Study of *E. coli* metabolic pathways for efficient production of commodity chemicals using synthetic biology and genome engineering

**vom Fachbereich Biologie
der Technischen Universität Darmstadt**

Dissertation

Of

Garima Goyal

First thesis advisor: Prof. Dr. Johannes Kabisch
Second thesis advisor: Prof Dr. Heribert Warzecha

Darmstadt, 2021

Garima Goyal: Study of *E. coli* metabolic pathways for efficient production of commodity chemicals using synthetic biology and genome engineering

Darmstadt, Technische Universität Darmstadt

Year thesis published in TUprints: 2022

Date of the viva voce 12.11.21

Acknowledgments

Firstly, I gratefully acknowledge my Ph.D. advisor, Prof Dr. Johannes Kabisch, for providing me with an opportunity to pursue my Ph.D. under his supervision. I am thankful for his scientific advice and evaluation of the thesis. I would like to acknowledge my second dissertation advisor Prof Dr. Heribert Warzecha, for his support and consideration of my Ph.D. work. I am grateful to my defense examiners Prof Dr. Dominik Niopek and Prof Dr. Beatrix Suess for their valuable time and scientific expertise.

All of my thesis work was supported by Joint BioEnergy Institute (JBEI). I want to express my deep gratitude towards my JBEI supervisor Dr. Nathan J Hillson, for his continuous support and encouragement. Nathan has been incredibly motivating and supportive towards me pursuing the Ph.D., for which I am very thankful. Despite his busy schedule, he made himself available and provided valuable guidance and feedback, which were vital for completing this thesis.

I would like to convey my deep appreciation to Dr. Harry J Beller for conceptualizing (along with Dr. Nathan J Hillson) and his mentorship in the FABIO project. Working with and mentoring my undergraduate researchers Kara Dane and Alexander Earsley during this work was a remarkable experience, and I thank them for their contribution to the project. Many people from the Berkeley Lab have supported this research over the last few years in many different ways. In particular, I would like to acknowledge Shipra Gupta, Robert Evans, Sebastian Palluk, Deepti Tanjore, Di Liu, Gina Geiselman, Jennifer Chiniquy, Nurgul Kaplan, Joshua McCauley, Xi Wang, Aram Kang, Jorge Alonso-Gutierrez, Taek Soon Lee, Zachary Costello, and Robert Haushalter for all their help and guidance. I extend my thanks to JBEI's admin and lab ops team for their unconditional support with routine lab activities. I

Words are not enough to express my deep sense of gratitude towards my parents, Surinder Kumar Goyal and Sneh Lata, for their unconditional love, support, and trust in me. I am forever indebted to you for teaching me to follow my instincts and stay positive. I would like to pay my highest regards to my elder brother, Varun Goyal, for encouraging me to pursue new directions and guiding me through thick and thick. To my sister-in-law, Aditi Goyal, niece, and nephew for their love and always being there for me. I would like to acknowledge my husband's sister Arika Goyal and her family for their continuous love and support.

Last but not least, to my husband Puneet Gupta and my daughter Myra Gupta, finishing this thesis wouldn't have been possible without your unconditional love and support. Puneet, thank you for being there with me during the highs and lows of this journey, for your guidance and continuous encouragement to stay focused. Myra, thank you for inspiring me unknowingly through your inquisitive mind and never-ending questions.

This milestone became a reality due to the undying love and motivation of my family. For that, I dedicate this thesis to them.

"I knew that if I failed, I wouldn't regret that, but I knew the one thing I might regret is not trying." --Jeff Bezos

Summary

Although bio-based fuels and chemicals offer an appealing and more sustainable alternative to traditional petroleum-based fuels and chemicals, their widespread acceptance has been partially obstructed by their limited industry-level titer and yield improvements. Historically, *E. coli* developed for chemical production has been modified with intuition-based biochemical changes of genes and their expression at the plasmid level on a trial-and-error basis. This type of work can be high cost and labor-intensive, and very unstable for industrial applications. Here, we present our work on the employment of advanced synthetic biology and gene editing tools that can accelerate the understanding and engineering of the microbial metabolic pathways directed towards the systematic and stable improvement of process, yields, and rates for biofuels and bio-products production.

The overall objective of the current study is focused on studying and optimizing *E. coli* metabolic pathways for the efficient production of target bio-based compounds. We performed engineering of two *E. coli* metabolic pathways for optimizing the titers of two target compounds through a shared systematic approach of utilizing gene editing, strain development and optimization, and synthetic biology tools. Chapter three of this thesis involves the parallel integration and chromosomal expansion of the isoprenoid pathway in the *E. coli* genome for improved isopentenol titers. In this study, we enabled integration and independent expansions of three pathway components across multiple loci. Suicide vectors were used to achieve high-efficiency site-specific integration of sequence-validated multigene components and a heat-curable plasmid was introduced to obviate *recA* deletion post pathway expansion. Through 3-dimensional expansion, we generated libraries of pathway component copy number variants to screen for improved titers. Machine learning studies were conducted to predict the gene expression for the

6

isopentenol titers. Polynomial regressor statistical modeling of the production screening data suggested that increasing copy numbers of all isopentenol pathway components would further improve titers. From the library of engineered strains with isoprenoid pathway components integrated and expanded in the genome, the best strain produced 344 mg/mL of isopentenol in the absence of selection pressure.

Chapter four focuses on the metabolic engineering of fatty acid biosynthesis pathway for efficient production of fatty alcohols in *E. coli*. Due to tight regulation of endogenous fatty acid biosynthesis pathway, growth essential pathway genes were removed from the genome via CRISPR-Cas9 and placed on the replicable and repressible bacterial artificial chromosome (BAC). Another BAC module that consisted of a heterologous pathway for fatty alcohols production was introduced. Both these modules in the defragged genome were confirmed to be orthogonal and independent of each other. BAC-A module was introduced to support the cell growth in the absence of native genes and BAC-B module's main function was to express heterologous pathway genes for producing high titers of fatty alcohols. CRISPR-Cas9 was developed for simultaneous and consequential multigene deletions of twenty-seven growth essential native pathway genes. A resulting engineered strain with defragged genome harboring BAC-A and BAC-B was tested for fatty alcohols under different express/repress conditions and performed 2.5 times better than an engineered non-FABIO strain (consists of BAC-B without repressing the growth essential genes).

Zusammenfassung

Obwohl biobasierte Kraftstoffe und Chemikalien eine attraktive und nachhaltigere Alternative zu den traditionellen erdölbasierten Kraftstoffen und Chemikalien darstellen, ist die Herstellung in industriellem Maßstab bis heute schwierig. Historisch gesehen wird *E. coli* für die Produktion genutzt, und biochemischen Veränderungen von Genen und deren Expression auf Plasmid-Ebene wurden meist auf Intuition basierend auf einer Trial-and-Error-Basis durchgeführt. Dieser Ansatz kann sehr kosten- und arbeitsintensiv sein, und ist für industrielle Anwendungen daher nur mäßig geeignet. Hier stellen wir unsere Arbeit zum Einsatz fortschrittlicher Werkzeuge der synthetischen Biologie und des Gene Editing vor, die das Verständnis und das Engineering der mikrobiellen Stoffwechselwege beschleunigen können, und somit systematisch und robuste Verbesserungen der Produktion von Biokraftstoffen und Bioprodukten ermöglichen.

Das Gesamtziel der Arbeit ist die Untersuchung und Optimierung von *E. coli* Stoffwechselwegen für die effiziente Produktion von biobasierten Chemikalien. Zwei *E. coli*-Stoffwechselwege wurden untersucht, und die Titer von zwei Zielverbindungen wurden mit Hilfe von Gene Editing, Stammentwicklung und Werkzeugen der synthetischen Biologie optimiert. Kapitel drei dieser Arbeit umfasst die parallele Integration und chromosomale Erweiterung des Isoprenoid-Stoffwechselwegs in das *E. coli*-Genom zur Verbesserung des Isopentenol-Titers. In dieser Studie nutzten wir die Integration und unabhängige Expansion von drei Stoffwechselwegkomponenten in mehreren Loci. Suizid-Vektoren wurden verwendet, um eine hocheffiziente ortsspezifische Integration von sequenzvalidierten Multigen-Komponenten zu erreichen, und ein hitzemodifizierbares Plasmid wurde eingeführt, um die recA-Deletion nach der Pathway-Expansion zu vermeiden. Durch 3-dimensionale Expansion

generierten wir Bibliotheken von Pathway-Komponenten mit unterschiedlichen Kopienzahlen, um verbesserte Titer zu erreichen. Maschinelles Lernen wurden genutzt, um die Genexpression für die Isopentenol-Titer vorherzusagen. Die statistische Modellierung der Produktions-Screening-Daten mit einem polynomialen Regressor legte nahe, dass eine Erhöhung der Kopienzahl aller Isopentenol-Stoffwechselweg-Komponenten die Titer weiter verbessern würde. Aus der Bibliothek der manipulierten Stämme mit genomisch integrierten und erweiterten Isoprenoid-Stoffwechselweg-Komponenten produzierte der beste Stamm 344 mg/mL Isopentenol wenn kein Selektionsdruck vorlag.

Kapitel vier konzentriert sich auf das Metabolic Engineering des Fettsäurebiosyntheseswegs zur effizienten Produktion von Fettalkoholen in *E. coli*. Aufgrund der engen Regulation des endogenen Fettsäurebiosyntheseswegs wurden die wachstumsrelevanten Gene mittels CRISPR-Cas9 aus dem Genom entfernt und auf einem replizierbaren und repressiblen bakteriellen künstlichen Chromosom (BAC) platziert. Ein weiteres BAC-Modul, bestehend aus einem heterologen Stoffwechselweg für die Produktion von Fettalkoholen, wurde eingeführt. Es wurde bestätigt, dass sich diese beiden Module im defragmentierten Genom orthogonal und unabhängig voneinander verhalten. Das BAC-A-Modul wurde eingeführt, um das Zellwachstum in Abwesenheit von nativen Genen zu unterstützen, die Hauptfunktion des BAC-B-Moduls war die Expression von Genen des heterologen Stoffwechselweges zur Produktion hoher Titer von Fettalkoholen. CRISPR-Cas9 wurde für die gleichzeitige Multigen-Deletion von siebenundzwanzig wachstumsrelevanten nativen Stoffwechselwegen entwickelt. Ein daraus resultierender Stamm mit defragmentiertem Genom, der BAC-A und BAC-B beherbergt, wurde unter verschiedenen Expressions- und Repressionsbedingungen auf Fettalkohole getestet und schnitt 2,5-mal besser ab als ein manipulierter nicht-FABIO-Stamm (der aus BAC-B besteht, ohne die wachstumsrelevanten Gene zu unterdrücken).

Graphical Summary
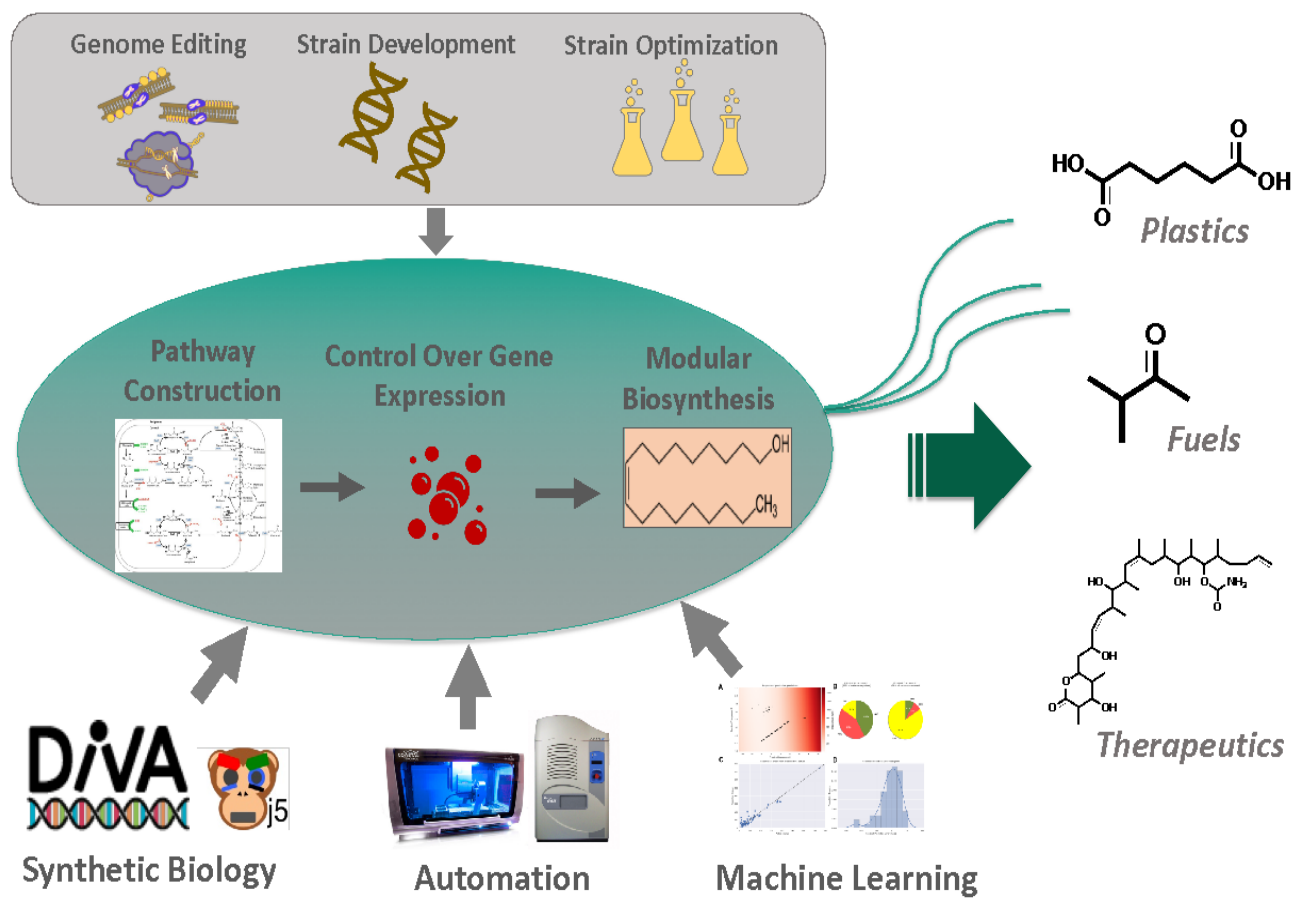
List of Publications

Publications generated from the present thesis

1. **Garima Goyal**, Zak Costello, Jorge Alonso-Gutierrez, Aram Kang, Taek Soon Lee, Hector Garcia Martin, and Nathan J. Hillson. "PIACE – Parallel integration and chromosomal expansion of metabolic pathways into *E. coli." ACS Synthetic Biology,* Oct 2018.

