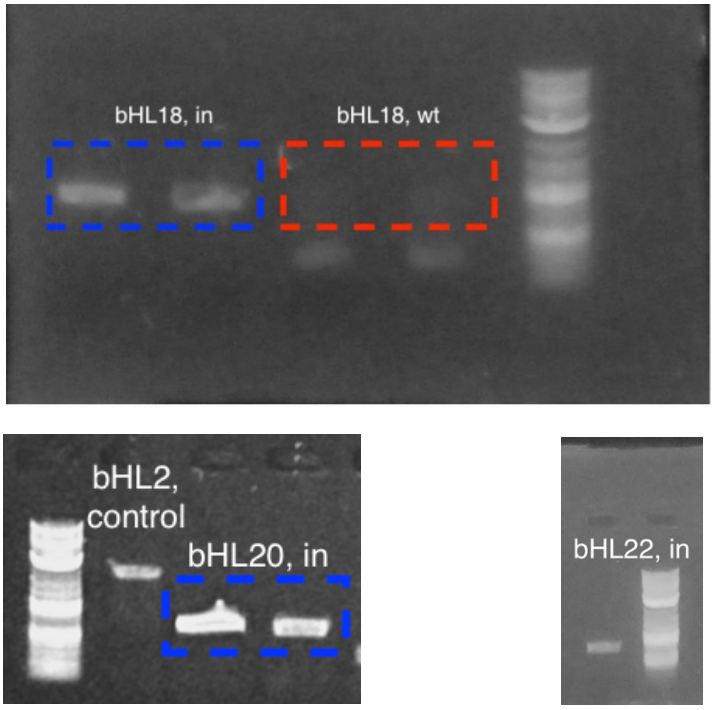


Figure 19 Gel electrophoresis following colony PCR of knockout transformants after induction with IPTG and L-arabinose to excise the resistance marker. a) Deletion of *rbsD, rbsA, rbsC, rbsB, rbsK*. Strain name: bHL18. b) Deletion of *manXYZ*, strain name: bHL20. c) Deletion of *fruB, fruK, fruA*. Strain name: bHL22 (In: Represents the successful insertion of the linear fragment without a successful recombination to excise the Neo/KanR cassette, wt: control, represents possibility of colony pcr failure if showed band, bHL2, wildtype *E.coli*.

We further checked the growth rate of these transformants. When fructose was provided as the single carbon source, strains showed a weak but increased growth rate. This was expected, given that fructose can be uptaken by *E. coli* via three routes[41, 42]. The predominant one involves the function of phosphoenolpyruvate: glycose phosphotransferase system (PTS), and fructose operon

which include *fruA, fruK* and *fruB*. However, in our case, this route is inaccessible given that we successfully deleted these genes. The second route involves the PTS, as well as the membrane-spanning proteins that recognize a variety of sugars processing the 3,4,5-D-arabino-hexose configuration, specified by *manXYZ*. For this route to be viable, fructose has to be supplied in relatively high concentrations(>2mM). We meet such criteria, given that a 22mM concentration is provided in culture. The third route devoid the first two routes and fructose diffuses into the cell via an isoform (PtsG-F) of the major glucose permease of the PTS. However, such mutants need to have elevated levels of mannosidase (Mak+) to allow the fructose uptake via membrane-spanning protein, in this case it is unlikely. In addition, according to the flux Constraint-Based Reconstruction and Analysis (COBRA) method[43], to achieve a complete incapability to metabolize fructose, *manXYZ* or *pgi, ptsH, malK* need to be deleted.

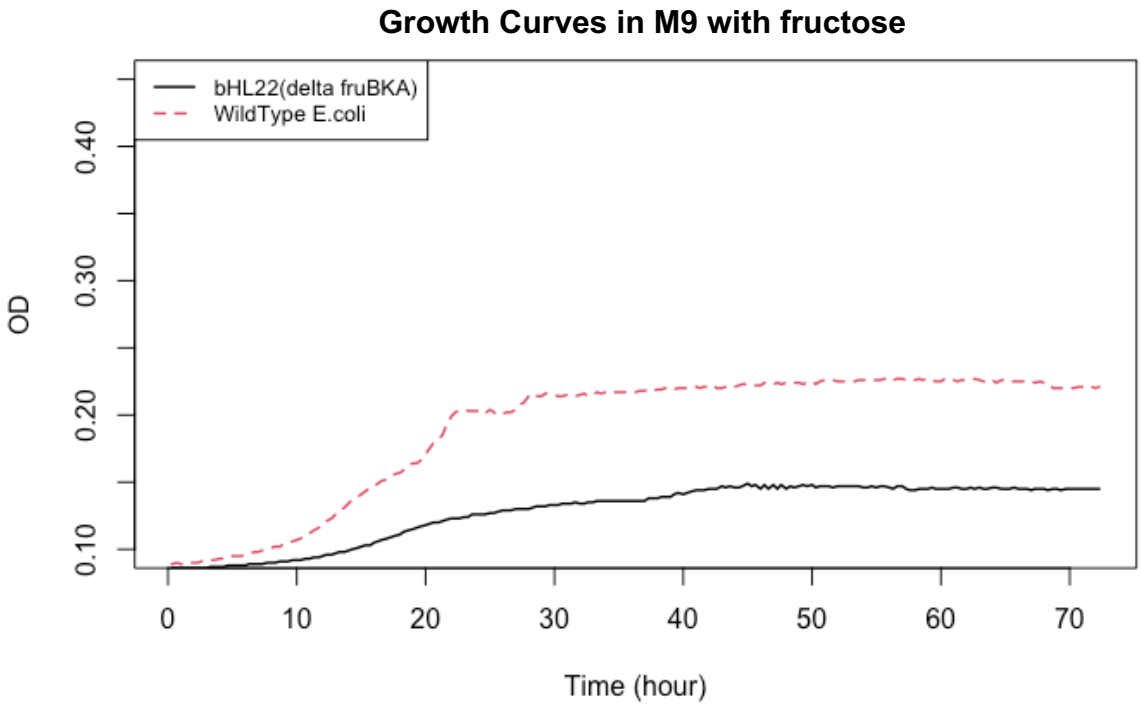


Figure 20 Growth Curves of strains of non-scarless knockout of genes involved in Fructose metabolism(bHL22). bHL22 were cultured in 96-well plates in the presence of Fructose paired with different carbon sources.

Likewise, in addition to *rbsABC,rbsD,rbsK,* which were removed, *alsABC* and x*ylFGH* need to be deleted due to their ability to transport ribose. Previous studies showed the deletion of *manXYZ* cause a defect on sole mannose media growth [44], and a complete incapability to grow on sole mannose media can be achieved when *ptsG* is also deleted. In scope of time, theses abovementioned deletions will be the future work.

3.4 Improved pCAGO methodology

As mentioned in Section 3.1, the unsatisfactory performance of the pCGAO method hindered us from performing the knockouts. Difficulties in creating the linear fragment designed for the gene knockouts resulting in high proportions of false transformants.

This section highlights a notable improvement in the scarless gene knockout method[16]. As illustrated in Figure 20, we improved the screening efficiency via incorporation of fluorescence protein genes. The construction of circular fragment containing an origin of replication ensure the successful creation of the linear fragment required for gene knockouts. In addition, the incorporation of the mRFP1 and sfGFP into the plasmids for amplification significantly enhanced the screening efficiency of the false transformants. Upon the growth of transformants, colonies exhibiting green or red fluorescence were discarded during the microscope screening. This fast and simple screening process perfectly aligned with the colony PCR results, helping us target the correct transformants. In conclusion, in contrast to the conventional approach proposed by Zhao et al.[16] of constructing a linear fragment and directly transforming it into E. coli, which yields few correct transformants, the method has proven efficient and time-saving.

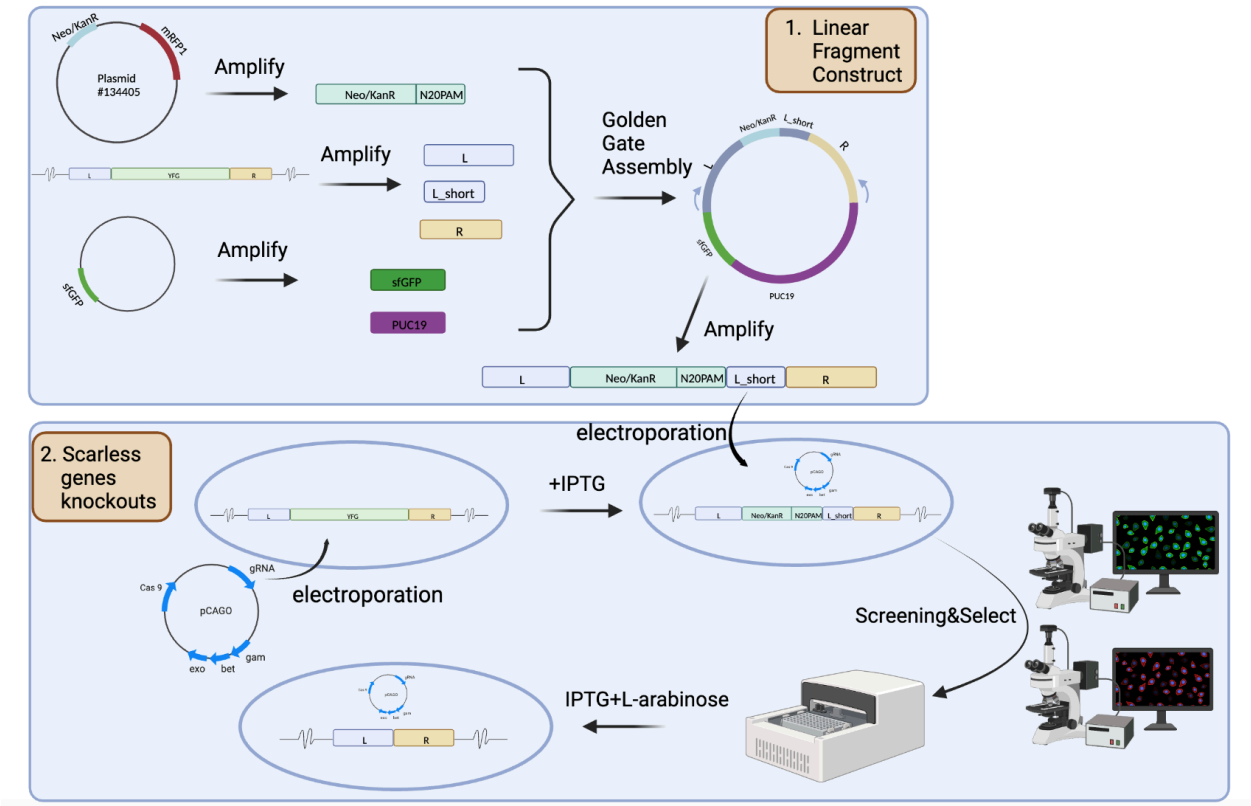


Figure 21 Schematic of genome editing process with improved pCAGO method. This is divided into two main parts. The first part is the linear fragment construction. The Neo/KanR region is amplified from the plasmid that have fluorescence protein genes(mRFP1), the linear fragment is amplified from the circular fragment that contains sfGFP, from which constructed via Golden Gate Assembly. The second part includes the pCAGO plasmid transformation, and then induced with IPTG to express the λ red system. Then the linear fragment created in part 1 is transformed into DH5aplha with pCGAO. After screening the fluorescence colonies via microscope, colony PCR was applied to verify the successful transformants. Then the editing cassette that was inserted at the target locus will be induced with both IPTG and L-arabinose to induce the expression of λ Red and CRISPR/Cas, which facilitated the removal of the editing cassette and yielded the final edited locus.

Chapter 2 Metabolism and coexistence

The main objectives of this chapter were to i) assess how catabolite repression affects the growth of *E. coli* metabolic knockouts in mixed-sugar environments and ii) explore how the metabolic requirements of different strains determine if and how they may coexist when competing for a shared set of resources. We investigated these two objectives using the Keio knockout collection. In exploring objective i), we discovered that maltose uptake is inhibited by the presence of either extracellular D-sorbitol, or intracellular D-sorbitol-6-phosphate.

Due to the difficulties encountered in Chapter 1, which involved improving a method for Cas9-assisted gene editing in *E. coli*, progress towards objective ii) was limited and will be the subject of future work.

1. Introduction

1.1 Catabolite repression

In the presence of multiple carbon sources, *E. coli* can either co-metabolize them or preferentially use one over the others[45]. In most cases, sequential sugar utilization occurs due to carbon catabolite repression (CCR)[46], which is a global regulatory mechanism that inhibits the expression of genes involved in metabolizing the less preferred sugars. Glucose repression is a well-known example. When glucose and lactose are both present, lactose is not utilized until the glucose concentration is low or almost depleted. This is because the Lac operon, the operon for lactose utilization, is regulated by a promoter system that responds to glucose levels. In addition, CCR has been observed to occur with sugar mixtures not involving glucose[47]. Therefore, a comprehensive understanding of the underlying molecular mechanisms of this process is crucial.

Previous studies have made progress in understanding the molecular mechanism underlying CCR, particularly for sugars transported by the phosphotransferase system (PTS). For instance, Brückner and Titgemeyer[45] demonstrated that PTS sugars lower the activity of the cAMP receptor protein (CRP), leading to inactivation of the utilization of alternative carbon sources. However, the metabolic regulation of microbial consortia that occurs when cultivated on mixtures of sugars (especially non-glucose ones) is less understood, even in model organisms such as *E. coli*. Aidelberg et al.[48] studied the hierarchy of non-PTS sugars and discovered that both sequential and simultaneous utilization of carbon sources can occur. Their study was mainly focused on modeling the relationship between promoter systems and cAMP levels. Understanding if and how catabolite repression affects the coexistence of multiple bacteria strains in a consortium is a very active research topic. Xia et al.[14] investigated how the simultaneous utilization of lignocellulosic hydrolysates affected the coexistence of engineered *E. coli* strains, but their work is limited to sugars in lignocellulosic hydrolysate.