   - Author Contributions: G.G., J.A.G., T.S.L., and N.J.H. conceived the project. G.G., Z.C., and N.J.H. designed the experiments. G.G. performed all the lab experiments. Z.C. performed the machine learning analysis. G.G., Z.C., and N.J.H. analyzed the results. T.S.L, H.G.M., and N.J.H. supervised the work. G.G., Z.C., T.S.L, H.G.M., and N.J.H. wrote the manuscript. All authors have given approval of the final version of the manuscript.

   - Sections of this publication were used in Section 3.3 and 3.4 and supplemental section 6.4 of this thesis.

2. Paul Opgenorth, Zak Costello, Takuya Okada**, Garima Goyal**, Yan Chen, Jennifer Gin, Veronica Benites, Markus de Raad, Trent R. Northen, Kai Deng, Samuel Deutsch, Edward E. K. Baidoo, Christopher J. Petzold, Nathan J. Hillson, Hector Garcia Martin, and Harry R. Beller. "Lessons from Two Design–Build–Test–Learn Cycles of Dodecanol Production in *Escherichia coli* Aided by Machine Learning." *ACS Synthetic Biology,* May 2019.

   - Author Contributions: H.R.B., N.J.H., G.G., P.O., and S.D. conducted Design activities; G.G., N.J.H., P.O., and T.O. conducted Build activities; P.O. and T.O. conducted BioLector incubations and ancillary analyses (e.g., glucose); P.O. and H.R.B. conducted GC−MS analyses; C.J.P., Y.C., and J.G. conducted proteomics analyses; M.dR., K.D., and T.N. conducted NIMS analyses; E.B. and V.B. conducted metabolite analyses; Z.C. and H.G.M. conducted Learn analyses (machine learning);

H.R.B., Z.C., and P.O. analyzed the data. The manuscript was written by H.R.B., P.O., Z.C., and H.G.M., and all authors contributed to or refined the text.

- Sections of this publication were used in 4.3 and 4.4 and supplemental section 6.4 of this thesis.

3. **Garima Goyal,** Robert Evans, Kara Dane, Shipra Gupta, Harry R Beller, and Nathan J Hillson. "Metabolic engineering of Fatty Acid biosynthesis pathway for independent and orthogonal oleochemicals production." In preparation.

- Author contribution: G.G., H.R.B., N.J.H. conceived the project. G.G. designed and performed all the experiments. G.G. supervised K.D. and S.G. on the preparation and confirmation of constructs. The manuscript is being written by G.G and N.J.H.

Other selected publications

1. Shen Long Tsai, **Garima Goyal**, Wilfred Chen, 2010. "Surface display of a functional mini-cellulosome by intracellular complementation using a synthetic yeast consortium: Application for cellulose hydrolysis and ethanol production." *Applied and Environmental Microbiology,* Nov 2010.

2. **Garima Goyal,** Shen Long Tsai, Bhawna Madan, Wilfred Chen, 2011. "Simultaneous cell growth and ethanol production from cellulose by an engineered yeast consortium displaying a functional mini-cellulosome." *Microbial Cell Factories,* Nov 2011.

3. Gregory Linshiz, Nina Stawaski, **Garima Goyal**, Nathan J Hillson. "PR-PR: Cross-Platform Laboratory Automation System." *ACS Synthetic Biology*, Jan 2014.

4. Steve CC Shih, **Garima Goyal**, Nathan J Hillson, Anup K Singh. "A Versatile Microfluidic Device for Automating Synthetic Biology." *ACS Synthetic Biology,* June 2015.

5. **Garima Goyal**, Nick Elsbree, Mike Fero, Nathan J Hillson, Gregory Linshiz. "Repurposing a microfluidic formulation device for automated DNA construction". *PloS One,* Nov 2020.

Contents

Chapter 1

Scientific Background

1.1 *Escherichia coli* as a microbial host

German bacteriologist Theodor Escherich originally discovered *Escherichia coli* 1885 [1-3]. It is a non-spore-forming, Gram-negative, and facultative anaerobic bacterium found in the human colon. *E. coli* has become the best-characterized organism due to extensive investigation and development in the last century [4]. It has also become one of the most applied micro-organisms for metabolic engineering [5-9] and synthetic biology because of its many physiological and biochemical characteristics such as rapid doubling time, easy culture conditions, metabolic plasticity, and ability to consume a variety of carbon sources [8].

The development of many molecular cloning techniques and genetic tools made it easy to manipulate *E. coli* genetically, enhancing our ability both to study its physiology and engineer new phenotypes. Rapid strain development allowed by these techniques has led to significant reduction of costs for industrial development, and *E. coli* and other micro-organisms are a preferred choice for commercial product formation. *E. coli* can naturally consume a wide range of substrates such as glucose, xylose, arabinose, galactose, and fatty acids and convert them into products of interest, making it an ideal candidate for biofuels and biochemicals production [2, 10-12]. The ability to break down waste substrates like lignocellulosic agricultural residues, residential residues, switchgrass, and oilseed crops etc. into simple sugars and metabolize them into product of interest is a strong advantage a microbial host like *E. coli* possesses [13-15]. Despite *E. coli*'s ability to consume various carbon sources, glucose is the preferred carbon source of *E. coli* and is readily obtained hydrolysis of either feedstock. It's important to understand the metabolic regulation of biosynthetic product pathways and toxic effects of fuel molecules and inhibitors [5, 8, 16, 17].

Recent advances in metabolic engineering [5, 8, 18, 19], synthetic biology [20-22], and systems biology [23-25] have played a major role in enabling us to improve the naturally existing pathways, introduce the heterologous pathways, and synthesize the de-novo pathways for high yields and titers of the desired chemicals. In addition, the development of new sequencing technologies to identify the genetic changes, understand the diversity, and characterize the genetic makeup of the organisms are facilitating the discovery of new compounds and robust production of existing ones at optimized levels.

1.1.1 Biofuels and biochemicals production

There are over 6000 products that are derived from petroleum and petroleum-based components. Other than gasoline and other fuels, everyday items like plastic, fragrances, cosmetics, dyes, and textiles also contain petroleum-based resources (Figure 1.1). Their demand is continuously increasing due to the ever-rising population. This increased demand of petroleum-derived products has raised concerns such as depletion of fossil fuel reserves, unequal distribution of reserves, and most importantly, unprecedented calamity to humankind due to increased carbon emission [26-28]. Their excessive usage is causing severe environmental problems and a climate crisis. To cope with these problems, we urgently need to find alternatives to reduce the dependency on fossil fuel reserves [29-31].
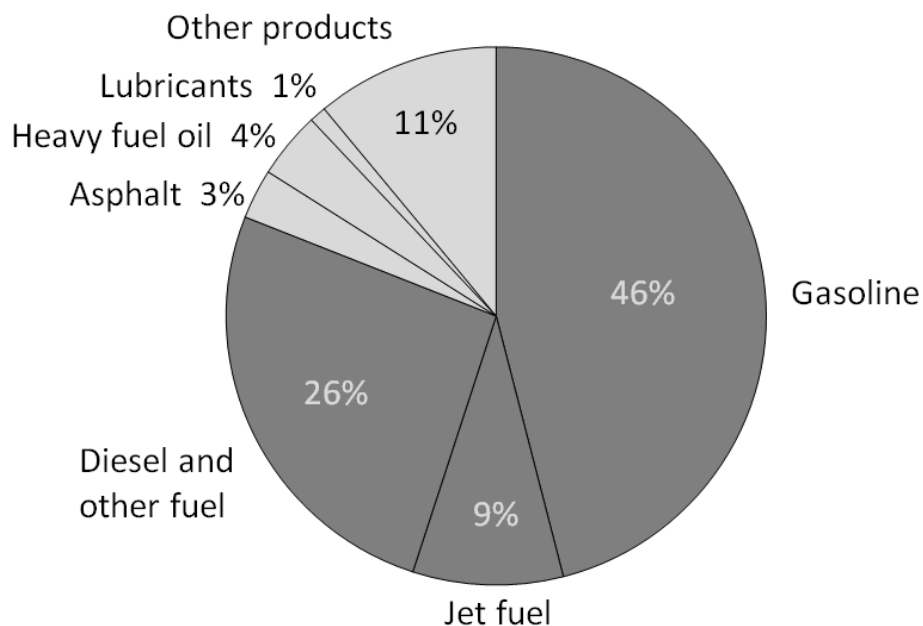
Figure 1.1. Distribution of fossil fuels-based products with 46% gasoline, 26% diesel and other fuels, 9% jet fuels, and 19% all other products. [35]

1.1.2 Sustainable and renewable energy

For the past few decades, scientists worldwide have been working tirelessly to find alternate renewable and sustainable sources for replacing fossil fuels. Among many different energy alternatives, biofuels are the most environmentally friendly energy source [32-37]. They are a favorable choice of fuel due to their renewability, biodegradability, and net-zero carbon emissions. In the recent developments, not just fuels like gasoline and diesel, other fossil-fuel based products are being replaced by sustainable and renewable products. Producing chemicals and materials from non-edible renewable biomass using microorganisms has been gaining momentum. Microbial production often requires mild conditions such as ambient temperature and pressure, can be cultured in a secured tank and doesn't pose any threat to mankind. No toxic

solvent or metal catalysts are added for their growth. Market Research Future reported in 2019 that the global market of bio-based chemicals is estimated to grow with a compound annual growth rate (CAGR) of 10.47% to reach USD 97.2 billion by 2023. This thesis focuses on discovering biologically derived components to find sustainable and high-quality replacements for fossil fuel-driven products.

1.2 Engineering of metabolic pathways in *E. coli*

Microbial fermentation techniques have traditionally been used for food products such as bread, wine, and cheese, etc., and provide valuable metabolites such as alcohols, ketones, fatty acids, and amino acids [38-40] . Research on the microbial production of useful materials has gained attention as an alternative production method of petrochemical products. Microorganisms producing these products have the advantages of generating them through their endogenous metabolic pathways. Scientists have further engineered microbes towards optimized and efficient productivities by modifying their metabolic pathways using synthetic biology techniques.

Biofuels can be produced through various metabolic pathways: ethanol pathway, keto acid pathway, isoprenoid pathway, CoA-dependent reverse beta-oxidation pathway, and fatty acid biosynthesis pathway [4, 5, 16, 17, 41, 42]. Other than bioethanol, the other most prevalent biofuels produced by engineering metabolic pathways are higher chain alcohols and hydrocarbon fuels. High chain alcohols and hydrocarbon fuels have high energy content and superior fuel properties, which make them a good contender to substitute gasoline, diesel, and jet fuel depending on the chain length. Still, it may not be possible for them to entirely replace fossil-based fuels due to less availability of non-food biomass.

E. coli is known to be one of the best-studied prokaryotes concerning its metabolome [43-46]. It has pretty robust and tunable metabolic pathways compared with other microorganisms. Thus, *E. coli* has been utilized as a model microbe and an industrial producer for a variety of chemicals [39, 47, 48]. The focus should be on the accumulation of the metabolic precursors to achieve high yield and productivity with engineered *E. coli*. For example, accumulation of acetyl-CoA would be desirable for pathways engineered to produce iso-propyl alcohol and n-butanol. Accumulation of precursors like acyl-CoA can target multiple metabolites like free fatty acids (FFA), fatty alcohols (FOH), and methyl ketones (MK), etc. For improving the accumulation of target precursors, enhancement of target metabolic flux is a commonly used approach. Simultaneously, strategies for decreasing the metabolic fluxes towards the competing compounds are also used. One common way to achieve this is by overexpression or deletion of relevant pathway genes when re-designing the endogenous pathway. Several previous reports have successfully used this strategy for improving the metabolic flux, thereby facilitating the optimized production of desired compounds. Metabolically engineered microbes could play an essential role in replacing the currently used non-renewable resources to produce the biochemical building blocks that potentially meet economic and productivity targets [49-52].

There are many successful examples of biochemicals produced from the metabolic engineering of *E. coli*. Some of them are ethanol [2], higher alcohols [53], fatty acids [6], amino acids [54], shikimate precursors [55], terpenoids [56], polyketides [57], and polymeric precursors of 1,4-butanediol [16]. In this thesis, the focus is mainly on the metabolic engineering of two *E. coli* pathways: isoprenoid pathway for producing isopentenol and fatty acid biosynthesis pathway for producing fatty acids.

1.2.1 Fatty Acid Biosynthesis pathway in *E. coli* for fatty alcohols production

Fatty alcohols, also known as oleochemicals, are composed of hydrophobic acyl chains with hydroxyl moieties that provide them with amphiphilic properties [58, 59]. Despite their widespread synthesis by organisms like marine bacteria, algae, terrestrial plants, and insects, fatty alcohols in their free form are relatively rare and found in low abundance in nature.

In recent years, significant effort has been made to produce these hydrophobic compounds (fatty alcohols) via fatty acid biosynthesis for use as a precursor for liquid transportation fuels or commodity chemicals [60, 61]. The amphiphilic nature of fatty alcohols makes them excellent chemical components of detergents, emulsifiers, and cosmetics (Figure 1.2). The current global market for fatty alcohol production is estimated at 6.8 billion USD and is expected to reach greater than 10 billion USD by 2023 (Global Info Research, 2020). Recent advancements in synthetic biology and metabolic engineering have shifted the focus to produce these specialized oleochemicals by engineered microbial cell factories [60, 62].

Fatty alcohols can be made by microorganisms endogenously via the fatty acid biosynthesis pathway [63-65]. Its formation depends on the availability of substrates fatty acyl-Coenzyme A (acyl CoA), fatty acyl-carrier protein (ACP), or free fatty acids and the redox donor NADPH. Engineering of the heterologous pathways can improve the availability of these substrates.
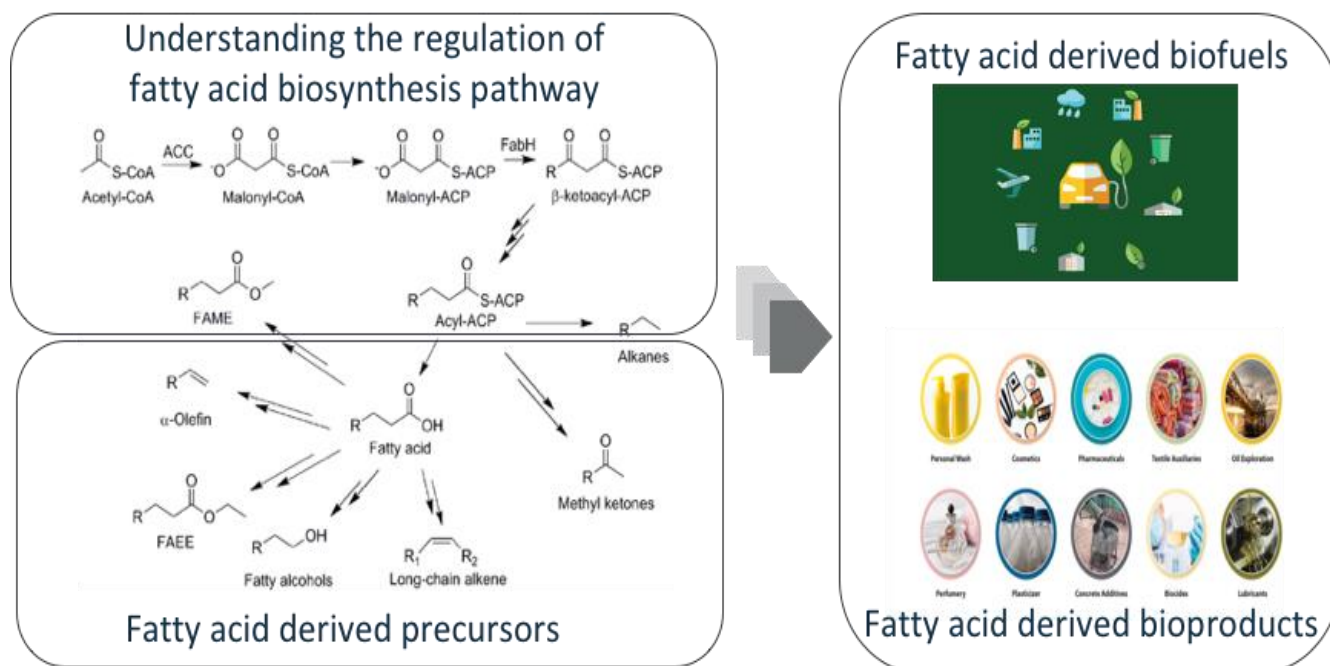
Figure 1.2. Summary of fatty acid pathway components, fatty acid derived precursors, and biofuels and bioproducts. [60]

1.2.1.1 Fatty acid biosynthesis and degradation

Endogenous fatty acid biosynthesis and degradation has three main components – (1) Initiation of the fatty acid metabolism from acetyl-CoA, (2) elongation of fatty acid chain from Malonyl-ACP and Acetyl-CoA, and (3) β-oxidation from Acyl-CoA to acetyl-CoA. In the first step, carboxylation of acetyl-CoA to malonyl CoA catalyzed by an ATP-dependent acetyl-CoA carboxylase (ACC) complex occurs. ACC complex is encoded by four genes, *accA, accB, accC*, and *accD*, also known as accABCD complex. Type II FAS complex composed of monofunctional enzymes encoded by nine separate genes (*fabA, fabB, fabD, fabF, fabG, fabH, fabI, fabZ, acpS*) is used in the elongation of fatty acid chain. Post elongation, fatty acyl-ACP, fatty acyl-CoA,

and free fatty acids are made from the fatty acyl chains and can act as substrates for fatty alcohol production. The length of the fatty acid chain is controlled by the enzymatic activity that releases the fatty acyl molecules from FAS. However, reports suggest that they may be further elongated by *FabF* in *E. coli* [63, 66-68].

Fatty acid degradation (FAD) takes place via cyclic β-oxidation pathway generating acetyl-CoA as the terminal product. The degradation pathway is catalyzed by the enzymes encoded by the fad regulon, which are responsible for the transport and activation of long-chain fatty acids and their β-oxidative cleavage into acetyl CoAs. β-oxidation exclusively uses acyl-CoA thioesters and not acyl-ACP. This process occurs in the cytosol by the combined action of *FadD* and *FadE* in *E. coli* [69].

1.2.1.2 Fatty acid regulation

E. coli tightly regulates the fatty acid biosynthesis and degradation pathway. Multilayer control, including transcriptional, translational, and substrate-level regulation limits the supply of substrates [67, 69]. Transcriptional control occurs via the FadR regulatory protein. When long-chain fatty acids of more than 14 carbons are present in the growth medium, they are converted to acyl-CoAs, which bind to FadR, causing the release of the regulatory protein from the operator and thus de-repression of transcription of the fad genes [70-72]. Thus, FadR functions as a repressor of many genes and operons mainly involved in fatty acid degradation and as a regulator of the cellular processes associated with it. Moreover, many fad genes are positively regulated by the cyclic-AMP receptor protein/cyclic-AMP system [73, 74]. On the other hand, FabR represses the transcription of *fabA* and *fabB* in the presence of unsaturated fatty acids. In addition, metabolic flux through the biosynthesis pathway is

limited by the long-chain acyl-ACP. They act as feedback inhibitor on genes *fabH*, *fabI* and ACC complex [75-79].

Table 1.1. Fatty alcohols production in *E. coli [93]*

Properties	Host strain	Product-forming enzyme	Mutations	Media	Shake flask titer (mg L^{-1})	Fed batch titer (mg L^{-1})	Productivity (mg L^{-1} h^{-1})	Yield (g/g)	References
Even chain C12–C18	C41 (DE3)	AcACR1	'tesA, fadD, ΔfadE	M9	60	ND	ND	ND	[6]
	BL21 (DE3)	AbACR1	BTE, fadD	M9	37.5	598	24.94	0.001	[80]
	MG1655	MaFAR	3X BTE, FadD	M9	ND	1650	13.75	0.134	[81]
	BL21(DE3)	MaFAR	'tesA, fadD	M9	422	1750	16	0.028	[82]
	BL21 (DE3)	MmCAR	'tesA, AHR, Bssfp	Minimal	350	ND	ND	0.05	[83]
	MG1655	MaFAR	fadD, 'tesA, ΔfadE, araD:acpS (C. glutamicum)	M9	3500	ND	50	0.126	[84]
	DH5α	SeAAR	ybbO, fadR, ΔplsX, fabZ	M9	270	1989	41.43	ND	[85]
	MG1655	MaFAR	ΔtesCB, ΔldhA, Δpta, ΔackA	LB	760	6330	126	ND	[86]
	MG1655	MaFAR	ΔfadE	M9 (2mM NH$_4^+$)	180	ND	ND	0.12	[87]
	DH5α	SeAAR	zwf, ybbO, Δedd, Δpps, ΔldhA, ΔaceA, ΔpoxB, Δpta, ΔpflB, ΔplsX	Modified mineral medium	1506	12500	ND	ND	[88]
Octanol	MG1655	MaFAR	ΔaraBAD, ΔfadE, ΔfadD::P$_{TRC}$CpFatB1*-lacI, ΔackA-pta::P$_{TRC}$-Mt-FadD6-lacI, ΔpoxB::P$_{TRC}$-Mt-FadD6, ΔFadBA::P$_{TRC}$Ma-ACR-lacI	Mineral medium 2% glycerol	1273		53	0.064	[89]
Octanol	BL21 (DE3)	Mm CAR	BsSFP, At tes3 AHR	LB	62		4.4	12	[90]
Odd (C$_{2n-1}$) chain	BL21 (DE3)	O. sativa αDOX	'tesA FadR, AHR	M9 yeast extract, glycerol	105	1950	105	0.018	[91]
Branched chain	UB1005	MaFAR	ΔfabH, ΔfadE, ΔldhA, fadR, ilvCD, S. aureus fabH, B. subtilis lplA, leu operon, and bkd operon	M9	170	350 branched chain, 737 total fatty alcohol	9	ND	[92]

1.2.2 Isoprenoid pathway in *E. coli* for isoprenoids production

Isoprenoids, also known as terpenoids, are hydrocarbon-based compounds with different carbon-chain lengths and mainly consist of a large variety of natural products in structure and function. Since petroleum-derived fuels such as gasoline, diesel, and jet fuels are a complex mixture of compounds with different chain lengths, saturation levels, and chemical structures, microbially produced isoprenoids are gaining attention as advanced biofuels due to their similar characteristics. Their additional properties like higher energy density,

higher octane numbers, lower water miscibility, and better low-temperature fluidity than other biofuels make them a stronger candidate for potential petroleum replacement [94, 95]. Additionally, they have a wide range of commercial applications as rubbers, flavor, fragrance, pharmaceuticals, biopesticides, and precursors to critical industrial chemicals. Approximately 95% of isoprenoids (derived from isoprene) are used for rubber synthesis, and approximately 20 million tons of rubber is produced annually [96]. Commercial isoprene used in synthetic rubber production is produced by cracking of petroleum which is an unrecoverable and environment-unfriendly resource [97]. Another option is producing them naturally from plants, but plant cultivation is generally season-dependent and requires land, nutrients, and water, which hinders eco-friendly large-scale production [94, 98, 99] in most cases, although rubber is now successfully made commercially from dandelions by a company called Continental. Metabolic engineering of microbial systems for isoprenoid production is found to be a very promising alternative to address these problems. A well-studied *E. coli* has been mostly used for producing isoprenoids by engineering the endogenous methylerythritol phosphate (MEP) pathway and introducing the heterologous mevalonate (MVA) pathway.