The mechanism underlying the metabolic interaction of microorganisms in the natural environment remains incomplete, and an analysis of an *E. coli* community in a mixed carbon source environment still remains a challenge[15]. To address this issue, we constructed the library of *E. coli* strains described in Chapter one. Future work in this project will perform additional knockout within the knockout strains already developed, to produce strains that cannot metabolize multiple sugars. This approach will allow us to investigate how coexistence is affected by metabolic preferences, knowing exactly the metabolic preferences and requirements of each strain. Additionally, once we understand the intricacies of catabolic repression, we can design strategies to optimize the utilization of multiple carbon sources by microbial consortia, which has numerous applications in biotechnology and industrial microbiology.

1.2 Species coexistence concept and current research

The Earth is home to an estimated 8.7 million species[49]. In humans' guts, an estimated 200 to 1,000 different species exist[50]. Despite the astonishing diversity of life on our planet, there has been a conceptual challenge to explain such diversity, known as the "paradox of plankton", that is, in most mathematical models of community dynamics, the number of coexisting species cannot exceed the number of resources available[51]. Yet, in most marine ecosystems, hundreds or more species of phytoplankton can coexist while competing for a very limited number of resources.

To address this paradox, Posfai et al.[12] built a resource-competition model that introduced trade-offs between species' resource utilization abilities. Their model showed that an unlimited number of species can coexist on a finite set of resources, contradicting traditional resource competition models that did not account for trade-offs in the metabolism of different resources. The model by Posfai et al.[12] is based on the mathematical analysis of consumer-resource models that include equations for the abundance of different species and the concentration of different resources. Without going into the details of the mathematics, the model can be explained using a geometrical description. For example, in their model (Fig. 22), one can mix three species with distinct metabolic strategies and investigate the conditions that lead to their coexistence. Each vertex in figure 22 denotes a specific resource. The proportion of supplemented resource concentrations in the system is represented by the distance between the black diamond and each vertex. The closer to the vertex, the more resource of that kind is provided in the system. For instance, when the black diamond is located at the midpoint, it indicates an equal amount of each resource is provided, while a location at the vertex signifies that the system is only supplied with one resource. Each colored point denotes a different species, which have different metabolic

strategies. Their abilities to utilize different resources are defined by their specific distribution of enzymes for resources utilization, which represented in this triangle are the distances between the colored point and each vertex. The shorter the distance to the vertex, the more inclined they are in utilizing the corresponding resource. In a theoretical investigation, Posfai et al.[12] observed that extinctions happen when the concentrations of certain nutrients fall below a threshold for certain species, causing their populations to decay toward zero. They used this and other observations to divide the triangle into sectors representing different coexistence scenarios based on the location of the resource supply (black diamond) and the convex hull of these colored points. In Figure 22, the red/green/blue species drives both of the others to extinction if the black diamond lies in the red/green/blue regions. The red & green/green & blue/blue & red species eradicate the third species if the black diamond lies in the orange/cyan/purple region. All three species coexist if the black diamond lies in the gray region. However, the prediction of this model lacks practical experimental testing and validation. Obtaining such experimental tests has been difficult, given that the metabolic relations between strains and recourses in natural communities are unclear and hard to dissect.

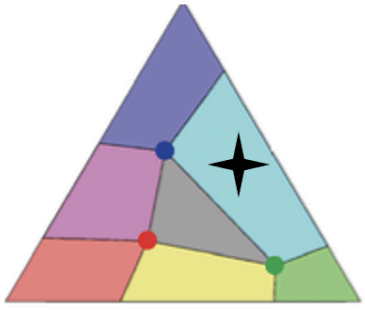


Figure 22 Three-species competition model developed by Posfai et al. [12] Each vertex denotes a specific carbon source. The black diamond represents the proportion of resources that this system provides. Each colored circle denotes each mutant strain and the distance between the colored point and each vertex represented their ability to utilize different resources.

Testing the predictions of this model using microbial communities from the wild is difficult, due to the difficulty of understanding the metabolic requirement of wild strains. Thus, we aimed to test the predictions of this model by engineering *E. coli* strains that have a prescribed set of carbon sources preferences. In this model, strains of our knockout library should be located somewhere at the edges of the triangle, a location determined by their ability to utilize different carbon sources, and inability of utilizing others. Co-culturing the knockout strains and varying the resource supply rate would allow us to test the coexistence criteria of Posfai et al.[12] For example, when the supply rate is within the convex hull of the rescaled uptake rates, coexistence should happen, while the supply rate located outside the convex hull, certain mutant strains should not be about to survive. However, as shown in chapter 1, the issues we encountered hindered us from getting a complete library of knockout strains. To get some intuition for the issues we might encounter in such an approach, for example a possible role for catabolite repression in controlling strain coexistence, we turned to utilizing strains from the *E. coli* Keio knockout collection[34] to investigate their coexistence in mixed sugar environments. The Keio collection contains a set of single-gene knockout mutants for all nonessential genes in *E. coli* K-12. However, it has not been adopted at first due to the fact that certain strains within the collection are still capable of importing carbon sources that they cannot metabolize. This complicates the analysis and therefore poses a challenge in the interpretation of the co-culture data.

Tong et al.[33] employed the Keio collection strains and conducted a comprehensive genome-scale investigation of the impact of gene deletion on growth in different carbon sources. They compiled their findings into Carbon Phenotype Explore (CarPE) (https://edbrownlab.shinyapps.io/CarPE/). In this website, users can explore experimentally

measured growth curves of *E. coli* strains by manipulating the targeted genes and carbon sources environment. This resource provided guidance in identifying the essential genes for the metabolism of various carbon sources. Based on the growth curves from CarPE, we identified twelve strains that were deleted with one of the chosen genes, and all of them were unable to grow when provided with specific sugars as their only source of carbon. By growing these strains in the presence of two or more carbon sources, we found that most mutant strains were not or only slightly affected by the presence of the other carbons sources, except for the sorbitol deletion mutant with maltose. After further investigation of the growth rate of these strains, *ΔgalK, ΔmalG*, and *ΔxylB* were mixed and cultured for a five-growth dilution cycle with varying proportions of carbon sources provided. To quantify the proportion of surviving strains during culturing, we transformed them with plasmids containing three different fluorescence protein genes (Venus YFP, mRFP1, cyOFP1), before culturing. The data collection is still ongoing and the results will be presented during my M exam.

2. Method

2.1 Choice of strains

The growth curves provided in CarPE allowed us to identify individual genes that are essential for the metabolization of different sugars. Specifically, the deletion of these genes from the *E. coli* genome led to an inability of the microbe to grow when exclusively supplied with the corresponding carbon source. For instance, the deletion of *galK* gene resulted in the incapacity of *E. coli* to grow when only galactose was provided, as shown in the flat growth curve in CarPE (Fig. 23). Using this selection criterion, we obtained twelve strains from the Keio collection library, each of which was deleted with a single gene that is essential for the metabolism of a specific carbon source. The knockout mutants are *ΔaceA, ΔxylA, ΔxylB, ΔmanY, ΔmanZ, ΔmalQ, ΔmalG, ΔmalE. ΔgalT, ΔgalK, ΔsrlD, Δrbks*.

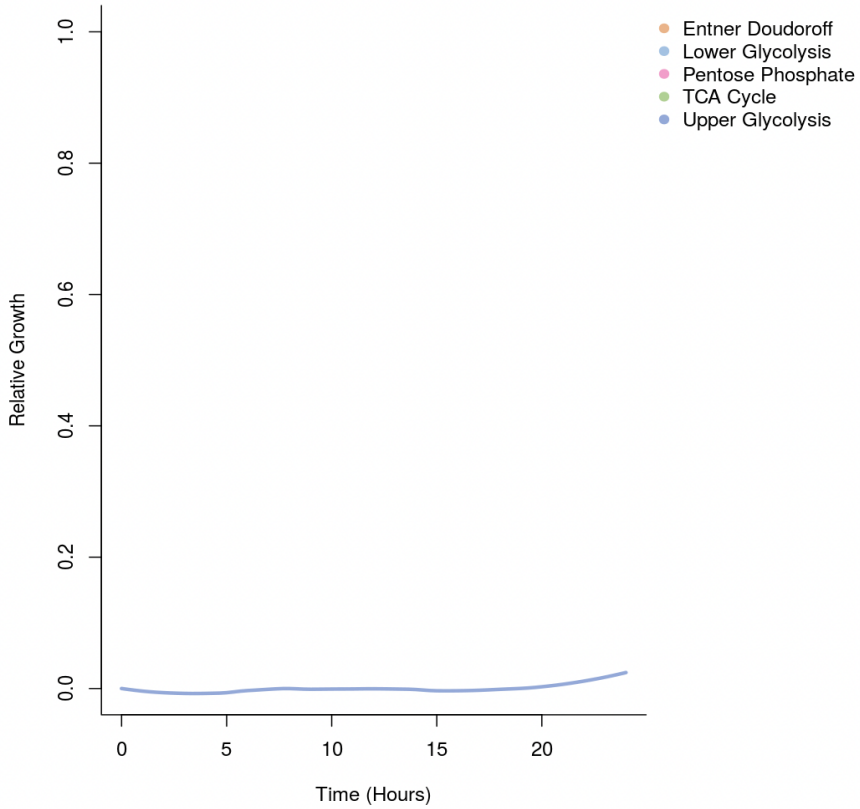


Figure 23 Growth curve of ΔgalK when only galactose was provided, picture from CarPE.

2.2 96 well-plate culture

2.2.1 Phenotype verification

Following the resuspension of cultures in PBS (Chapter 2, section 2.5.2), 5 uL of the suspension were inoculated into a single well of a 96-well plate containing 150 uL M9 minimal medium and a carbon source (0.2% w/v) that the strain was unable to metabolize. For instance, the strain with *xylA* deleted was cultured in M9 minimal medium supplemented with xylose only. Then the plate was placed in a plate reader at 37°C for 48 hours to monitor cell growth.

2.2.2 Paired carbon source culture

The phenotype verification results showed a general incapability for the chosen *E. coli* Keio collection strains to grow when the corresponding (single) carbon source was provided, in agreement with CarPE, except *for strains ΔmanY, ΔmanZ, ΔmalE*. Therefore, we focused our investigation on the remaining knockout mutants.

Following the resuspension of cultures in PBS (as outlined in procedure 2.5.2), 1.5uL of PBS were inoculated into a single well in 96-well plate that contained 150uL of M9 minimal medium without any carbon source. Next, each strain was provided with two types of carbon sources (0.2% w/v), one of which was the carbon source that each strain was unable to metabolize, while the other was one of the remaining four carbon sources (Table 6). For instance, the strain with *galK* deleted was grown in M9 medium supplemented with galactose and one among maltose, ribose, sorbitol, or xylose, as well as in M9 medium containing a mixture of all these carbon sources (each carbon sources have the same final concentration 0.2% w/v). The 96 well-plate was grown in a plate reader at 37°C with agitation for 72 hours.

Table 6 Categories of carbon sources provided for different strains from the Keio collection. 0.2% w/v of each carbon source were provided in each well according to the table. G= galactose, M = maltose, R = ribose, S= sorbitol, X=xylose

Strain Types	*ΔgalK*	*ΔmalG*	*ΔmalQ*	*ΔrbsK*	*ΔsrlD*	*ΔxylA*	*ΔxylB*
Carbon Source	All sugars	All sugars	All sugars	All sugars	All sugars	All sugars	All sugars
	GG	MG	MG	RG	SG	XG	XG
	GM	MM	MM	RM	SM	XM	XM
	GR	MR	MR	RR	SR	XR	XR
	GS	MS	MS	RS	SS	XS	XS
	GX	MX	MX	RX	SX	XX	XX

In addition to the Keio collection strains, strains obtained via gene deletions following the improved pCAGO method outlined in Chapter one (bHL6, bHL21) were cultured in a 96-well plate provided with mixed carbon sources. The procedure was the same as described above.

Table 7 Categories of carbon sources provided for different strains using the improved pCAGO method. 0.2% w/v of each carbon source were provided in each well according to the table. (bHL6 = *ΔMalGFEK, lamB, malM*; bHL21 = *ΔgalTKM*)

Strain Types	bHL21(*ΔgalTKM)*	bHL6*(ΔmalGFEK-ΔlamB-ΔmalM*)
Carbon Source	All sugars	All sugars
	GG	MG
	GM	MM
	GR	MR
	GS	MS
	GX	MX

2.3 Representation of knockout strains in the Posfai et al. model

Based on the experiments described above, three strains were selected based on the selection criteria that strains for which the deletion of genes had the least impact on the growth rate across

different carbon source environments. Strains with genes *galK, malG*, and *xylB* deleted were chosen to investigate coexistence in mixed carbon sources. Correspondingly, galactose, maltose and xylose were chosen as the resources to represent the vertexes of the triangle (Chapter 2, section 1.2). At the time of writing, we did not have the equipment required to test the uptake rate of different sugars with precision, so we adopted an alternative way to map out the triangle, that is, to represent where each strain lies within this geometrical model. Several points were identified as shown in Fig. 24 (diamonds), and their locations within the triangle (tube number in Table 6) represent different proportions of the amount of each carbon source provided.