All isoprenoids are derived from the universal C_5 prenyl phosphate precursors isopentenyl diphosphate (IPP) and, its isomer, dimethylallyl diphosphate (DMAPP). More than 50,000 compounds can be generated from these two precursors [100]. These precursors can be generated from the MEP pathway present in most prokaryotes and chloroplasts and the MVA pathway present in most eukaryotes. The MVA pathway generally supplies precursors for the production of sesquiterpenes like farnensenes, triterpenes like brassinosteroids, and dolichol. The MEP pathway generally supplies precursors for the biosynthesis of diterpenoids, tetraterpenes like carotenoids, and plant-derived gibberellins, and chlorophylls [101, 102].
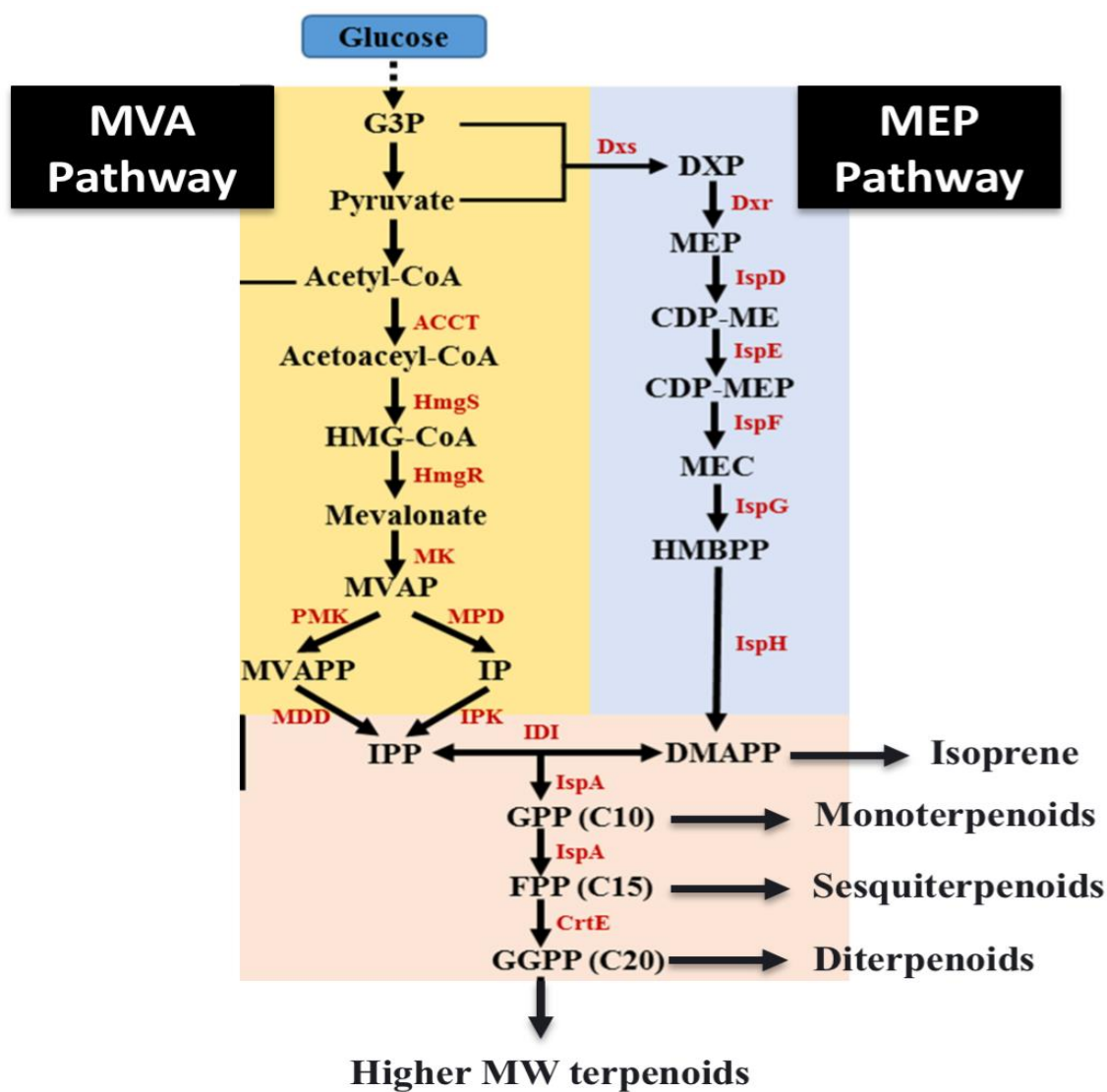
Figure 1.3. The two isoprenoid pathways to central metabolism, MVA pathway and MEP pathway (adapted from Wang et al 2020) [104].

Then, downstream steps utilize IPP and DMAPP to produce several isoprenoids [100] [103]. As shown in the figure, glyceraldehyde 3-phosphate (G3P) and pyruvate serve as precursors for the MEP pathway and three acetyl-CoA are

the precursors for the MVA pathway. Precursors for both pathways for the synthesis of IPP are derived from glycolysis. Once synthesized, IPP and DMAPP can be converted to isoprene and many other products.

Metabolic engineering of microbial cell factories to improve isopentenol and isoprene productivity has been investigated for more than a decade now. In fact, the research performed for enhancing the isopentenol and isoprene production is also instructive and meaningful for engineering the bio-production of other isoprenoids. All the isoprenoids are biologically synthesized via either the MEP or the MVA pathway and metabolic engineering methods for the biosynthesis of isoprenoids are mutually referential. However, metabolic engineering of the pathway into native microbial hosts can result in the accumulation of IPP and DMAPP, which are toxic for bacterial growth [105]. Therefore, balancing the accumulation and consumption of IPP and DMAPP is very important for isoprenoid bio-production [103]. Unlike other isoprenoids, isoprene and isopentenol are direct derivatives of IPP and DMAPP, which makes controlling the accumulation of IPP and DMAPP easier than other isoprenoids.

1.2.2.1 Challenges in isoprenoid pathway engineering

Balancing catalytic efficiencies, energy demands, and carbon is critical in achieving high titers of isoprenoid after introducing pathway engineering strategies. Avoiding the accumulation of toxic intermediates and overcoming pathway regulation are vital points in successfully engineering the MEP and MVA pathways in *E. coli*. Both pathways fulfill their carbon demands from central carbon metabolism; therefore, they must compete with many other reactions for the necessary precursors [106-109]. Ensuring a good precursor supply is of great importance for a well-functioning isoprenoid pathway. For the energy requirement, both pathways need a similar amount of energy in the

form of ATP/CTP. Consumption of reducing cofactors is higher in the MEP pathway with 3 NADPHs affecting three steps of the pathway compared to the MVA pathway with 2 NADPHs affecting only one step of the pathway. Balancing enzyme expression is very important in improving the overall flux and avoiding toxic intermediates. It can be done if rate-limiting steps of the pathways can be identified e.g., in the case of the MVA pathway, these would be HMG-CoA reductase (HMGR) and HMG-CoA synthase (HMGS).

1.3 Challenges of strain engineering

Overexpression of the rate-limiting genes to direct more carbon flux into the desirable endogenous or heterologous pathway and prevent the accumulation of toxic intermediate metabolites is needed to better optimize biochemical productivity. To accomplish this, most strategies include a plasmid-based gene expression system due to its easy and fast manipulation and controllable expression via plasmid copy number, and promoter and RBS strength. This strategy, however, suffers from the extensive metabolic burden on the cellular machinery under stressful fermentation conditions. More importantly, this leads to genetic instability, especially for heterologous genes, resulting from segregational and structural instability [110, 111]. Owing to these limitations of the plasmid-based system, researchers have been focusing on genome engineering of the metabolic pathways resulting in stable and consistent product formation. Chromosomal gene editing is also preferable because antibiotics and other selective agents are not needed to maintain the genetic changes, decreasing the overall biomanufacturing cost and environmental impacts. Thus far, significant progress has been made in the development of genetic manipulation tools that allow modifications of homologous and heterologous DNA in micro-organisms.

There are several widely used genomic recombination methodologies available for *E. coli*. Some of the traditional gene-editing techniques are still in use. Methods like SacB based approach using counterselection markers, a phage derived recombinases (RecET and λ-Red) [112, 113], applying double-stranded [114, 115] or single-stranded donor DNAs [116], or inducing double-stranded breaks (DSBs) in a chromosomal target using I-SceI [113, 117] have been studied used extensively for manipulating metabolic pathways at the genomic level. Some of the most common genome editing techniques that have emerged in recent years are zinc-finger nucleases (ZFNs) [118-121], transcription activator–like effector nucleases (TALENs) [121-127], and the RNA-guided CRISPR-Cas nuclease system [128-133].

One of the most commonly used methods employs λ-Red mediated homologous recombination. Three phage-derived λ-Red proteins are necessary for carrying out dsDNA recombination: Gam, Exo, and Beta. It requires as small as 35 bp homology arms on each side of the desired alteration for making precisely defined insertions, deletions, and point mutations in *E. coli*. Various strategies like I-SceI cleavage-facilitated recombination, knock-in/knock-out (KIKO) [134], CIChE [111], PIACE [110], vector aided insertion have been developed based on the λ-Red system. λ-Red based strategies are very efficient for chromosomally integrating DNA molecules of size less than 2000 bp. However, recombination efficiency drops significantly for chromosomal integration of DNA molecules larger than 2000 bp. To integrate larger DNA modules, they need to be sub-divided into segments and then inserted into *E. coli* genome iteratively. This method has the potential to integrate large synthetic pathways with multiple rounds of iteration for integration.

A recent and prevalent genome engineering technique is known as CRISPR-Cas9. It is short for clustered regularly interspaced short palindromic repeats

and CRISPR-associated protein 9. CRISPR-Cas9 is proven to be faster, cheaper, and more accurate and efficient than other existing genome editing methods, which has led to a lot of excitement in the scientific community. It has been developed very efficiently in several prokaryotes and eukaryotes, including (but not limited to) *E. coli [135], Saccharomyces cerevisiae* [136], *Streptomyces* spp. [137], higher plants [138], *Bombyx mori [139]*, *Drosophila* [140], and human cell lines [141-143]. In CRISPR-Cas9, the cas9 nuclease is guided by an sgRNA consisting of a 20-nucleotide guide sequence and a scaffold and makes a double-stranded break (DSB) at a target genomic locus. After cleavage by Cas9, two possible outcomes happen for the damage DNA repair at the target site (1) Non-Homologous End Joints (NHEJ) (2) Homology Directed Repair (HDR). NHEJ process pathway, which is known to be error-prone, works when the repair template is absent. Double stranded breaks introduced by Cas9 are re-ligated which leaves scars in terms of random insertion/deletion mutations at the target site. Alternatively, if a repair template is provided in the form of a linear DNA or plasmid, high fidelity and precise editing at the junction site takes place through the HDR pathway. HDR typically has much lower and variable frequencies than NHEJ. Most eukaryotes can very efficiently repair DSBs via NHEJ. In eukaryotes, small insertions and/or deletions (indels) are introduced at the cut site during NHEJ repair which leads to gene disruptions in the target gene. Most bacteria including *E. coli* lack the repair mechanism by NHEJ. DSBs introduced in these organisms normally lead to cell death due to the lack of NHEJ. To escape the cell death, DNA damage is primarily repaired via HDR with sister chromatids, where the template DNA replace the damaged DNA fragment by recombination. Although the Cas9 system is more commonly used in *E. coli,* other less common Cas systems are also being explored and utilized currently.
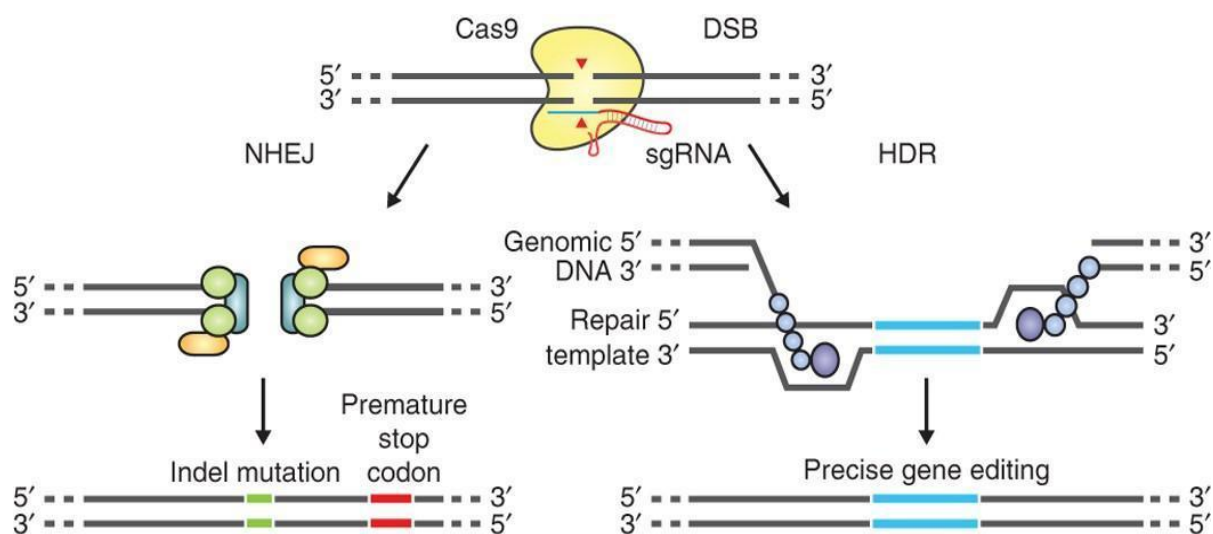
Figure 1.4. CRISPR Cas9 generated repair mechanisms. DSB induced by Cas9 can be repaired by one of the two pathways: (1) Non-Homologous End Joints (NHEJ) (2) Homology Directed Repair (HDR) [144]

1.5 Synthetic biology tools to aid the metabolic engineering

The advent of synthetic biology has accelerated the engineering of microbes for a wide range of applications. The key-enabling technologies in this space include the ever-decreasing cost of DNA synthesis and sequencing, standardization of DNA assembly procedures, genome editing, computer-aided design (CAD), synthetic biology workflows, development and use of

automation and high throughput instrumentation, and machine learning methods. These new approaches, along with engineering principles, are being utilized for systematic, product-independent metabolic engineering strategies. The global bio-foundries frequently use Design-Build-Test-Learn (DBTL) cycles that utilize the above-mentioned technologies to accelerate strain engineering for biochemical production. The DBTL cycle represents a framework that facilitates systematic metabolic engineering targeted towards increased efficacy and generalizability. There are several benefits of using DBTL infrastructure for rapid design construction and characterizing the metabolic pathways, (1) Large scale bioprocesses can be examined in each DBTL round (2) use of automation and high throughput in the design of experiment and construction will boost the speed and efficiency of iterations. (3) standard assembly procedures will improve the repeatability and comparability across different batches of experiments. Lastly, (4) an enormous amount of data acquisition enables the development of machine learning tools to explore and predict gene and protein-related properties.