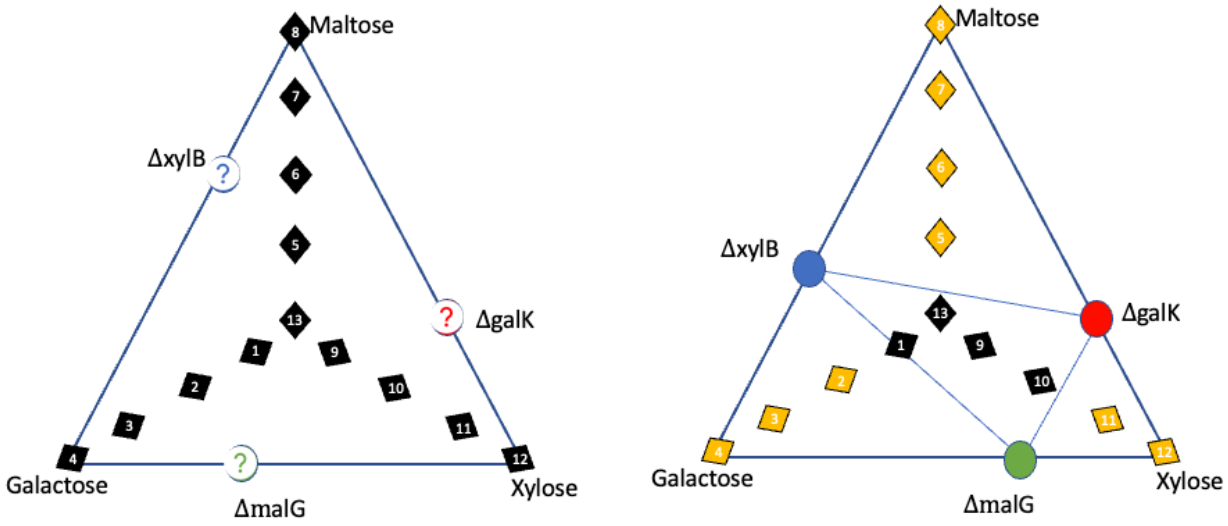


Figure 24 Triangle model for a three-species competition model in the presence of three resources. Each vertex denotes a specific carbon source. The three resources are maltose, galactose, and xylose. The numbers on the black diamonds correspond to the tube number in table 6, which represent the proportion of different resources that this system provides. Each colored disk denotes each mutant strain. The distance between each colored disk and each vertex represents the relative uptake rate of each resource by that particular strain. Question marks are used to indicate that the precise location is unknown. Yellow diamonds indicate the theoretical prediction that no coexistence should be observed with such proportion of carbon sources provided.

Table 8 Final concentrations of carbon sources added in M9 minimal medium.

	Concentration(%w/v)												
Tube Number	1	2	3	4	5	6	7	8	9	10	11	12	13
Galactose	0.15	0.2	0.25	0.3	0.075	0.05	0.025	0	0.075	0.05	0.025	0	0.1
Maltose	0.075	0.05	0.025	0	0.15	0.2	0.25	0.3	0.075	0.05	0.025	0	0.1
Xylose	0.075	0.05	0.025	0	0.075	0.05	0.025	0	0.15	0.2	0.25	0.3	0.1

Coexistence experiments were performed by co-inoculating species with varying carbon source fractions and propagating them through five growth-dilution cycles. After incubating the three chosen strains in LB medium for an overnight at 37^{o}C, 250uL of each overnight culture was mixed and centrifuged at 4.0g for 1 minute. The resulting pellet was resuspended twice in 750uL of PBS.

On day zero, 10 uL of PBS were inoculated into an Eppendorf tube containing 1 mL aliquots in corresponding falcon tubes, then 150 uL of the mixed solution was transferred into a single well in a 96-well plate. During each cycle, cells were cultured for 24 h and then diluted by a factor 31 into fresh M9 minimal medium with the correct mixture of sugars. Repeated dilutions were adopted to allow cells to grow for multiple generations. After 24 hours, 5 uL of the solution in wells were transferred into a new 96-well plate using a multichannel pipette. The transfer was made to the corresponding location, which contained the identical sugar type and amount. 75 uL of a 50% glycerol solution were added to all wells of the older plate, then the plates were stored at -80^{o}C. Every combination of a given strain and sugar mixture was replicated four times.

2.4 Coexistence criteria identified via fluorescence

To quantify the proportion of strains in a mixed carbon source environment after culturing, three plasmids containing three different fluorescence protein genes (mRFP1, Venus YFP, and CyOFP1) were transformed into the *ΔgalK, ΔmalG* and *ΔxylB* knockout strains, respectively. The presence and relative abundance of each strain was measured using a fluorescence plate reader, calibrated using monocultures, as well as in the flow cytometer. The excitation and emission were set according to FPbase (https://www.fpbase.org) (table 9) to achieve the least overlap among different wavelengths. To ensure distinct differentiation of each strain can be discerned via fluorescence plate reader, the fluorescence intensity was first measured independently for each strain. Then the colonies were mixed as Chapter 2, 2.3 and measured for fluorescence every 20 minutes during the five days dilution cycle. Every 24 hours, 2 uL of the cultures in 96-well plates were inoculated into 96-well U-bottom plate with 9mmol of Tris (tris(hydroxymethyl)aminomethane), and measured in different fluorescence channels via flow cytometer.

Table 9 Wavelength chosen for fluorescence protein gene to excite and emit.

Fluorescence Protein Gene	Excitation	Emission
Venus YFP	510 nm	536 nm
mRFP1	586 nm	612 nm
CyOFP1	472 nm	601 nm

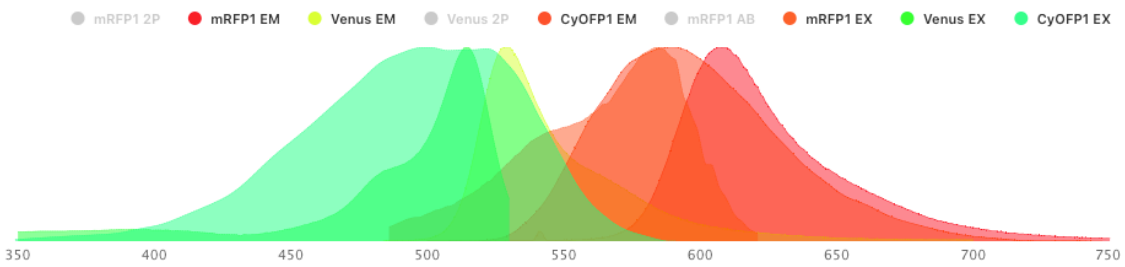


Figure 25 Wavelength comparing: mRFP1, Venus, CyOFP1. From FPbase

2.5 Materials and Protocols

2.5.1 M9 minimal medium for *E. coli*

To prepare 100 mL of M9 minimal medium, the following components were mixed: 20mL of M9 salts (5X), 2 mL of 20% carbon source solution, 200uL of 1M $MgSO_4$ solution, 10uL of 1M $CaCl_2$ solution, and 78 mL of MilliQ water. To prepare 200 mL of 5X M9 salts solutions, the following steps were followed: $Na_2HPO_4 \cdot 7H_2O$ (12.8 g), KH_2PO_4 (3 g), NaCl (0.5 g), NH_4Cl (1.0 g), MilliQ water (200 mL) were prepared and mixed in a specified order. Water first, followed by $CaCl_2$ solution, then the remaining components. The solution was mixed thoroughly and autoclaved to ensure complete dissolution of the components.

2.5.2 Plate reader inoculation

To ensure comparable starting concentrations, the strains were cultured in LB medium for an overnight under 37°C. Subsequently, they were centrifuged at 4.0g for 1 minute and resuspended with 1ml of PBS twice. Then the resuspended solution was inoculated into the 96-well plate as needed. For growth rate measurements, the 96-well plates were set in a plate reader at 37°C for desired length of time. Growth rate measurements were taken every 20 minutes. Before each measurement, the 96-well plate was shaken for 10s to ensure the samples were homogenous.

2.5.3 Strains transformation with plasmid expressing fluorescent proteins

To ensure that the fluorescence signal produced by different strains could be detected independently, we selected the fluorescent proteins Venus YFP (excitation at 510 nm, emission at 536 nm), mRFP1(excitation at 586 nm, emission at 612 nm), and cyOFP1(excitation at 472 nm, emission at 601 nm). The genes encoding for mRFP1 and CyOFP1 were introduced in pbAG1 via FastCloning[52], pbAG1 contained the fluorescent protein CFP expressed via the trc promoter (inducible by addition of IPTG). Then these created plasmids were transformed into *ΔgalK* and *ΔxylB*, respectively. Gene encoding for Venus YFP is expressed via the promoter, and the plasmid pbAG2 was transformed into *ΔmalG*.

3. Results and Discussion

3.1 Growth on single carbon sources of the majority of the *E. coli* Keio collection strains correspond with the CarPE results

As anticipated, the majority of the strains did not display growth signals after 48 hours of culturing when provided with carbon sources that strains were incapable to metabolize according to the CarPE dataset. In contrast, the wildtype strains exhibited weak growth after approximately 20 hours of culture in M9 minimal medium.

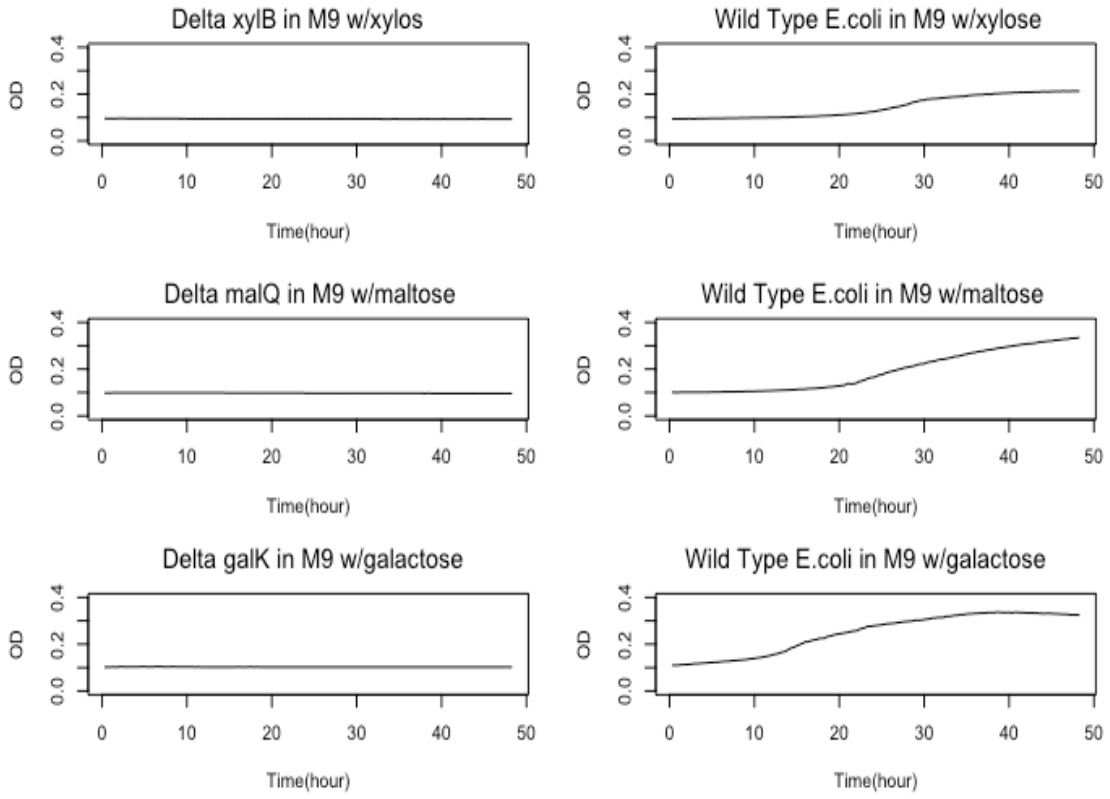


Figure 26 Growth curves of different types of engineered *E. coli* strains under the single carbon sources environment. On the left are *ΔxylB, ΔmalQ* and *ΔgalK* respectively. Figures on the right are wildtype DH5a in the same carbon source environment.

However, exceptions were observed for *ΔmanY, ΔmanZ and Δ malE*. Strains *ΔmanY* and *ΔmanZ* did not display apparent differences in their growth when compared with the wildtype. For *ΔmalE*, a moderate amount of growth was observed after 24 hours.

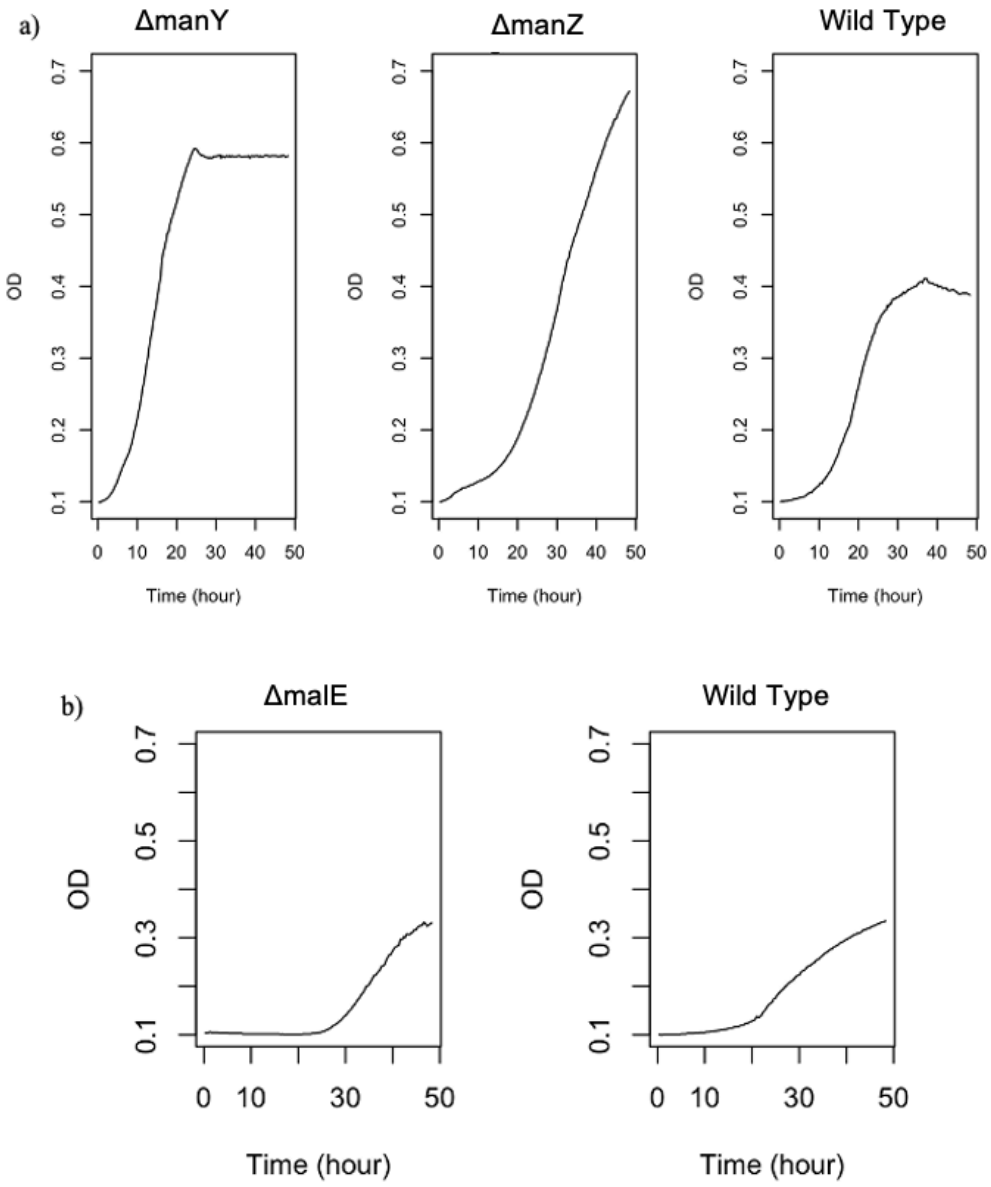


Figure 27 48 hours of growth curves of *E. coli* Keio collection strains a) Δ *manY* (a, left), *ΔmanZ* (a, middle) and Wildtype (a, right) in M9 minimal medium with Mannose. b) *ΔmalE* (b, left), and Wildtype (a, right) in M9 minimal medium with Maltose

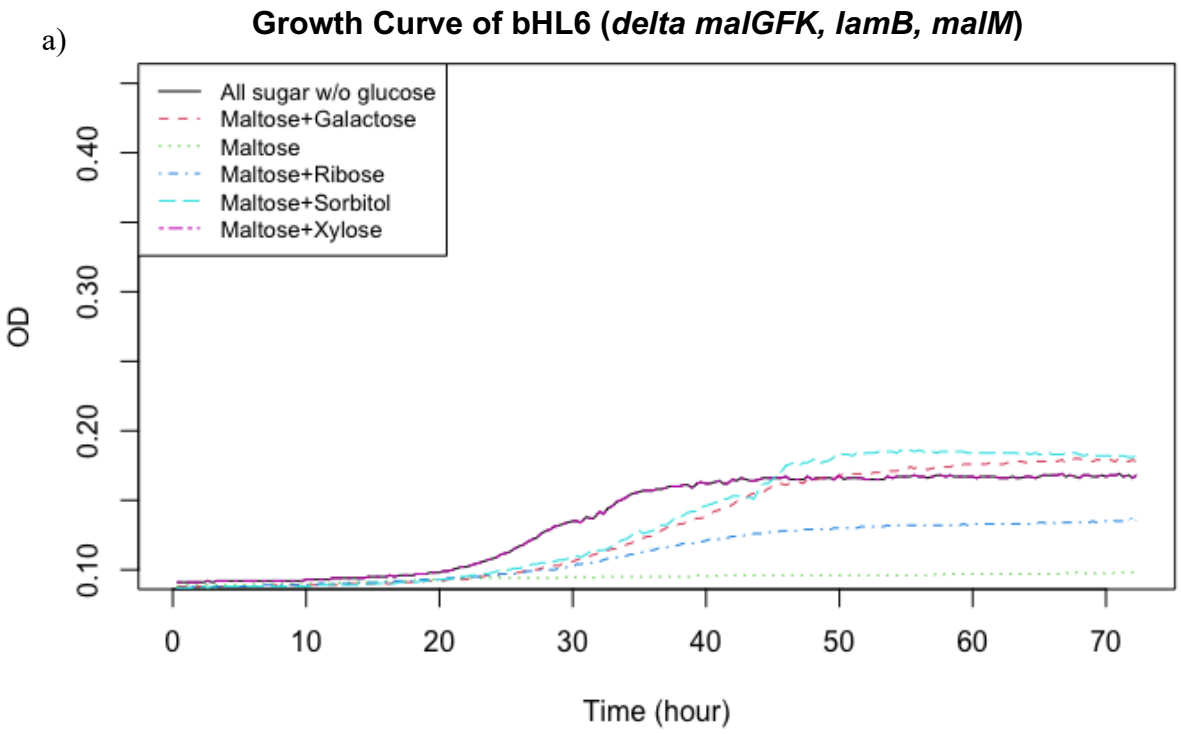
3.2 Catabolite repression in the Keio collection metabolic knockouts

3.2.1 Strains with metabolic genes deleted using the improved pCAGO method

In Chapter one, we successfully created scarless knockouts of genes involved in the metabolism and uptake of maltose (*malG, malF, malE, malK, lamB, malM, strain name* bHL6),

galactose(*galT, galK, galM, strain name bHL21*), and non-scarless knockouts of genes involved in the metabolism and uptake of fructose(*fruB, fruK,fruA, strain name bHL22*), mannose(*manX, manY, manZ, strain name bHL20*) and ribose(*rbsD, rbsA, rbsC, rbsB, rbsK, strain name bHL18)*. For these latter genes, the recombination process to remove the selection marker has not yet been completed. To investigate their catabolite repression, they were cultured in a 96-well plate with various carbon sources for catabolite repression analysis. As expected, bHL6 (*ΔmalG, ΔmalF, ΔmalE, ΔmalK, ΔlamB, ΔmalM*), and bHL21a (*ΔgalT, ΔgalK, ΔgalM*) did not grow when maltose/galactose was provided as the single carbon source. When they were provided with different carbon sources, they uptake the other sugars and showed abilities to grow. However, the impeded growths were noticed in these strains (Fig. 28). In the Keio collection *E. coli* strains, a typical 10 hours of lag phase was observed, and exponential phase was reached between approximately 10 to 30 hours after initiating culture. This could be because they were adapting to the new environment, as the wildtype *E. coli* also exhibited 10 hours of lag phase. However, the strains that are deleted via pCAGO method displayed an approximately 20 hours of lag phase. Additionally, after 60 hours of culture, the highest OD reached by bHL6, bHL21a, among all carbon sources compositions, were 0.15 and 0.2. These values were significantly lower than that of wildtype *E. coli* and strains with single gene deletions (0.5).

a)



b)

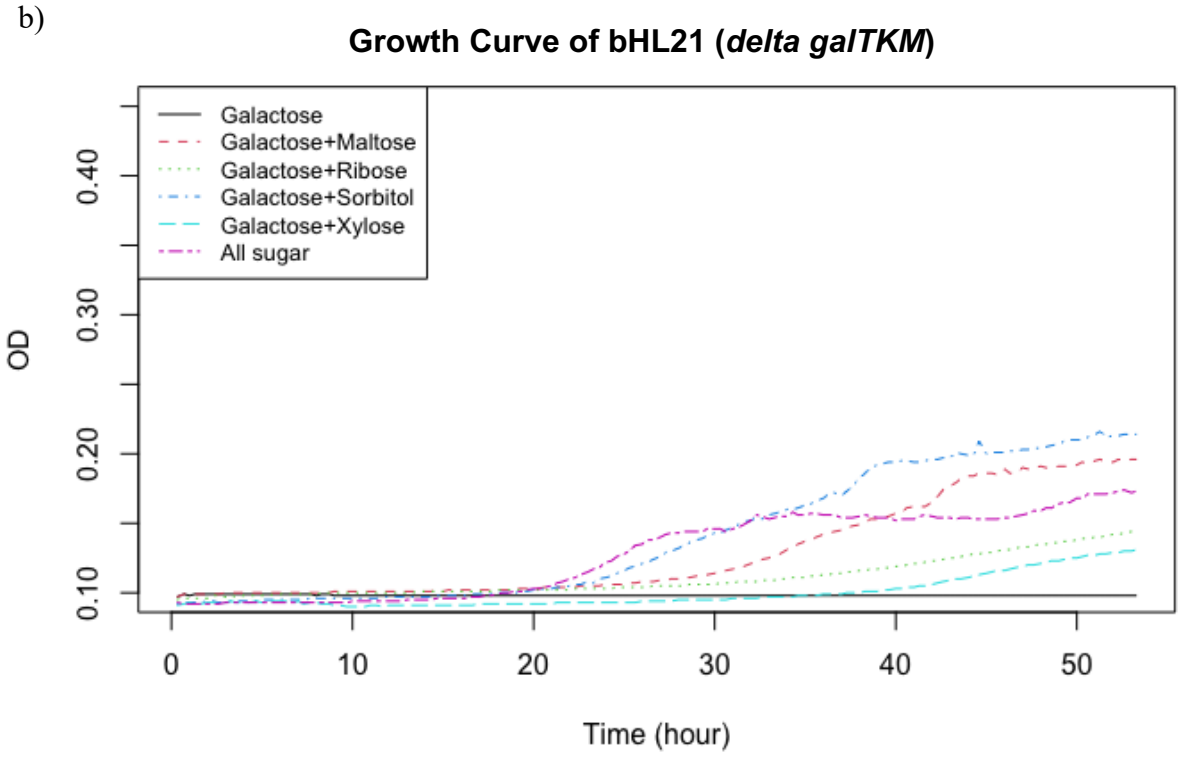


Figure 28 60 hours of growth curves of strains with knockouts (measured in OD) using pCAGO method. a) bHL6(*ΔmalG, ΔmalF, ΔmalE, ΔmalK, ΔlamB, ΔmalM*) b) bHL21 (*ΔgalT, ΔgalK, ΔgalM*)

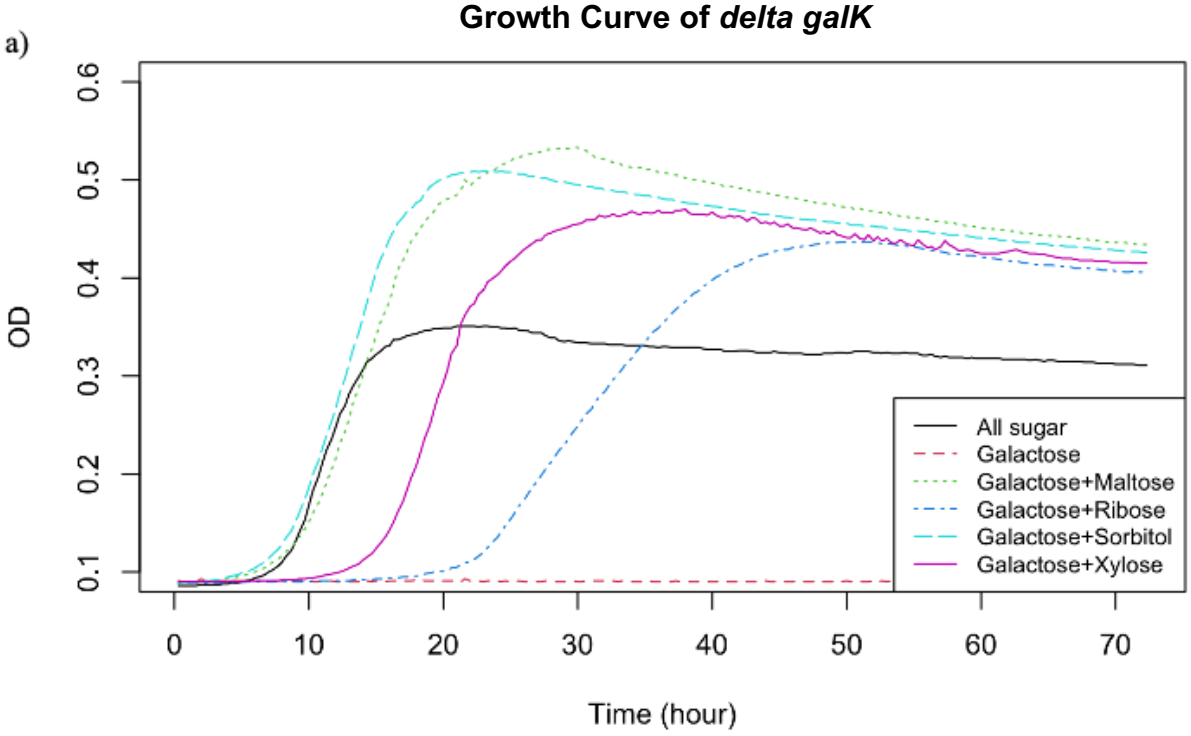
3.2.2 *E. coli* Keio collection growth rates provided with multiple combination of Carbon Sources

From results in section 3.1, we excluded the strains that showed capability to grow in carbon source that they cannot metabolize. Therefore, strains deleted with *galK, malG, malQ, rbsK, srlD, xylA, xylB* were chosen to investigate the carbon catabolic repression with paired genes provided.