Some of these DBTL approaches are applied in the present thesis work for accelerated strain engineering. We have attempted to achieve standardization in the method development for DNA construction and validation by using high-throughput robotic workflows. Machine learning tools were developed to accomplish the predictability of the experimental data and guide the new set of experiments. Gibson and golden-gate assembly methods were used for all the plasmid assembly strategies which improved the repeatability across the experimental batches. Efforts were made towards systematizing the genome editing approaches like CRISPR-Cas9 and the λ-red system.

Chapter 2

Materials and Methods

2.1 Materials

2.2.1 List of disposables, reagents, consumables, and devices

Table 2.1. List of commonly used Disposables

Consumables	Vendor	Stock Number
Nunc 96-well culture plates with lids	Fisher Scientific, Pittsburgh, PA, USA	12-565-331
Nunc 384-well plates, with lid	Fisher Scientific, Pittsburgh, PA, USA	GSS-164688
Q-trays with divider	Molecular Devices	X6029
384-well PCR plates	BioRad Hercules, CA, USA	HSP3801
384-well Echo PP plates	Labcyte, San Jose, CA, USA	P-05525
96-well hard-shell PCR plates	Bio-Rad Hercules, CA	
96-well semi-skirted PCR plates	Eppendorf, Enfield, CT, USA	89049-178
Tips (P10, P20)	Rainin, Oakland, CA, USA	RT-L10F(SR-L10F)
Tips (P100, P200)	Rainin, Oakland, CA, USA	RT-L200F(SR-L200F)
Tips (P1000)	Rainin, Oakland, CA, USA	RT-L1000F
Tips, Matrix, 1250 μL	Matrix Technologies	8045
Kimwipes (140 ct)	Kimberly-Clark	

96-well no-skirted PCR plates	VWR, West Chester, PA, USA	
96-well half-skirted PCR plates	USA Scientific, Enfield CT, USA	1402-9220
Marsh Seals	Fisher Scientific, Pittsburgh, PA, USA	1044394
Foil Seals	Fisher Scientific, Pittsburgh, PA, USA	FS-100
Airpore Seals	Qiagen, Hilden, Germany	19571
Weigh boats 3X3	VWR, West Chester, PA, USA	89106-766
Centrifuge 50 mL tubes	VWR, West Chester, PA, USA	
Biomek AP384 P30 Tips, 30 µl	ISC Bioexpress, Kaysville, UT, USA	P-3692-3
Biomek AP96 P200 tips	ISC Bioexpress, Kaysville, UT, USA	
Biomek AP96 P20 Tips, Sterile Barrier, 20 µl	USA Scientific, Enfield CT, USA	1061-2810
BioRobotix Tips for Picker	VWR, West Chester, PA, USA	72830-226
300 uL Nimbus filter tips	Coastal Genomics, Burnaby, British Columbia, Canada	235903
Kimwipes EX-L (280 ct)	Kimberly-Clark, Irving, TX, USA	34155

COOLRACK(R) XT PCR96	Sigma Aldrich St. Louis, MO, USA	BCS-529

Table 2.2. List of commonly used reagents

Reagents	Vendor	Stock Number
Agencourt AMPure XP, 5 ml	Beckman Coulter, Brea, CA, USA	A63880
Carbenicillin	Sigma-Aldrich, St. Louis, MO, USA	C1389-1G
Chloramphenicol, 20 mL, 10 mg/mL	Teknova, Hollister, CA, USA	C0312
Kanamycin sulfate	Sigma-Aldrich, St. Louis, MO, USA	60615-25G
Streptomycin sulfate	Sigma-Aldrich, St. Louis, MO, USA	S6501-5G
PCR purification kit	Promega, Madison, WI, USA	A9282
Qubit BR assay kit	Invitrogen, Carlsbad, CA, USA	Q32853
NEB Gibson assembly master mix	New England Biolabs (NEB) Ipswich, MA, USA	E2611L
NEB Golden-gate HF Assembly Kit	New England Biolabs (NEB) Ipswich, MA, USA	E1601L

LB Media + 7.5% glycerol	ISC Bioexpress, Kaysville, UT, USAG	MBLE-7970
SOC media	GrowCells, Irvine, CA, USA	MBLE-8003/G-3276-50
LB Broth	Thermo Fisher Scientific, Waltham, MA, USA	10855001
Isopropyl β- d-1-thiogalactopyranoside (IPTG)	Thermo Fisher Scientific, Waltham, MA, USA	15529019
L-(+)-Arabinose	Sigma-Aldrich, St. Louis, MO, USA	5328-37-0
QuantiFluor® ONE dsDNA System	Promega, Madison, WI, USA	E4871
mix n go competent cells (DH10B)	Zymo Research, Irvine, CA, USA	T3020
Nuclease Free Water	Ambion, Austin, TX, USA	AM9932
Q5 high fidelity PCR master-mix	New England Biolabs (NEB) Ipswich, MA, USA	M0492L
Plasmid miniprep kit	Qiagen, Hilden, Germany	27106
DpnI enzyme	New England Biolabs (NEB) Ipswich, MA, USA	R0176S
Competent cells TOP10	Invitrogen, Carlsbad, CA, USA	C404050
Loading Dye 6X	New England Biolabs (NEB) Ipswich, MA, USA	B7021S
Generuler mix DNA ladder	MBI Fermentas, Vilnius, Lithuania	SM0334

SyBR safe	Invitrogen, Carlsbad, CA, USA	S33102
1%, 1.5%, 2% Nimbus Analytical gels	Coastal Genomics, Burnaby, British Columbia, Canada	
0.5%, 1%, 1.5%, 2% Nimbus size selection gels	Coastal Genomics, Burnaby, British Columbia, Canada	CG-10600-01
Dual dyes with buffer 100-1000bp, 300-3000bp, 1-10kb	Coastal Genomics, Burnaby, British Columbia, Canada	CG-10700-01
Ultrapure Agarose	Invitrogen, Carlsbad, CA, USA	15510-027
50X TAE buffer	Invitrogen, Carlsbad, CA, USA	24710-030
1% bleach	Fisher Scientific, Pittsburgh, PA, USA	S66362
Ethanol (100%)	Sigma-Aldrich, St. Louis, MO, USA	459844-500ML
Isopropanol	Millipore Sigma, Burlington, MA, USA	67-63-0
n-Hexane	Millipore Sigma, Burlington, MA, USA	110-54-3
Dodecane	Sigma-Aldrich, St. Louis, MO, USA	112-40-3
Dodecanol	Sigma-Aldrich, St. Louis, MO, USArai	
PCR Purification Kit	Qiagen, Hilden, Germany	28104

Gel DNA Purification Kit	Zymo Research, Irvine, CA, USA	D4007
Plasmid DNA Miniprep Kit	Qiagen, Hilden, Germany	27106
Genomic DNA Prep Kit	Promega, Madison, WI, USA	A1120
Nextera XT Library Preparation Kit	Illumina, San Diego, CA, USA	
Nextera XT Index Kit v2	Illumina, San Diego, CA, USA	
SsoAdvanced Universal SYBR Green Supermix (2X)	Bio-Rad, Hercules, CA, USA	
Ampure XP magnetic beads	Beckman Coulter, Indianapolis, IN, USA	A63880
Qubit DNA HS Assay	Invitrogen, Carlsbad, CA, USA	

Table 2.3. List of commonly used devices

Equipment	Vendor	Stock Number
Pipettes (P10, P20, P100, P200, P1000)	Rainin, Oakland, CA, USA	RL-10/20/100/200/1000

Multichannel pipettor 2-10 µl, 20-200 µl, 100-1000 µl	Rainin, Oakland, CA, USA	
Microcentrifuge for 1.5 mL microcentrifuge tubes	Fisher Scientific, Pittsburgh, PA, USA	05-090-128
Vortex Genie 2	VWR, West Chester, PA, USA	G-560
Bench-top centrifuge	-	-
Water bath (42 °C)	-	-
MilliQ Water Dispenser	Millipore Sigma, Burlington, MA, USA	-
Plate-Loc Sealer	-	-
Magnet for ampure beads	EpiGentek, Farmingdale, NY, USA	Q10002-1
Incubators and shakers (30, 37, 60 °C) 20-400 rpm, 1 inch orbit	Eppendorf, Enfield, CT, USA	-
Biomek FXp, Fx, NX, NXs	Beckmann Coulter, Brea, CA, USA	-
Nimbus Size Selection	Hamilton/Coastal Genomics, Burnaby, British Columbia, Canada	-
QP460 Plater/Picker	Molecular Devices, San Jose, CA, USA	-
Echo 550 Liquid Handler	Labcyte, San Jose, CA, USA	-

Biolector Pro - bioreactor	M2p-labs, Baesweiler, Germany	-
GC-FID	Agilent Technologies, Santa Clara, CA, USA	-
ZAG DNA Analyzer	Agilent Technologies, Santa Clara, CA, USA	-
Miseq System	Illumina Inc, San Diego, CA, USA	-
-Veriti thermocycler	Applied Biosystems	4375786
Bioanalyzer	Agilent Technologies, Santa Clara, CA, USA	-
F200 Pro microplate reader	Tecan, Männedorf, Switzerland	

2.1.2 Software tools

DNA software tools, including DeviceEditor (v5.1.16) [145], j5 [146], ICE (v5.9.4) [147], DIVA (v5.1.16), and VectorEditor (v16.1.35) were used for designing and visualizing DNA constructs and for designing the DNA assembly protocols used for their construction. Bio::BioStudio (v2.10) was used for designing the CRISPR target sites.

2.2. Media

All media used in this study are sterilized by autoclaving at 121 ^{0}C for 20 minutes. All media were supplemented with appropriate antibiotics for bacterial growth.

2.2.1. LB media

Lysogeny Broth (LB) media was used for routine *E. coli* growth. Its composition is 10 g/L Tryptone, 10 g/L NaCl, and 5 g/L Yeast Extract. It was prepared by adding 25 g of LB in 1 L of dH_2O.

2.2.2. EZ rich media

Table 2.4. Composition of EZ rich media

Volume (mL)	Component	Final Concentration
100	10X MOPS Mixture	1X
10	0.132 M K_2HPO_4	1.32 mM K_2HPO_4
100	10X ACGU	1X
200	5X Supplement EZ	1X
100	20% Glucose	2%
490	sterile H_2O	-

2.2.3. M9 MOPs media

Table 2.5. Composition of M9 MOPs media

Volume	Component	Final concentration

200 mL	20% Glucose	2%
400 mL	5x M9	1X
115.4 mL	1.3M Mops	75mM
4 mL	1M MgSO$_4$.7H$_2$O	2mM
200 µL	10mg/mL thiamine.HCl	1mg/L
2 mL	1000X micronutrient	1X
2 mL	1000X FeSO$_4$•7H$_2$O	1X
to 2000mL	H$_2$O	
200uL (last in)	1M CaCl$_2$.2H$_2$O	0.01mM

2.3 Molecular biology methods

2.3.1 PCR Amplification

All fragments used for plasmid construction were amplified using Q5 hot start high fidelity 2X Master Mix and touchdown PCR conditions. All the genomic modifications were confirmed by colony PCRs using Taq DNA Polymerase. Veriti thermocycler was used for all the touchdown and colony PCR reactions. All the PCR compositions and thermocycler conditions are listed in Tables 2.6 to 2.9.