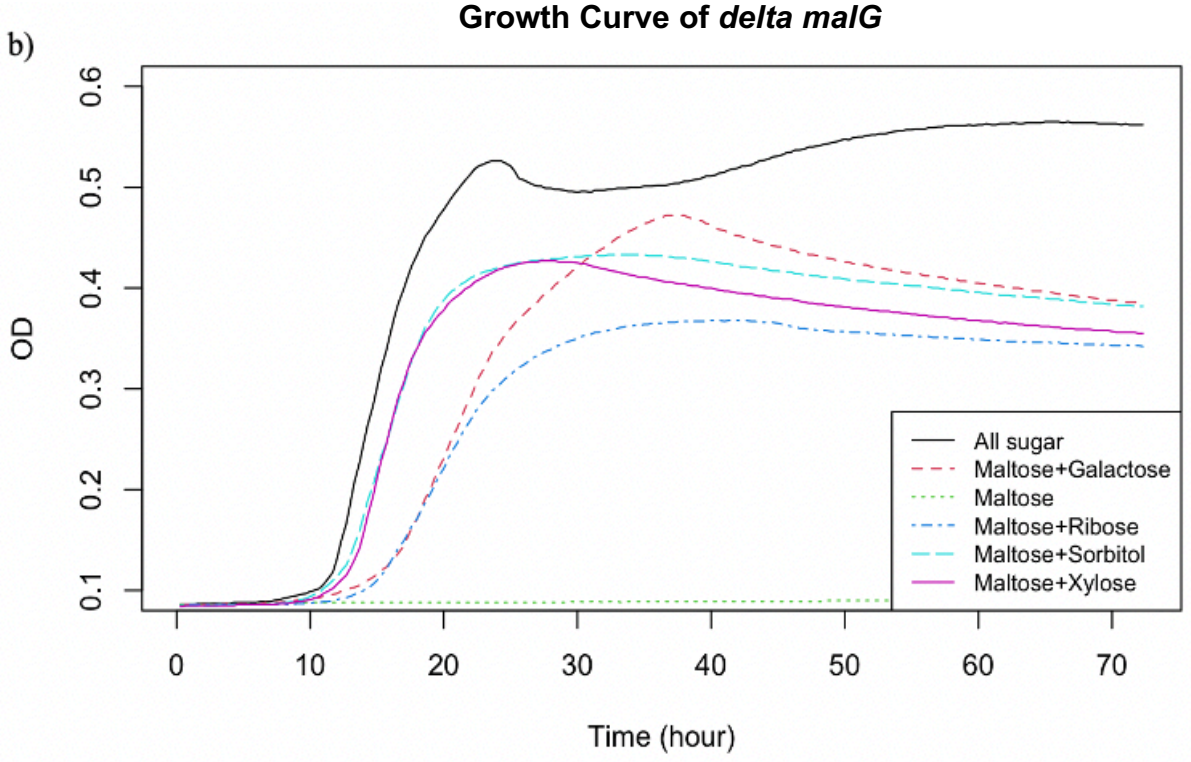
Among the tested engineered strains, catabolic repression did not occur except for the *ΔsrlD* mutant, which exhibited this phenomenon in the presence of maltose. The other engineered strains were able to grow when provided with other carbon sources, albeit with varying degrees of lag time. When five carbon sources (galactose, maltose, ribose, sorbitol and xylose) were present, the growth rate was the highest compared to paired carbon source provided, except for *ΔgalK*. Sorbitol and xylose were found to be good resources when provided as the second carbon source, as all tested strains showed a relatively high growth rate when these sugars were present. In contrast, when ribose was present as the additional carbon source, strains took a relatively long time to grow, resulting in significantly lower OD compared to the others.

Due to the weak performance of pCAGO deleted engineered *E. coli* strains on growth in M9 minimal medium, our study pivoted to testing the predictions of Posfai's model using the Keio collection strains. The selection criteria for the strains were based on identifying strains for which the deletion of genes had the least impact on the growth rate across different carbon source environments. Specifically, we targeted strains that exhibited a carrying capacity of 0.5 OD and no obvious sign of lag while they grew. Based on the results shown in figure 28, we chose *ΔgalK, ΔmalG, ΔxylB*. They were verified uncapable of metabolizing galactose, maltose and xylose, respectively.

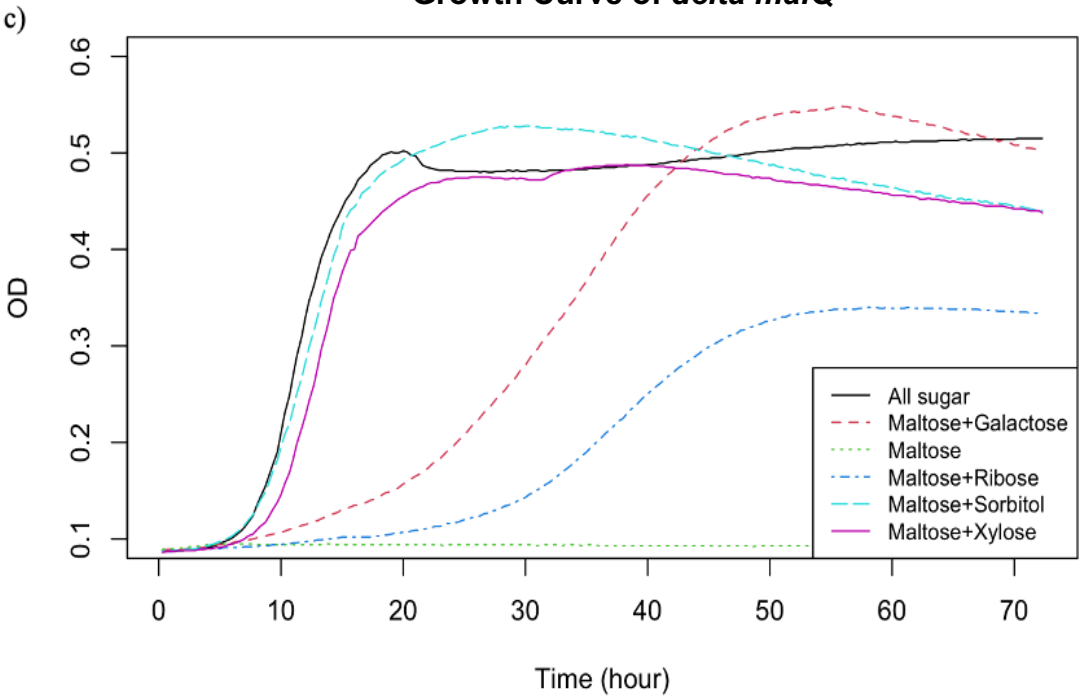
a)


Growth Curve of *delta galK*
OD
0.1
0.2
0.3
0.4
0.5
0.6
0
10
20
30
40
50
60
70
Time (hour)
All sugar
Galactose
Galactose+Maltose
Galactose+Ribose
Galactose+Sorbitol
Galactose+Xylose

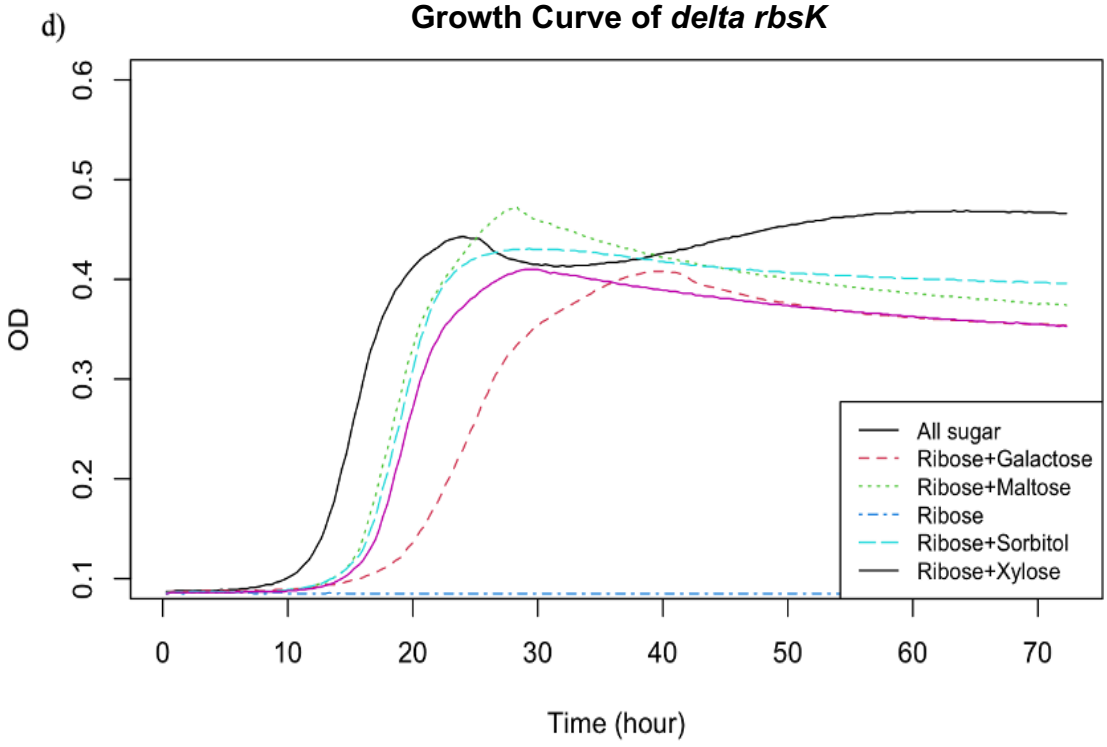
b)


Growth Curve of *delta malG*
OD
0.1
0.2
0.3
0.4
0.5
0.6
0
10
20
30
40
50
60
70
Time (hour)
All sugar
Maltose+Galactose
Maltose
Maltose+Ribose
Maltose+Sorbitol
Maltose+Xylose

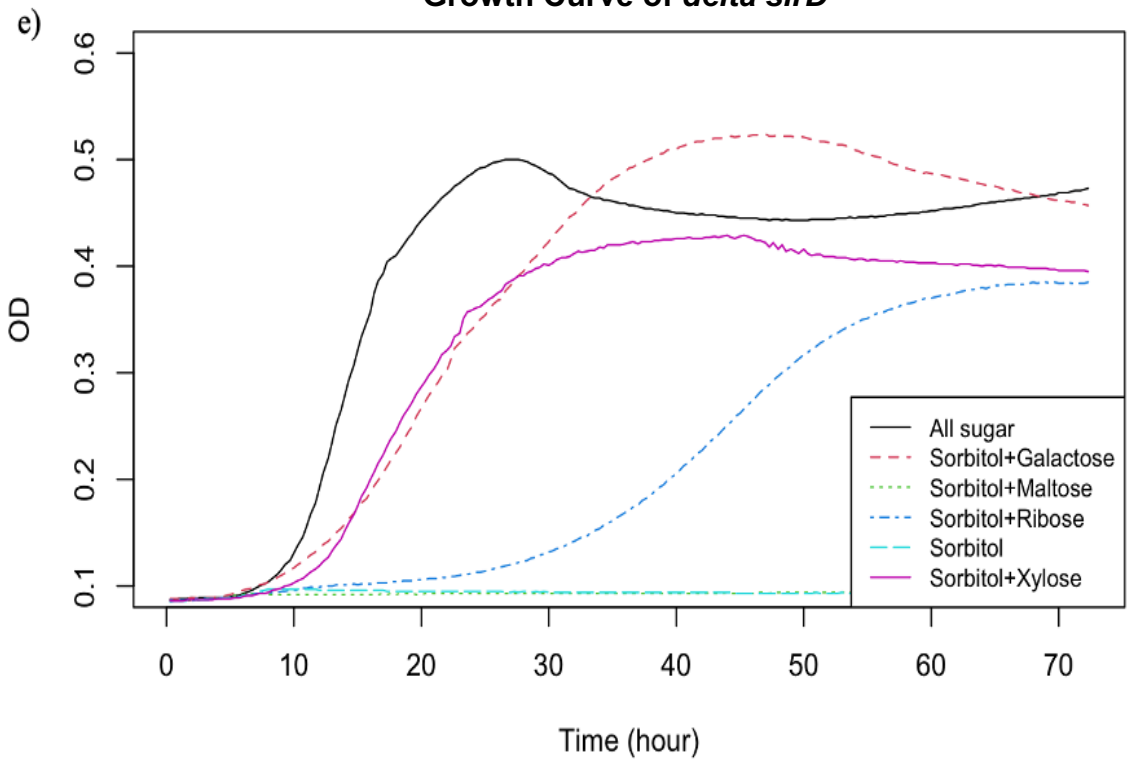
c)


Growth Curve of *delta malQ*
OD
0.1
0.2
0.3
0.4
0.5
0.6
0
10
20
30
40
50
60
70
Time (hour)
All sugar
Maltose+Galactose
Maltose
Maltose+Ribose
Maltose+Sorbitol
Maltose+Xylose