Table 2.6. PCR composition for amplifying DNA parts for assembly reactions

Component	Volume/Reaction	Final concentration

Q5 high fidelity PCR MasterMix (2X)	25 µL	1X
Forward Primer (10 µM)	2.5 µL	0.5 µM
Reverase Primer (10 µM)	2.5 µL	0.5 µM
DNA Template	varies	5-10 ng
Nuclease Free Water	varies	-
Total Volume	50 µL	-

Table 2.7. Touchdown PCR thermocycler conditions

Cycle step	Temperature [^{0}C]	Duration [s]	number of cycles
Initial denaturation	98	30	1
Denaturation	98	10	10
Annealing	based on primer Tm with decrease in annealing temperature of 0.5 °C per cycle starting from cycle 2	30	
Elongation	72	20 s/kb	
Denaturation	98	10	25
Annealing	based on primer Tm	30	
Elongation		20 s/kb	
Final elongation	72	120	1
Hold	10	-	1

Table 2.8. Colony PCR composition for screening genomic edits

Component	Volume/Reaction	Final concentration
Taq Polymerase (5U/µL)	0.2 µL	1 U
Taq DNA polymerase Buffer (10X)	2 µL	1X
Forward Primer (10 µM)	1 µL	0.5 µM
Reverase Primer (10 µM)	1 µL	0.5 µM
DNA Template	varies	5-10 ng
Nuclease Free Water	varies	-
Total Volume	20 µL	-

Table 2.9. Colony PCR thermocycler conditions

Cycle step	Temperature [◦C]	Duration [s]	number of cycles
Initial denaturation	95	180	1
Denaturation	95	10	
Annealing	55-65	30	30
Elongation	72	20 s/kb	
Final elongation	72	300	1
Hold	10	-	1

2.3.2 Real time qPCR for gene copy number quantification

Estimation of gene copy number was carried out using qPCR. Relevant genes were amplified from the genome using SYBR Green Super Master Mix. BioRad CFX 96 thermocycler was used for qPCR reactions. Threshold cycles (Ct) were determined with CFX Manager (BioRad, Hercules, CA) software for all samples. For the normalization of qPCR data, *hcaT* reference gene was used. Reaction composition and thermocycler conditions for the qPCR are listed in the table 2.10 and 2.11.

Table 2.10. qPCR composition for quantifying gene copy numbers

Component	Volume/Reaction	Final concentration
SYBR green Super Mix (2X)	10 µL	1X
genomic DNA	2 µL	50 ng
Forward Primer (10 µM)	1 µL	0.5 µM
Reverase Primer (10 µM)	1 µL	0.5 µM
Nuclease Free Water	varies	-
Total Volume	20 µL	-

Table 2.11. qPCR thermocycler conditions

Cycle step	Temperature [∘C]	Duration [s]	Number of cycles
Initial denaturation and enzyme activation	95	180	1
Denaturation	95	15	40

Annealing	55	30	
Extension	72	30	
Melt curve	55-95 in 0.5 °C increments	30	1

2.3.3 DpnI digestion and DNA purification

Following PCR amplification, residual (methylated) DNA template in each PCR reaction was DpnI digested and purified using a NIMBUS size selection robot. Gel purified DNA fragments were cleaned up with AMPure magnetic beads to remove the buffer that was used for elution by the NIMBUS robot; this step increased the DNA concentration by eluting in less reagent water to improve the Gibson assembly efficiency. DNA quantification was done using Qubit DNA High Sensitivity kit.

2.3.4 Gibson Assembly and *E. coli* transformation

Gibson assembly [148, 149] was performed by mixing the purified DNA parts in an equimolar ratio. The composition of the assembly reaction is listed in Table 2.12. Assembly was then incubated at 50 °C for 1 hour in a Veriti thermocycler.

Table 2.12. Gibson assembly reaction composition

Component	Volume/Reaction	Final concentration
Gibson assembly Mix (2X)	5 µL	1 X
DNA parts	-	equimolar

Nuclease free water	-	-
Total volume	10 µL	-

Two µL of the Gibson assembly reaction was transformed into chemically competent Top10 cells. To transform the assembled plasmid, TOP 10 cells were thawed out on ice and 2 µL of the assembled DNA was added. After incubating for 15 minutes, cells were subjected to heat shock at 42 ^{0}C for 30 seconds. Then, cells were recovered at 37 ^{0}C for 1 hour with shaking at 200 rpm and plated on LB agar plates containing respective selection markers. Next day, colonies were picked after incubating plates at 37 ^{0}C overnight.

Eight colonies per construct were selected and grown overnight in 1 mL Lysogeny broth (LB) with the appropriate antibiotic in a 96-well plate. Cell cultures were used for DNA sequencing on a MiSeq system for sequence verification.

2.3.5 Chromosomal integration of gene of interest

Strains containing pKD46 were cultured from a single colony in LB media with 100 µg/mL ampicillin at 30 °C and 200 rpm to an OD600 nm of 0.25−0.3. Arabinose was then added (to 10 mM) to the cell cultures to induce the expression of λ-red proteins, and the cells were then grown for another 30 min (OD of 0.5−0.6). Electrocompetent cells were then prepared by washing the cell pellets once with chilled sterilized water, and twice with 10% chilled and sterilized glycerol. Cells were transformed with the corresponding suicide vector. Following electroporation, cells were grown out in 1 mL of SOC media (no antibiotic) and incubated at 37 °C for 3 h. Cells were then plated on LB agar plates with the addition of the appropriate antibiotic overnight at 37 °C. Suicide vector integrations were verified by colony PCR following protocol and thermocycler conditions in the Tables 2.3.3 and 2.3.4 and then by Sanger

sequencing of the amplicons. pKD46 was cured from the verified integrants through outgrowth at 37 °C, isolation of colonies on LB agar (no ampicillin), and subsequent counter-screening of colonies on LB agar with ampicillin (no growth indicating the absence of pKD46).

2.3.6 Re-sequencing Plasmids via Next-Generation Sequencing

All the plasmids and amplicons generated from colony PCR of gene modifications were confirmed via Miseq sequencing. Nextera libraries were constructed using the Illumina Nextera XT Library Preparation Kit and the Nextera XT Index Kit v2. Liquid transfers were carried out on the BioMek NX, Biomek FX, and Labcyte Echo 550 acoustic liquid dispensing system. During tagmentation, DNA was fragmented by transposases, and transposon terminal sequences were appended as adapters. Unique barcode combinations were added during PCR amplification. The tagmentation and PCR amplification steps were performed with several modifications to the standard Nextera protocol: 1) all the steps were performed in a single 384-well PCR plate, 2) the tagmentation reaction volume was reduced to 1 μL using the Echo 550, and 3) a heat-kill step (15 minutes at 70 °C) was added to circumvent the need for purification after tagmentation. For the qPCR step, 7.5 μl SsoAdvanced Universal SYBR Green Supermix (2X) and 5.5 μl Nuclease free water were added to the tagmentation reaction. qPCR thermocycler conditions were followed as described in table X. The end point fluorescence values from qPCR on a CFX384 were used to normalize concentrations using the Echo 550. This enabled Ampure XP bead purification of the pooled library in a single tube. The purified library was quantified using the Qubit DNA HS assay, and the fragment size profile was determined using the Bioanalyzer 2100. Sequencing was performed on the MiSeq using 2×300 cycles. MiSeq reads were demultiplexed using the embedded MiSeq Reporter (MSR) software. Read mapping and variant calling were performed using BWA26, and GATK27.

Table 2.13. Composition of tagmentation reaction

Component	Volume [µl]
Tagmentation Buffer	0.50
Tagmentation Enzyme	0.25
DNA template/ Glycerol Stock	0.25
Total	1.00

Cycle step	Temperature [^{0}C]	Time [min]
DNA Tagmentation	55	5
Heat Kill	70	15
Hold	4	-

Table 2.14. qPCR thermocycler conditions

Cycle step	Temperature [$\circ$C]	Duration [s]	number of cycles
	72	180	1
Denaturation	98	30	1
Denaturation	98	10	
Annealing	63	30	19
Extension (plate read)	72	180	
Hold	4	-	1

2.4 Analytical methods

2.4.1 Production Runs in a BioLector Pro Microbioreactor

Plasmids were transformed into the production strain and colonies were picked, grown overnight in LB at 37 °C, and the plasmids were re-sequenced with a MiSeq system (Illumina). In preparation for the final production run, strains were acclimated overnight in M9-MOPS growth medium with 2% glucose at 30 °C. The fatty alcohols production experiments were run in triplicate on a BioLector Pro microbioreactor with a 48-well flat-bottom plate. The cultures were inoculated with 50 µL of stationary-phase preculture, grown in M9-MOPS medium with 2% glucose, and were induced with 0.1 mM IPTG at the start of the incubation period. The total culture volume was 1 mL, with a 200 µL overlay of dodecane spiked with 500 µg (2.5 µg/µL) of the internal standard for fatty alcohol and aldehyde analysis, 1-dodecan-d25-ol (98 atom % D). The BioLector Pro was run at 1000 rpm and 3 mm orbit at 30 °C and ambient chamber pressure. The total time for each run was 40 hours, at which time the cultures were harvested by centrifugation at 20 817g for 4 min at 4 °C. The dodecane overlay was used for fatty alcohol and aldehyde analysis and the supernatant was sampled for glucose.

2.4.2 Fatty Alcohols measurement by Gas Chromatography−Mass Spectrometry (GC−MS)

Fatty alcohols captured in the dodecane overlay were analyzed by GC−MS. Before analysis, 1 µL of the dodecane overlay was diluted 100- fold with hexane. Electron ionization (EI) GC−MS analyses with a quadrupole mass spectrometer were performed with a model 7890A GC (Agilent Technologies, Santa Clara, CA) coupled to a HP-5ms fused silica capillary column (30-m length, 0.25-mm inner diameter, 0.25-µm film thickness; Agilent) and an HP

5975C series mass selective detector (Agilent); 1 µL injections were performed by a model 7683B autosampler (Agilent). The GC oven was programmed from 40 °C (held for 3 min) to 295 °C at 15 °C/min; the injection port temperature was 250 °C, and the transfer line temperature was 280 °C. The carrier gas, ultra-high purity helium, flowed at a constant rate of 1 mL/min. Injections were split-less, with the split turned on after 1 min. For full-scan data acquisition, the MS scanned from 50 to 600 atomic mass units at a rate of 2 scans per s. Internal standard quantification was performed. The internal standard, perdeuterated dodecanol, or 1-dodecand25-ol (98 atom % D), was spiked into the dodecane overlay such that a 1:100 dilution in hexane for GC− MS analysis would result in a 25 ng/µL concentration in the diluted extract. GC−MS standards (C12, C14, C16, C18 fatty alcohols) at three concentration levels (5, 20, and 50 ng/µL) all contained 25 ng/µL of the deuterated internal standard.

2.4.3 C5 alcohols detection by Gas Chromatography−Mass Spectrometry (GC−MS) and Gas Chromatography−Flame ionization detector (GC-FID)

Isopentenol was measured by GC-MS and GC-FID [9]. Three colonies were picked for each construct and grown overnight at 37°C in 5 ml of LB medium with antibiotic. The overnight cultures were inoculated into 5 ml of EZ Rich defined medium (Teknova) with 0.2% glucose and antibiotic to an initial OD_{600} of 0.05. Cultures were grown at 37°C for 3 h, induced with 0.5 mM IPTG, and grown for 20 h at 30°C. A 700-µl volume of culture was sampled, mixed with a solution of chloroform-methanol (80:20) spiked with 50 µg/ml of butanol (Sigma-Aldrich) as an internal standard, vortexed for 15 min, and centrifuged for 1 min at 13,000 × *g*. The chloroform layer was removed for analysis by GC.

The GC-MS data were collected in full-scan mode (*m/z* 50 to 300) using a Tr-Wax column (0.25 mm by 30 m, 0.25-µm film thickness, Thermo Electron) on a PolarisQ GC-MS apparatus with a TriPlus autosampler (Thermo Electron). The carrier flow rate was 1.2 ml min^{-1}, and the inlet temperature was set to 200°C. The oven program was 40°C for1.20 min, an increase from 40 to 130°C at 25°C min^{-1}, and an increase from 130 to 220°C at 35°C min^{-1}. The solvent delay was set at 3.40 min. Samples were normalized using the butanol internal standard and quantified using authentic standards. 3-Methylbutanol standards were purchased from Sigma-Aldrich. 3-Methyl-3-butenol and 3-methyl-2-butenol were purchased from the Tokyo Chemical Industry Co., Ltd.

The GC-FID data were collected using a Tr-Wax column (0.25 mm by 30 m, 0.25-µm film thickness; Thermo Electron) on a Focus GC apparatus with a TriPlus autosampler (Thermo Electron). The carrier was set at a constant pressure of 300 kPa, and the inlet temperature was set to 200°C. The oven program was 40°C for 1.50 min and then an increase from 40 to 110°C at 15°C min^{-1}. Samples were normalized using the butanol internal standard and quantified using authentic standards.