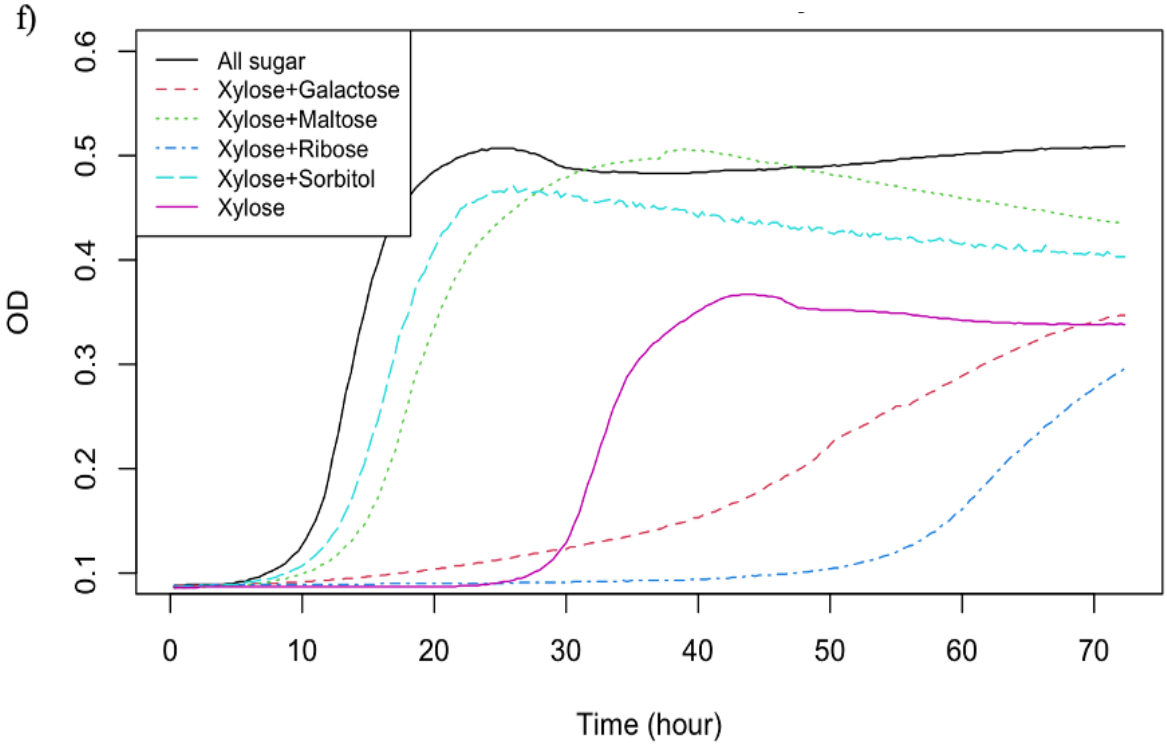
d)


Growth Curve of *delta rbsK*
OD
0.1
0.2
0.3
0.4
0.5
0.6
0
10
20
30
40
50
60
70
Time (hour)
All sugar
Ribose+Galactose
Ribose+Maltose
Ribose
Ribose+Sorbitol
Ribose+Xylose

e)


Growth Curve of delta slrD
OD
0.1
0.2
0.3
0.4
0.5
0.6
0
10
20
30
40
50
60
70
Time (hour)
All sugar
Sorbitol+Galactose
Sorbitol+Maltose
Sorbitol+Ribose
Sorbitol
Sorbitol+Xylose

f)


Growth Curve of delta xylA
OD
0.1
0.2
0.3
0.4
0.5
0.6
0
10
20
30
40
50
60
70
Time (hour)
All sugar
Xylose+Galactose
Xylose+Maltose
Xylose+Ribose
Xylose+Sorbitol
Xylose

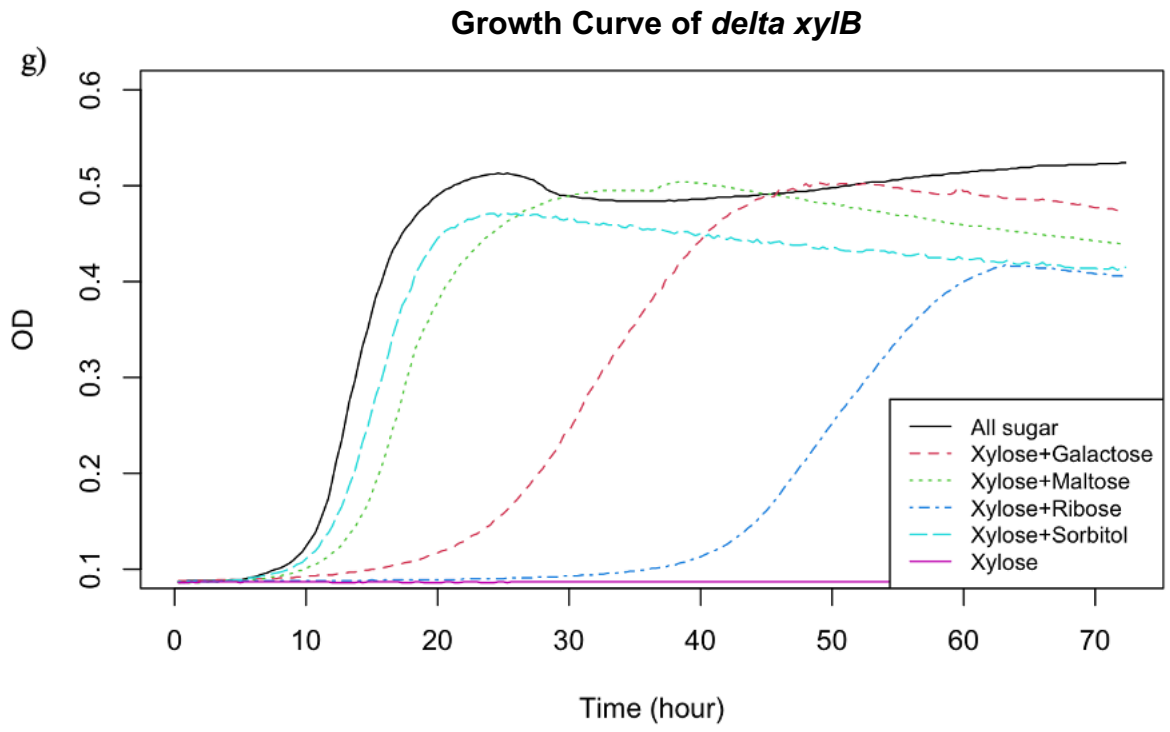


Figure 29 Growth curves of *E. coli* Keio collection strains (measured in OD) under different carbon sources combinations in 72 hours. a) *ΔgalK*, b) *ΔmalG*, c) *ΔmalQ*, d)ΔrbsK e)*ΔsrlD*, f)*ΔxylA* g)*ΔxylB*

3.3 Either D-sorbitol or D-sorbitol 6-phosphate repress the maltose system

We found that the strain from the Keio knockout collection in which gene *srlD* is deleted is unable to grow in the presence of both maltose and sorbitol, despite this strain is capable of growing in the presence of maltose alone.

To the best of our knowledge, there are no previous reports of this growth inhibition of *ΔsrlD* in the presence of both maltose and sorbitol. Hints that the presence of sorbitol may inhibit maltose catabolism are found in Aidelberg et al.[48], they found that expression of the maltose catabolism system is almost fully repressed by the presence of sorbitol. However, no further exploration of the mechanism of such regulation was mentioned in that paper, and the regulatory relationship between sorbitol and maltose is currently unknown.

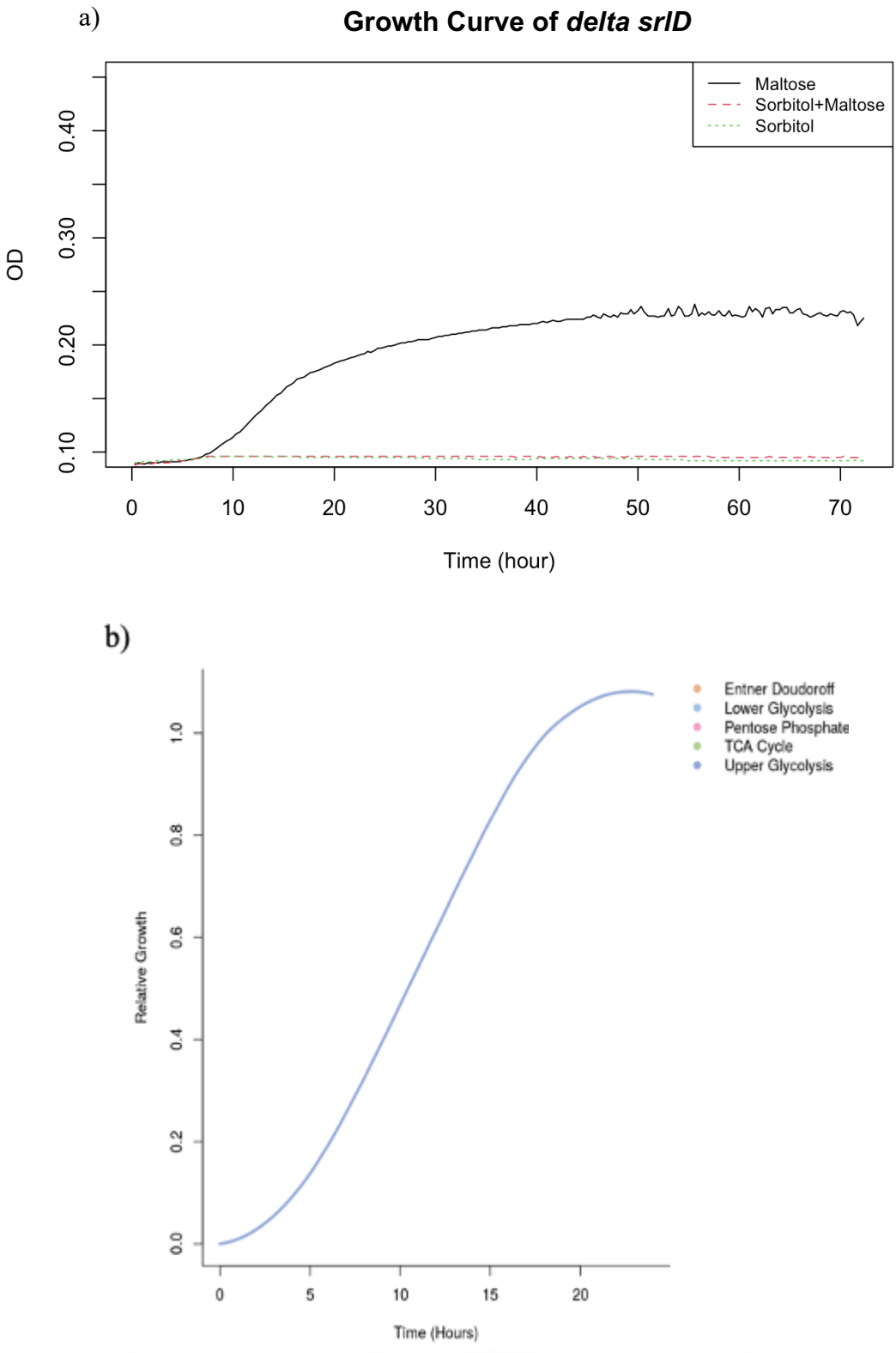


Figure 30 Growth curve of *ΔsrlD* a) in the presence of sorbitol (red dashed line), sorbitol and maltose (black line). B) in the presence of maltose, provided by CarPE

Here, we proposed that it is either the extracellular D-sorbitol or the intracellular D-sorbitol-6-phosphate that repress the maltose system. Genes *srlA*, *srlB*, and *srlE* encode for the sorbitol PTS permease which transports and simultaneously phosphorylates sorbitol. The next step is D-sorbitol degradation II (D-sorbitol 6-phosphate to keto-D-fructose 6-phosphate), which is encoded by *srlD* (Fig. 30). In the *ΔsrlD* strain, D-sorbitol degradation cannot occur. Therefore, our data suggests that the reactant that represses the maltose system expression would either be extracellular D-sorbitol or intracellular D-sorbitol-6-phosphate. To verify our hypothesis, one can check if cells grow on sorbitol-6-phosphate and maltose, but this is deemed outside of the scope of this thesis.

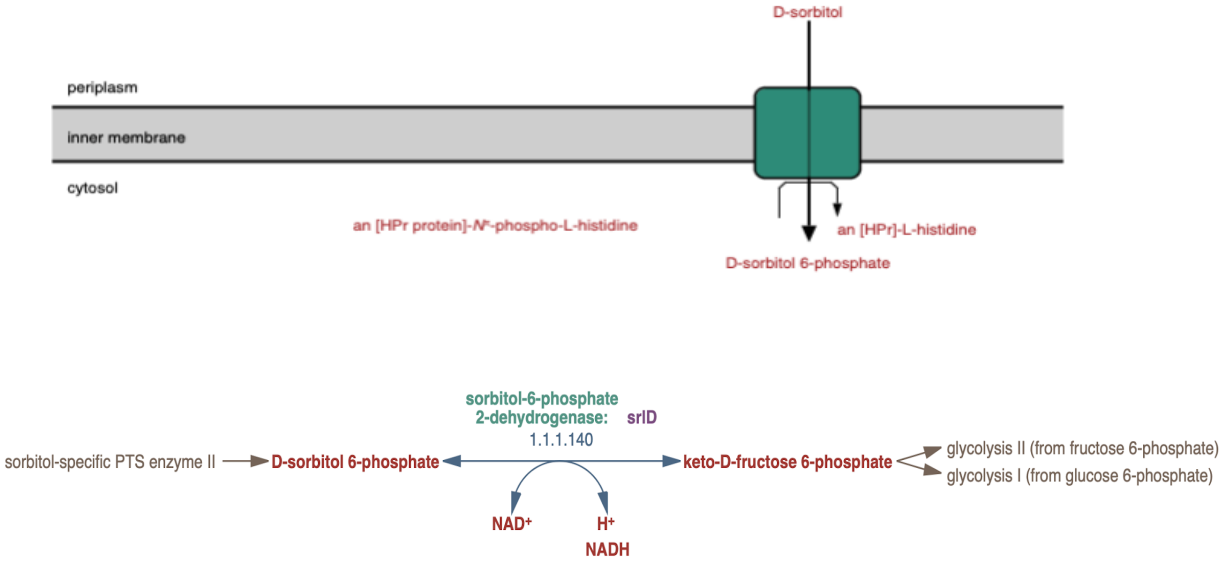


Figure 31 D-Sorbitol transport process, provided by Ecocyc.