Chapter 3

Optimization of Isopentenol titer by Parallel Integration and Chromosomal Expansion of Isopentenol pathway in *E. coli*

3.1 Abstract

There is an urgent demand for stable and predictable expression of metabolic pathway genes for the industry-level bio-products formation from the biomass-derived sugars. Genomic changes related to pathway engineering are known to be more robust as opposed to plasmid-based systems which often suffer from instability and result in highly variable product titers and yields. Here, we report development of Parallel Integration and chromosomal expansion (PIACE) that enables independent expansions of pathway components across multiple chromosomal loci. We applied the PIACE strategy to isoprenoid pathway for improved and stable isopentenol titers. Isoprenoid pathway was expanded in the genome across three different loci in *E. coli* DH1 using suicide vectors to achieve high-efficiency site-specific integration of multi-components and heat-curable plasmid to facilitate the λ-red homologous recombination and remove *recA* deletion post pathway expansion. Three-dimensional pathway component expansion yielded libraries of 125 copy number variants that were screened for improved titers of isopentenol. Three expression configurations namely chromosomal, hybrid and plasmid-based systems were constructed and compared for isopentenol production, and gene stability. Expectedly, the chromosomal system outperformed the hybrid and plasmid-based systems in gene stability and achieved 105 mg/L of isopentenol titers. For further improvement of titers, polynominal regressor statistical modeling suggested increasing the copy number of all pathway components in the genome.

3.2 Introduction

Engineering of heterologous isoprenoid pathway [7, 42, 97, 150] in *E. coli* has been of greatest interest for more than a decade due to its ability to produce a remarkably diverse group of compounds. Isopentenol (branched five carbon alcohol), produced from the heterologous isoprenoid pathway, is an attractive target compound for microbial production due to its desirable fuel properties and importance as potential platform chemicals [56, 98].

Many metabolic engineering tasks require importing heterologous pathways into host organisms. Plasmids are preferred over chromosomal genetic manipulations for such tasks directed towards developing metabolic pathways and synthetic biology tools in *E. coli* because manipulating plasmids is faster, easier, and less expensive than the single copy chromosomal integration [151-155]. Plasmids also provide high expression of pathway genes due to multiple copies which can yield high product titers. However, very often replicating plasmids can suffer from extensive genetic instability due to structural and allele distribution instability under high metabolic pressure [156]. Chromosomal integration of heterologous pathways provides better long-term stability and reduced cell to cell variability in copy number and expression levels in the absence of selection pressure, making such strains more suitable for industrial-scale processes or downstream engineering. DNA integration in *E. coli* has been established since a long time ago by using site-specific recombination between *attP* and *attB* sites. However, such strategies are limited to chromosomal positions containing these sites. Also, overall recombination efficiencies decrease sharply with the increased size of the inserted DNA fragment [157-159].

There are multiple chromosomal integration methods available for *E. coli* that have been developed/improvised over the years [132, 160, 161]. Most common approaches are flippase-flippase recognition targets (FLP-FRT) recombination [162], phage mediated site-specific recombination via conditional-replication integration and modular (CRIM) plasmids [163, 164], transposon-mediated gene transposition [165, 166], λ-red-mediated homologous recombination [158, 167-169], and knockin/knockout (KIKO) vectors [134]. While most of these approaches suffer from limited gene expression due to only a single copy of pathway integration, only a small subset provides the rapid and controllable chromosomal integration of multiple gene copies and those are CRIM, chemically inducible chromosomal evolution (CIChE) [111] and chromosomal integration of genes with multiple copies (CIGMC)[170]. CIGMC uses the FLP-FRT approach to achieve variable (colony-to-colony) numbers of chromosomal insertions in a single step. While CIGMC has been shown to achieve up to 15 integrated copies of a gene, average copy numbers were in the range of 2−4, with higher copy numbers occurring at very low frequencies. Copy number of genes in CIGMC approach occurs randomly which provides less control over the pathway expression level. In contrast, CIChE monotonically controls gene copy number through antibiotic titration and has been shown to achieve up to 60 gene copies, which is comparable to multicopy plasmids. PIACE is based on the CIChE approach. We chose CIChE (as a good trade-off between process control and labor/time requirements) for our efforts to genetically stabilize the multiple copies of isopentenol pathway in the genome and therefore, optimize the isopentenol titers.

3.3 Materials and Methods

3.3.1 Plasmid and Strain Construction

Cloning methods for constructing plasmids and strains used in this study are described in section 2.3. *E. coli* DH10B was used as the cloning host for the plasmids pRedi2RecA and pMevTop_AtoB_mvaS_mvaA. *E. coli* PIR1 was used as the cloning host for all the suicide vectors constructed for homologous recombination. *E. coli* DH1 (Supplemental Table 6.1) was used as the production host for isopentenol and RFP/GFP. DNA oligos and templates, PCR conditions, and the DNA assembly method type used to construct each plasmid are listed in Tables 3.1 and 3.2. DNA fragments were amplified following PCR composition and touchdown thermocycling conditions described in Tables 2.3.1 and 2.3.2. Following PCR amplification, residual (methylated) DNA template in each PCR reaction was *DpnI*-digested and purified using a Qiagen gel extraction kit or Nimbus Size Selection, as specified. Gibson and CPEC assembly methods were used for constructing the plasmids as described in table 2.6.1. All suicide vectors have a R6K origin of replication, which can be propagated only in a PIR1 strain background, as well as a 1 kb homologous sequence targeting their respective locus for chromosomal integration. Chloramphenicol (pPIACE_Cm_RFP and pPIACE_Cm_PMK-MK), spectinomycin (pPIACE_Cm_GFP and pPIACE_Sp_NudB-PMD), and gentamicin (pPIACE_Gent_AtoB-mvaS-mvaA) resistance genes were chosen as selectable markers due to their ability to titrate antibiotic resistance roughly linearly (as opposed to an all-or-nothing threshold) with copy number expansion. Homologous recombination was performed in freshly prepared *E. coli* DH1 electrocompetent cells, as described in section 2.3.5. Transformations were otherwise performed in chemically competent cells. Integration of each gene cassette was confirmed by colony PCR and Sanger sequencing (of its colony PCR product, using the corresponding colony PCR primers). Colony

PCR composition and thermocycler conditions are described in table 2.3.3 and 2.3.4. The complete list of DNA oligos used for colony PCR is provided in Table 3.3.

Table 3.1. List of plasmids and their construction details

Plasmid Name	Template	Annealing Temperature ^{0}C	DNA Size (bp)	Gel Purification Method	Assembly Method
pPIACE_Cm_RFP	pAH143	66.7	476	Qiagen gel purification kit	
	E. coli DH1 gDNA	61.8	1025	Qiagen gel purification kit	Gibson
	pBbE5c-RFP	67.3	3261	Qiagen gel purification kit	
pPIACE_Sp_GFP	pAH143	65.7	470	Qiagen gel purification kit	
	E. coli DH1 gDNA	69.0	1029	Qiagen gel purification kit	Gibson
	pCIChE_Spec_GFP	66.8	863	Qiagen gel purification kit	
	pCIChE_Spec_GFP	65.7	1185	Qiagen gel purification kit	
pPIACE_Sp_NudB-PMD	pPIACE_Sp_GFP	62.7	2626	Qiagen gel purification kit	Gibson
	pTrc99A-NudB-PMD	67.2	3727	Qiagen gel purification kit	
pPIACE_Cm_PMK-MK	pPIACE_Cm_RFP	66.7	2459	Qiagen gel purification kit	
	pPIACE_Cm_RFP	64.0	123	Qiagen gel purification kit	Gibson
	pKWG0012	65.7	2778	Qiagen gel purification kit	
pPIACE_Gent_AtoB-mvaS-mvaA	pAH143	68.7	459	Nimbus Size Selection/ 1.5% gel and 300-3000bp marker	Gibson
	E. coli DH1 gDNA	66.7	1053	Nimbus Size Selection/ 1.5% gel and 300-3000bp marker	

	pMCS-4	64.5	850	Nimbus Size Selection/ 1.5% gel and 300-3000bp marker
	pMTSA	65.0	3983	Nimbus Size Selection/ 1% gel and 1-10kb marker
pREDI2RecA	pREDI	60.2	4465	Qiagen gel purification kit
	pREDI	60.8	4000	Qiagen gel purification kit	Gibson
	E. coli MG1655 gDNA	62.4	1099	Qiagen gel purification kit
pCIChE_Spec_GFP	pBbE5a-RFP	64.4	2518	Qiagen gel purification kit
	pCM62-Prha-T7SL-RFP	64.3	270	Qiagen gel purification kit
	pBbE5c-RFP	62.0	1476	Qiagen gel purification kit	CPEC
	pBbB5k-GFP	61.4	863	Qiagen gel purification kit
	MAGE-01	60.5	1115	Qiagen gel purification kit
	pCM62-Prha-T7SL-RFP	62.6	280	Qiagen gel purification kit

Table 3.2. List of oligos for plasmid construction

Plasmid Name	Part Name	Forward Oligo Name	Forward Oligo	Reverse Oligo Name	Reverse Oligo
pPIACE_Cm_RFP	R6K_ori	j5_00001_(R6K_ori)_forward	GGGCCTTTCTGCGTTTATACCATGGCTAATTCCCATGTCAGCCG	j5_00002_(R6K_ori)_reverse	CTTTTACGAGATCGCGGCCGCAAGATCCGGC
	intA3prime_HR	j5_00003_(intA3prime_HR)_forward	ATCTTGCGGCCGCGATCTCGTAAAAGCAGCAGATAGTGGT	j5_00004_(intA3prime_HR)_reverse	GGCGGGGCGTAAATATTCCTGCGATGTACCATGGGA

	CIChE_Cm_RFP	j5_00005_(CIChE_Cm_RFP)_forward	CATCGCAGGAATATTTACGCCCCGCCCTGCCA	j5_00006_(CIChE_Cm_RFP)_reverse	TGGGAATTAGCCATGGTATAAACGCAGAAAGGCCCACCCG
pPIACE_Sp_GFP	R6K_ori	j5_00001_(R6K_ori)_forward	CATCCGGATATCCCCATGGCTAATTCCCATGTCAGCCG	j5_00002_(R6K_ori)_reverse	TGGAGCGAGACCGGCGGCCGCAAGATCCGGC
	fdhF_HR	j5_00003_(fdhF_HR)_forward	ATCTTGCGGCCGCCGGTCTCGCTCCAGTTAATCA	j5_00004_(fdhF_HR)_reverse	GGGGTGCCTAAACGATAGAGTAAAGTGGCGCACGGT
	pLacUV5_GFP	j5_00005_(pLacUV5_GFP)_forward	CGTGCGCCACTTTACTCTATCGTTTAGGCACCCCAGGC	j5_00006_(pLacUV5_GFP)_(stop_codonTAA)_reverse	CGAGTTTGGATCCTTATTTGTAGAGCTCATCCATGCC
	Bba_1006_term_SpecR	j5_00007_(Bba_1006_term_SpecR)_forward	GGATGAGCTCTACAAATAAGGATCCAAACTCGAGTAAGGATCT	j5_00008_(Bba_1006_term_SpecR)_reverse	TGGGAATTAGCCATGGGGATATCCGGATGAAGGCACGA
pPIACE_Sp_NudB-PMD	PIACE_R6K_Spec_BB	DVA00149_(pPIACE_Sp_GFP_backbone)_forward	TGGTCTACCAAAGGAATAAGGATCCAAACTCGAGTAAGGATC	DVA00150_(pPIACE_Sp_GFP_backbone)_reverse	GCCGCGTTGCTGGAGAGTAAAGTGGCGCACGGTC
	pTrc99A_NudB_PMD	DVA00151_(pTrc99A_NudB_PMD)_forward	GCCACTTTACTCTCCAGCAACGCGGCCTTTTTACGG	DVA00152_(pTrc99A_NudB_PMD)_reverse	CCTTACTCGAGTTTGGATCCTTATTCCTTTGGTAGACCAGTCTTTGCG
pPIACE_Cm_PMK-MK	PIACE_R6K_Cm_BB	DVA00153_(PIACE_Cm_BB)_forward	TGGACGTCTTAATTAGGATCCAAACTCGAGTAAGGATCTCCAGGC	DVA00154_(PIACE_Cm_BB)_reverse	GTGCCTAAACGATGCGCAACGCAATTAATGTAAGTTAGC
	prom_LacUV5	DVA00155_(prom_LacUV5)_forward	AATTGCGTTGCGCATCGTTTAGGCACCCCAGGC	DVA00156_(prom_LacUV5)_reverse	AGATCCTTTAGATCCAGATCTTTTGAATTCTGAAATTGTTATCCGC
	PMK_MK	DVA00157_(PMK_MK)_forward	CAGAATTCAAAAGATCTGGATCTAAAGGATCTAGGAGGGAGA	DVA00158_(PMK_MK)_reverse	TCCTTACTCGAGTTTGGATCCTAATTAAGACGTCCACGGCAGATTGGTGTTACC
pPIACE_Gent_AtoB-mvaS-mvaA	R6K_ori	DVA02036_(R6K)_forward	GCGTTTATACCTAGGCCATGGCTAATTCCCATGTCAGCCG	DVA02037_(R6K)_reverse	GCGGCCGCAAGATCCGGC