3.4 Coexistence

In this system, three carbon sources (maltose, galactose and xylose) and the three corresponding knockout strains *(ΔmalG, ΔgalK, ΔxylB)* are included in the triangle model.

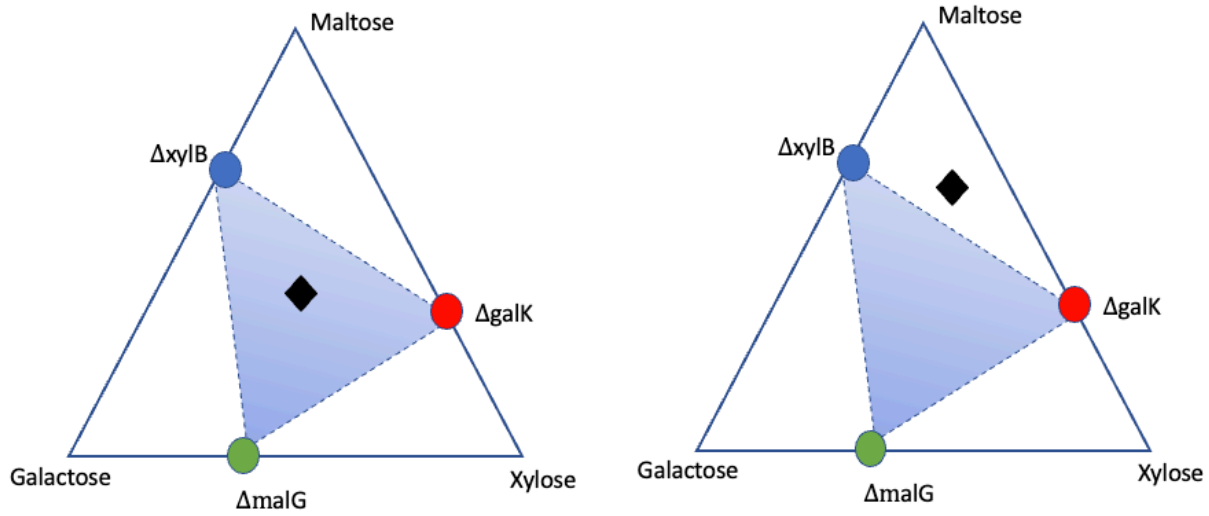


Figure 32 Three-species competition model. Each vertex denotes a specific carbon source. Three resources are maltose, galactose, xylose. Black diamond represents the proportion of resources that this system provides. Each colored circle denotes each mutant strain, the distance between the colored point and each vertex represented their ability to utilize different resources.

In order to empirically tested Posfai's model[12], the carbon sources uptake rate of these engineered strains needs to be determined in order to pinpoint the location of the colored points on the triangle. However, at the time of writing, our laboratory lacked the necessary equipment for measuring these uptake rates. An alternative method was thus devises. We established a sugar supplement matrix within the triangle where several points were identified as shown in figure 24, and their locations within the triangle represent different proportions of the amount of carbon sources provided. In doing so, we could identify the coexistence for each black diamond, therefore an approximate range of the shaded triangle broader could be drawn.

3.4.1 Strains detection

To quantify the proportion of species present, plasmids containing different fluorescence protein genes were individually transformed into the chosen strains, bHL27(*ΔgalK*, mRFP1), bHL29(*ΔmalG*, Venus YFP), bHL30(*ΔxylB*, Cyofp1). To establish a correlation between the data obtained from the plate reader and the true co-culture fluorescence data, the strains were cultured separately in M9 minimal media with glucose for changes in fluorescence over time. Along with the culture growth, the increase of each strain fluorescence can be measured and easily distinguished in different fluorescent channel. Such single-strain growth curves allowed us to identify that Venus YFP and mRFP1 both had bleed-through in the CyOFP1 channel, as shown in Figure 33. These bleed-through can be compensated by a linear transformation (Figure 34), which allows measuring the fluorescence intensity of each individual strain at the plate reader, despite the bleed-through. Specifically, compensated fluorescence values in the three channels would read:

$$\boldsymbol{F} = \begin{pmatrix} F_Y \\ F_C \\ F_R \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ S_Y^c & 1 & S_R^c \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} D_Y \\ D_C \\ D_R \end{pmatrix} = \boldsymbol{S^{-1}D},$$

where S_Y^c is the ratio between fluorescence intensity values in the CyOFP1 channel and those in the Venus YFP channel for the strain expressing Venus YFP, and similarly S_R^c is the ratio between fluorescence intensity values in the CyOFP1 channel and those in the mRFP1 channel for the strain expressing mRFP1. In the equation above, $\boldsymbol{D} = (D_Y, D_C, D_R)$ are fluorescence intensity values measured at the plate reader. Figure 34 shows the rescaled fluorescence intensity values $\boldsymbol{F} = (F_Y, F_C, F_R)$.

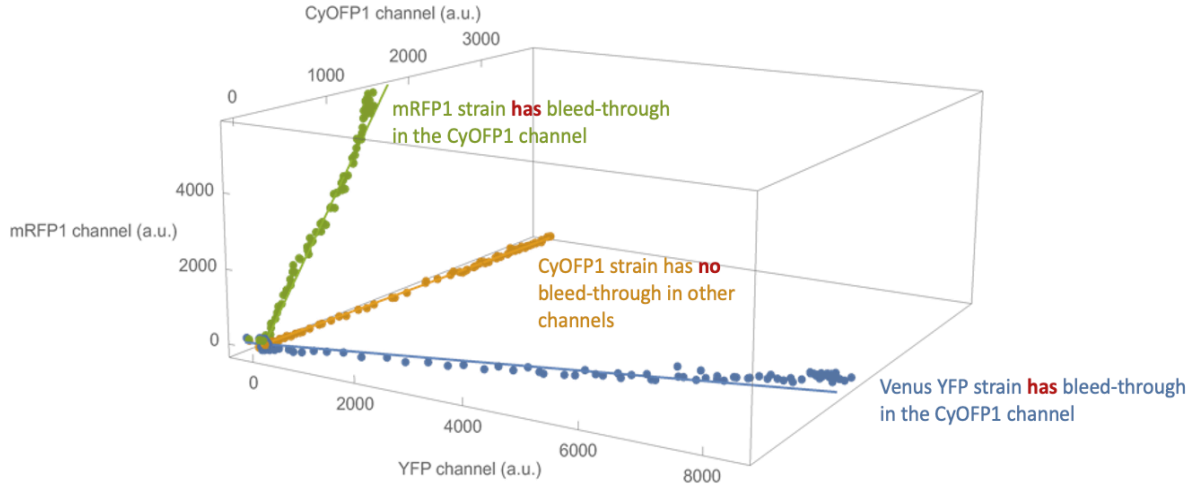


Figure 33 Single-strain growth curves measured in the three fluorescent channels at the plate reader.

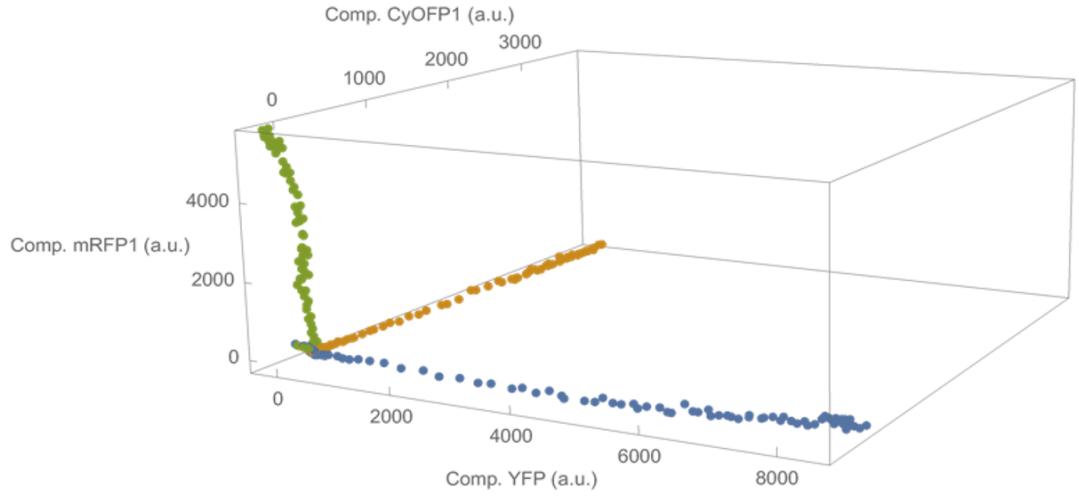


Figure 34 No bleed-through in compensated fluorescence values

3.4.2 Growth rate and acetate overflow metabolism can affect the coexistence model

According to the results obtained from flow cytometer analysis, we found that bHL30 (*ΔxylB*, with CyOFP1) exhibited a decrease in concentration towards the end of the first day (Table 12, Figure. 35), all channels were primarily dominated by bHL27 (*Δgalk* with mRFP1 fluorescence) except for column C1, which was only provided with galactose (Fig. 35). This observation might be attributed to different growth rates, which despite having similar initial OD, resulted in the

loss of competitive ability or possessed a competitive advantage of one strain in comparison to the other two. To address this issue, future work will focus on acclimating the strains to achieve a similar growth rate prior to inoculation, in order to obtain more accurate and reliable results.

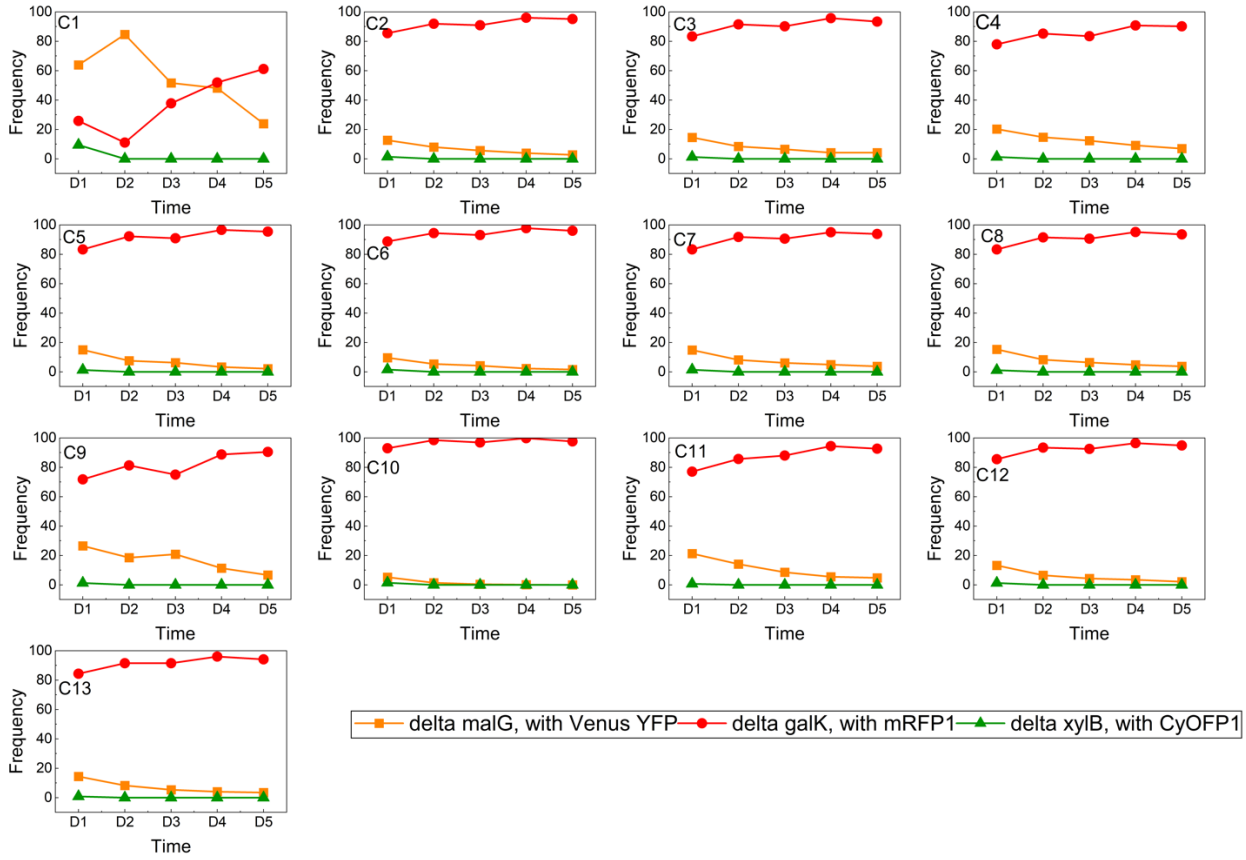


Figure 35 Daily frequencies of fluorescence protein shown in flow cytometer. C1-C13 represents the location in 9 well plate, the number correspond to the number in table 8.

In addition, Even in C1, where *ΔgalK* was not expected to survive, the frequency of mRFP1 gradually increased from day one to day five (Fig. 36). To investigate the reason for its growth capabilities, we restreaked cells from the last time point from the -80°C glycerol stock onto an LB plate with ampicillin and IPTG, and picked three single colonies that showed red fluorescence. Then, cells from these colonies were cultured in M9 medium with galactose, ampicillin and IPTG to test whether they could grow in the presence of galactose as the only carbon source. The inability of these strains to grow in the presence of galactose as the only

carbon source suggested that strain bHL27 may be metabolizing acetate. This is supported by the fact that *E. coli* excretes acetate upon growth on fermentable sugars[53].

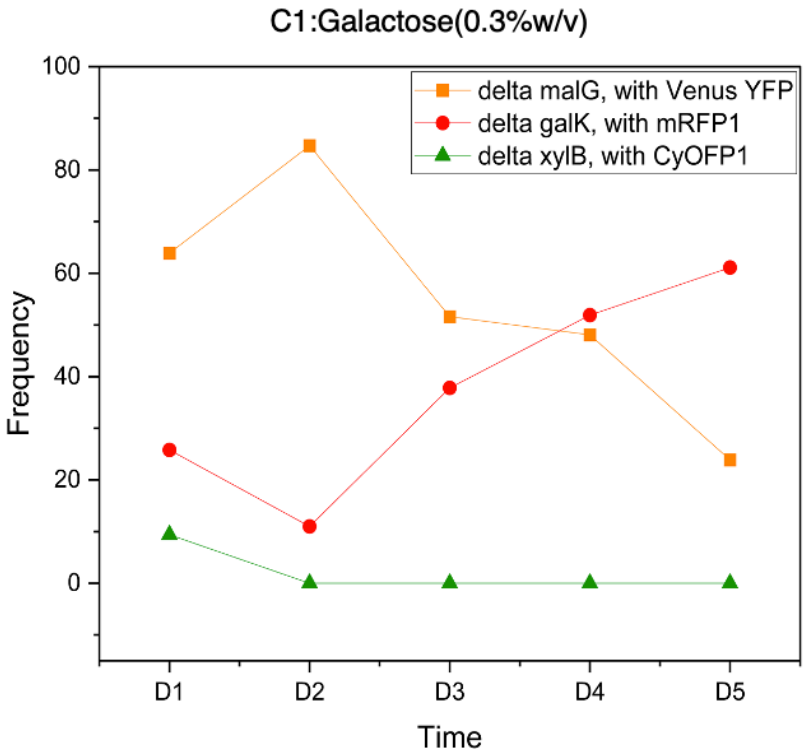


Figure 36 Channel C1, 0.3% (w/v) of galactose was provided.

The above-mentioned finding indicated that the model proposed by Posfai et al.[12] may require more comprehensive consideration, such as the possibility of having acetate overflow.

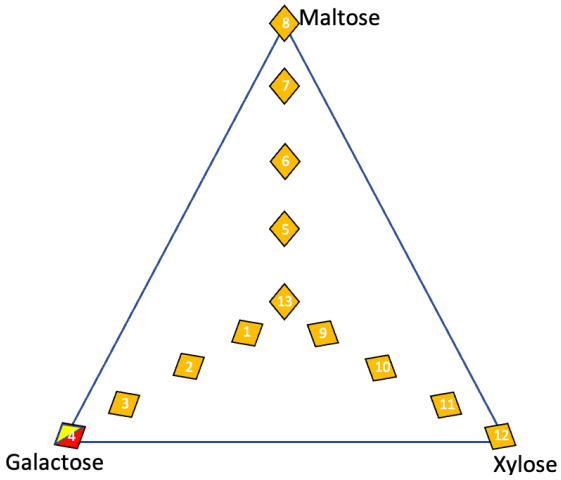


Figure 37 Triangle model in preliminary experiment. No coexistence existence at any location.

4. Conclusions and future plans

In chapter one, we successfully created the knockout of genes involved in the metabolism of maltose (*malG, malF, malE, malK, lamB, malM*) and galactose (*galT, galK, galM)*. Genes involved in the metabolism of fructose (*fruB, fruK, fruA*), mannose (*manX, manY, manZ)* and ribose (*rbsD, rbsA, rbsC, rbsB, rbsK)* were also deleted, but the selection marker is not yet removed. The reason for the difficulty to achieve the excision of the resistance cassette requires further investigation. Some of the engineered strains require other genes to be deleted to achieve a complete inability to metabolize certain carbon sources. Future work in this project will be dedicated towards performing these knockouts. We have improved the scarless genetic knockout system of Zhao et al.[16] Specifically, we incorporated fluorescence protein genes into different plasmids used in the scarless genetic knockout system, which allows for the detection of false positive transformants via a fluorescence microscope. Future knockouts will benefit from this improved methodology.

In chapter two, we established that catabolite repression should not be of concern for the engineered microbial consortia, given that the majority of Keio knockout strains investigated could grow in the presence of any other sugar. This is of interest even beyond our project, for example in applications in which researchers plan to adopt similar knockout strains to metabolize complex carbon sources[14]. One exception to this rule is represented by the growth of *ΔsrlD* mutants in the presence of both sorbitol and maltose, for which we found that growth did not occur in the presence of these carbon sources. Our results suggest that either D-sorbitol or D-sorbitol-6-phosphate that represses the maltose system. In this context, future work may be dedicated to discriminating which substance that represses the maltose system in *E. coli*. To perform a preliminary test of consumer-resource models of microbial coexistence, we designed

an experiment in which we would grow different Keio knockout strains in the presence of multiple resources, some of which they cannot metabolize. We developed a three-plasmid system to measure the relative frequency of different strains at either the fluorescence plate reader or flow cytometer. In addition, we found that the growth rate and acetate overflow metabolism can significantly impact the coexistence model establishment.

In the future, this project will adopt High-Performance Liquid Chromatography (HPLC) to determine the uptake rates of different carbon sources by different strains, which will allow the quantitative test of the prediction of consumer-resource models of microbial coexistence. This project will also need to acclimate strains to attain a similar growth rate prior to inoculation, and consider the fact of acetate overflow metabolism in this model.

Reference

1. Monod, J. & Jacob, F. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 389–401 (1961).
2. McAdams, H. H. & Arkin, A. Towards a circuit engineering discipline. *Curr. Biol.* **10**, R318-20 (2000).
3. McAdams, H. H. & Shapiro, L. Circuit simulation of genetic networks. *Science* **269**, 650–656 (1995).
4. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
5. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339–342 (2000).
6. Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. Combinatorial synthesis of genetic networks. *Science* **296**, 1466–1470 (2002).
7. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat. Rev. Microbiol.* **12**, 381–390 (2014).
8. Brenner, K., You, L. & Arnold, F. H. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol.* **26**, 483–489 (2008).
9. Sabra, W., Dietz, D., Tjahjasari, D. & Zeng, A.-P. Biosystems analysis and engineering of microbial consortia for industrial biotechnology. *Eng. Life Sci.* **10**, 407–421 (2010).
10. Zhang, Y., Hsu, H.-H., Wheeler, J. J., Tang, S. & Jiang, X. Emerging investigator series: emerging biotechnologies in wastewater treatment: from biomolecular engineering to multiscale integration. *Environ. Sci.: Water Res. Technol.* **6**, 1967–1985 (2020).
11. Duncker, K. E., Holmes, Z. A. & You, L. Engineered microbial consortia: strategies and applications. *Microb. Cell Fact.* **20**, 211 (2021).
12. Posfai, A., Taillefumier, T. & Wingreen, N. S. Metabolic Trade-Offs Promote Diversity in a Model Ecosystem. *Phys. Rev. Lett.* **118**, 028103 (2017).
13. Friedman, J., Higgins, L. M. & Gore, J. Community structure follows simple assembly rules in microbial microcosms. *Nat. Ecol. Evol.* **1**, 109 (2017).
14. Xia, T., Eiteman, M. A. & Altman, E. Simultaneous utilization of glucose, xylose and arabinose in the presence of acetate by a consortium of Escherichia coli strains. *Microb. Cell Fact.* **11**, 77 (2012).

15. Manhart, M. & Shakhnovich, E. I. Growth tradeoffs produce complex microbial communities on a single limiting resource. *Nat. Commun.* **9**, 3214 (2018).
16. Zhao, D. *et al.* CRISPR/Cas9-assisted gRNA-free one-step genome editing with no sequence limitations and improved targeting efficiency. *Sci. Rep.* **7**, 16624 (2017).
17. Uddin, F., Rudin, C. M. & Sen, T. CRISPR gene therapy: applications, limitations, and implications for the future. *Front. Oncol.* **10**, 1387 (2020).
18. Li, H. *et al.* Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduct. Target. Ther.* **5**, 1 (2020).
19. Andolfo, G., Iovieno, P., Frusciante, L. & Ercolano, M. R. Genome-Editing Technologies for Enhancing Plant Disease Resistance. *Front. Plant Sci.* **7**, 1813 (2016).
20. Gaj, T., Sirk, S. J., Shui, S.-L. & Liu, J. Genome-Editing Technologies: Principles and Applications. *Cold Spring Harb. Perspect. Biol.* **8**, (2016).
21. Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14**, 49–55 (2013).
22. 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).
23. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, Engl)* **151**, 2551–2561 (2005).
24. Westra, E. R. *et al.* The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu. Rev. Genet.* **46**, 311–339 (2012).
25. 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).
26. Asmamaw, M. & Zawdie, B. Mechanism and Applications of CRISPR/Cas-9-Mediated Genome Editing. *Biologics* **15**, 353–361 (2021).
27. Ryan, D. E. *et al.* Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs. *Nucleic Acids Res.* **46**, 792–803 (2018).
28. Aquino-Jarquin, G. Current advances in overcoming obstacles of CRISPR/Cas9 off-target genome editing. *Mol. Genet. Metab.* **134**, 77–86 (2021).

29. Arroyo-Olarte, R. D., Bravo Rodríguez, R. & Morales-Ríos, E. Genome Editing in Bacteria: CRISPR-Cas and Beyond. *Microorganisms* **9**, (2021).
30. Chen, L.-Q., Cheung, L. S., Feng, L., Tanner, W. & Frommer, W. B. Transport of sugars. *Annu. Rev. Biochem.* **84**, 865–894 (2015).
31. Shimizu, K. Metabolic Regulation of a Bacterial Cell System with Emphasis on Escherichia coli Metabolism. *ISRN Biochem.* **2013**, 645983 (2013).
32. Nakrani, M. N., Wineland, R. H. & Anjum, F. Physiology, Glucose Metabolism. in *StatPearls* (StatPearls Publishing, 2022).
33. Tong, M. *et al.* Gene Dispensability in Escherichia coli Grown in Thirty Different Carbon Environments. *MBio* **11**, (2020).
34. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
35. Keseler, I. M. *et al.* EcoCyc: a comprehensive database of Escherichia coli biology. *Nucleic Acids Res.* **39**, D583-90 (2011).
36. Sheets, M. B., Wong, W. W. & Dunlop, M. J. Light-Inducible Recombinases for Bacterial Optogenetics. *ACS Synth. Biol.* **9**, 227–235 (2020).
37. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3**, e3647 (2008).
38. New England Biolabs. Golden Gate Assembly. https://www.neb.com/applications/cloning-and-synthetic-biology/dna-assembly-and-cloning/golden-gate-assembly.
39. barrick lab gg. https://barricklab.org/twiki/bin/view/Lab/ProtocolsBTKDesignANewPart.
40. El Karoui, M., Amundsen, S. K., Dabert, P. & Gruss, A. Gene replacement with linear DNA in electroporated wild-type Escherichia coli. *Nucleic Acids Res.* **27**, 1296–1299 (1999).
41. Kornberg, H. L. Routes for fructose utilization by Escherichia coli. *J. Mol. Microbiol. Biotechnol.* **3**, 355–359 (2001).
42. Kornberg, H. & Lourenco, C. A route for fructose utilization by Escherichia coli involving the fucose regulon. *Proc Natl Acad Sci USA* **103**, 19496–19499 (2006).
43. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, 74 (2013).

44. Götz, A. & Goebel, W. Glucose and glucose 6-phosphate as carbon sources in extra- and intracellular growth of enteroinvasive Escherichia coli and Salmonella enterica. *Microbiology (Reading, Engl)* **156**, 1176–1187 (2010).
45. Brückner, R. & Titgemeyer, F. Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS Microbiol. Lett.* **209**, 141–148 (2002).
46. Görke, B. & Stülke, J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiol.* **6**, 613–624 (2008).
47. Ammar, E. M., Wang, X. & Rao, C. V. Regulation of metabolism in Escherichia coli during growth on mixtures of the non-glucose sugars: arabinose, lactose, and xylose. *Sci. Rep.* **8**, 609 (2018).
48. Aidelberg, G. *et al.* Hierarchy of non-glucose sugars in Escherichia coli. *BMC Syst. Biol.* **8**, 133 (2014).
49. How many species on Earth? About 8.7 million, new estimate says -- ScienceDaily. https://www.sciencedaily.com/releases/2011/08/110823180459.htm.
50. Yang, J. *et al.* Species-Level Analysis of Human Gut Microbiota With Metataxonomics. *Front. Microbiol.* **11**, 2029 (2020).
51. Levin, S. A. Community equilibria and stability, and an extension of the competitive exclusion principle. *Am. Nat.* **104**, 413–423 (1970).
52. Li, C. *et al.* FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnol.* **11**, 92 (2011).
53. Enjalbert, B., Millard, P., Dinclaux, M., Portais, J.-C. & Létisse, F. Acetate fluxes in Escherichia coli are determined by the thermodynamic control of the Pta-AckA pathway. *Sci. Rep.* **7**, 42135 (2017).