	lacZ_dh1	DVA02038_(lacZ_dh1)_forward	GCATCGTGGCCGGATCTTGCGGCCGCAGCGGTGCCGGAAAGCTG	DVA02044_(lacZ_dh1)_reverse	GATATCGACCCAAGTACCGCCACCTAACGTGTACCACAGCGGATGG
	GentR	DVA02045_(GentR)_forward	TTAGGTGGCGGTACTTGGGTCG	DVA02046_(GentR)_reverse	GTAAAGCCTGGGGTGCCTAAACGATGGTTGCGATCTTTTCTACGGGG
	AtoB-mvaS-mvaA	DVA02042_(MevTop)_forward	CCATCGTTTAGGCACCCCAGGC	DVA02043_(MevTop)_reverse	GCTGACATGGGAATTAGCCATGGCCTAGGTATAAACGCAGAAAGGCCCACC
pREDI2RecA	predi-I	GG_00046_(pRedi)_forward	CGTCGATAGCCATATGTGATCCTGCTGAATTTCATTACGACC	GG_00073_(predi-I)_reverse	ACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGC
	predi-II	GG_00075_(predi-II)_forward	GTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTG	GG_00077_(predi-II)_reverse	CAGAAACTAACGAAGATTTTTAATAATAAGTCGACCCATGGATTCTTCGTCTGTTTC
	recA	GG_00048_(recA)_forward	AGAATCCATGGGTCGACTTATTATTAAAAATCTTCGTTAGTTTCTGCTACGCCTTCGC	GG_00078_(recA)_reverse	CAGCAGGATCACATATGGCTATCGACGAAAACAAACAGAAAGCG
pCICHE_Spec_GFP	BB_colE1	GG_00121_(BB_colE1)_forward	CCCAACAGTGAAGCCCGGGGAGCAAAAGGCC	GG_00122_(BB_colE1)_reverse	CAACATTGCTTGCTTATATTCCCCAGAACATCAGGTTAATGG
	Tet5prime	GG_00123_(Tet5prime)_forward	CTGGGGAATATAAGCAAGCAATGTTGTCCAGGC	GG_00124_(Tet5prime)_reverse	GGAAAGCGGGCAGTGATCAGCGATCGGCTCGTTGC
	LacI	GG_00125_(LacI)_forward	GAGCCGATCGCTGATCACTGCCCGCTTTCCAG	GG_00126_(LacI)_reverse	GCCTAAACGATCTCGATCCTCTACGCCGG
	GFP_LacUV5	GG_00127_(GFP_LacUV5)_forward	GGCGTAGAGGATCGAGATCGTTTAGGCACCCCAGG	GG_00128_(GFP_LacUV5)_reverse	GTAGTCGGCAAATAATTATTTGTAGAGCTCATCCATGCC
	SpecR	GG_00129_(SpecR)_forward	GCTCTACAAATAATTATTTGCCGACTACCTTGGTG	GG_00130_(SpecR)_reverse	CCATTATTATCATGAGGATATCCGGATGAAGGCAC

Tet3prime	GG_00131_(Tet3 prime)_forward	TTCATCCGGATA TCCTCATGATAA TAATGGTTTCTT AGCACCC	GG_00132_(Tet 3prime)_reverse	CTTTTGCTCCCCG GGCTTCACTGTTG GGGCCGTG

Table 3.3. List of oligos for colony PCR

DNA Oligo Name	DNA Oligo Sequence	Annealing Temperature ˚C
intA_Cm_colPCR_fwd_junctionI	aaaacgacgtgatctcgcct	60
intA_Cm_colPCR_rev_junctionI	aagatgtggcgtgttacggt	60
Cm_intA_colPCR_fwd_junctionII	agggcttctcagtgcgttac	60
Cm_intA_colPCR_rev_junctionII	acaaaaatgaggtgcgtcgc	60
fdhF_lacIq_colPCR_fwd_junctionI	gcaacctgatgcgcaatagg	60
fdhF_lacIq-colPCR_rev_junctionI	ccctgagagagttgcagcaa	60
SpecR_fdhF_colPCR_fwd_junctionII	gtaccaaatgcgggacaacg	60
SpecR_fdhF_colPCR_rev_junctionII	cacaattacaggtcgcagcg	60
lacZ_AtoB_colPCR_fwd_junctionI	tcactggccgtcgttttaca	60
lacZ_AtoB_colPCR_rev_junctionI	taacgcttcagccgttgagt	60
mvaS_lacZ_colPCR_fwd_junctionII	tggctcaggttcagttgttga	60
mvaS_lacZ_colPCR_rev_junctionII	actcgggtgattacgatcgc	60

3.3.1.1 Strains and Plasmids availability

Strains and plasmids developed for this study (Supplemental Table 6.1), along with annotated DNA sequences were deposited in the public instance of the Joint BioEnergy Institute Registry (https:// public-registry.jbei.org/folders/354). Plasmid maps used in this chapter are given in Figure 6.1.

3.3.2 Chromosomal Integration of Suicide Vectors

Suicide vectors were integrated at their respective chromosomal loci using helper plasmid pKD46. Strains containing pKD46 were cultured from a single colony in LB media with 100 μg/mL ampicillin at 30 °C and 200 rpm to an OD600 nm of 0.25−0.3. Arabinose was then added (to 10 mM) to the cell cultures to induce the expression of λ-red proteins, and the cells were then grown for another 30 min (OD of 0.5−0.6). Electrocompetent cells were then prepared, and cells were transformed with the corresponding suicide vector. Following electroporation, cells were grown out in 1 mL of SOC media (no antibiotic) and incubated at 37 °C for 3 h. Cells were then plated on LB agar plates with the addition of the appropriate antibiotic overnight at 37 °C. Suicide vector integrations were verified by colony PCR using oligos listed in Table 3.3 and then by Sanger sequencing of the amplicons. pKD46 was cured from the verified integrants through outgrowth at 37 °C, isolation of colonies on LB agar (no ampicillin), and subsequent counter-screening of colonies on LB agar with ampicillin (no growth indicating the absence of pKD46).

3.3.3 Gene Copy Number Expansion of Genes of Interest

Following suicide vector integration(s), gene copy (CIChE) expansion was carried out using helper plasmid pREDI2RecA. For the RFP/GFP experiments

and hybrid plasmid/chromosomal systems, expansions were carried out in parallel with stepwise simultaneous antibiotic (chloramphenicol and spectinomycin) increases (concentrations doubling per step). For the fully chromosomal system, we observed that simultaneous titration increases of all three antibiotics significantly impaired cell growth. As such, for the fully chromosomal system, expansions were carried out sequentially starting with gentamicin followed by chloramphenicol and spectinomycin. Strains were grown in 5 mL of LB medium supplemented with 100 mM L-rhamnose (inducing the expression of *recA*) and increasing concentrations of antibiotics (one or more of gentamicin, starting concentration 12.5 mg/ L; chloramphenicol, starting concentration 15 mg/L; and spectinomycin, starting concentration 25 mg/L) in 15 mL culture tubes at 30 °C until stationary phase, as previously described. Fifty microliters of culture were then inoculated into 5 mL of fresh LB media with double the concentration of one or more of antibiotics (see above), and the process was repeated as desired or until the gene expansions were saturated (no further growth observed).

3.3.3.1 Curation of Helper Plasmids

pREDI2RecA was cured from all the strains following gene copy (CIChE) expansions. Strains were grown at 42 °C and shaken at 200 rpm overnight in liquid LB media. Thereafter, 5−10 μL of each strain was plated on LB agar, and single colonies were isolated, and replica streaked on LB agar with or without ampicillin, with no growth on LB with ampicillin indicating the loss of pRedi2RecA.

3.3.4 Growth Conditions and Isopentenol Production

Cultures of *E. coli* DH1 harboring heat-sensitive auxiliary plasmids pRedi2RecA or pKD46 were grown overnight in LB medium containing

ampicillin (100 mg/L final concentration) at 30 °C and shaken at 200 rpm in rotary shakers. Sequential integration experiments were performed with appropriate antibiotics in LB media at 37 °C, with one or more of the antibiotics gentamicin, chloramphenicol, and/or spectinomycin provided at final concentrations of 12.5, 15, and 25 mg/L, respectively. Expansion experiments were performed in LB media induced with 100 mM L-rhamnose at 30 °C and shaken at 200 rpm in rotary shakers. Production assays were performed in an EZ-Rich defined medium containing 1% glucose. Briefly, starter cultures were used to inoculate 10 mL of EZ-Rich medium in a culture tube to an OD600 of 0.1. Production cultures were grown in rotary shakers (200 rpm) at 37 °C to an OD600 of 0.6 and induced with 500 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Following induction, cultures were moved to 30 °C for the duration of the assay. Samples were taken after 48 h for isopentenol analysis by GC-FID as described in section 2.4.3. All cultures were grown in 10 mL medium volumes in culture tubes.

3.3.4.1 Fluorescence Measurement by Micro-plate Reader

Red fluorescence proteins and green fluorescence proteins were measured using a Tecan F200 Pro microplate reader. Measurement of RFP fluorescence was performed with excitation/emission at 590/635 nm. Measurement of GFP fluorescence was performed with excitation/emission at 485 nm/530 nm. Captured data were transferred directly into an Excel file for direct data analysis. Fluorescence values were corrected for background fluorescence (as a function of OD600) of the negative control (wildtype DH1).

3.3.4.2 Gene Copy Number Measurement by qPCR analysis

qPCR analysis was used to measure gene copy numbers. qPCR composition and thermocycler conditions are described in tables 2.10 and 2.11. A complete

list of DNA oligos used for qPCR is provided in Table 3.4. *hcaT* gene which maintained one copy in all *E. coli* strains used in this study, was used as a reference gene for normalization for qPCR.

Table 3.4. List of qPCR DNA oligos.

DNA Oligo Name	DNA Oligo Sequence	Annealing Temperature °C
PMK_MK_fwd_qPCR	cctggagaacaacgtggaca	55
PMK_MK_rev_qPCR	tgttccagtcttcttcgtcca	55
NudB_fwd_qPCR	cagcgtggaagagggtgaaa	55
NudB_rev_qPCR	gtgaaaacgatctgccgctc	55
RFP_Cm_fwd_qPCR	gggattggctgagacgaaaa	55
RFP_Cm_rev_qPCR	ataccacgacgatttccggc	55
GFP_Sp_fwd_qPCR	atgtcattgcgctgccattc	55
GFP_Sp_rev_qPCR	gcgagctttgatcaacgacc	55
AtoB_mvaS_mvaA_fwd_qPCR	tcgctagctgacttcgcatc	55
AtoB_mvaS_mvaA_rev_qPCR	tcctgaacgtaaacgctcttga	55
Refgene_HcaT_fwd_qPCR	gatgcactggcgaatacgtg	55
Refgene_HcaT_rev_qPCR	ctgaatcgtcggacggatga	55

3.3.4.3 Stability Tests of three expression configurations and RFP/GFP expressing strains

Stability tests were performed on chromosomal, hybrid plasmid/chromosomal, and two-plasmid isopentenol-producing strains, and on chromosomal RFP/GFP-expressing strains. Strains were grown overnight in LB media with antibiotics and in the absence of *recA* at 37 °C in 10 mL cultures. Cultures were

diluted 1000-fold into fresh LB media without antibiotics and this process was repeated for three cycles. Glycerol stocks were maintained for each cycle. At the end of the experiment, RFP/GFP strains were tested for stability by measuring RFP/GFP fluorescence by microplate reader and *rfp/gfp* copy number by qPCR. Isopentenol producing strains were tested for stability by measuring isopentenol production by GC-FID and copy number of each gene cassette by qPCR.

3.3.5 Machine Learning for Predicting Isopentenol Titer

Machine learning analysis was applied to the prediction of isopentenol titer from the number of plasmid-borne or chromosomally integrated copies of three gene cassettes (GOI1−3). Sixty-five strains with varying copy numbers for each gene cassette were used as a training set (as indicated in Supplemental Table 6.1). Three regression models -Polynomial, random forest, and support vector regressors- were applied to this training set. All regressors were implemented in python using the scikit-learn library, and a jupyter notebook is available from https://github.com/JBEI/PIACE, from which the analyses can be reproduced, and in which the specific implementation details of these models are described. Each model's performance was evaluated using 10-fold cross validation and the comparison of error residuals of the resulting predictions. One data point was held back (strain 7P7N, as indicated in Supplemental Table 6.1) during model development and was tested against fitted models to protect against/assess the extent of overfitting the data.

3.4 Results and Discussion

3.4.1 PIACE Strategy

For PIACE strategy (Figure 3.1), we are using inducible λ-red homologous recombination functionality. pKD46, an arabinose inducible and heat curable plasmid consisting of λ-red proteins (Figure 3.1), is transformed into a *recA*-negative *E. coli* base strain DH1. Three gene cassettes GOI1, GOI2, and GOI3 are delivered sequentially in the form of suicide vectors (R6K origin) with 1 kb homology sequence into the genome. These gene cassettes are integrated with one copy at three different *E. coli* genomic sites *lacZ*, *intA*, and *fdhF* respectively via homologous recombination.
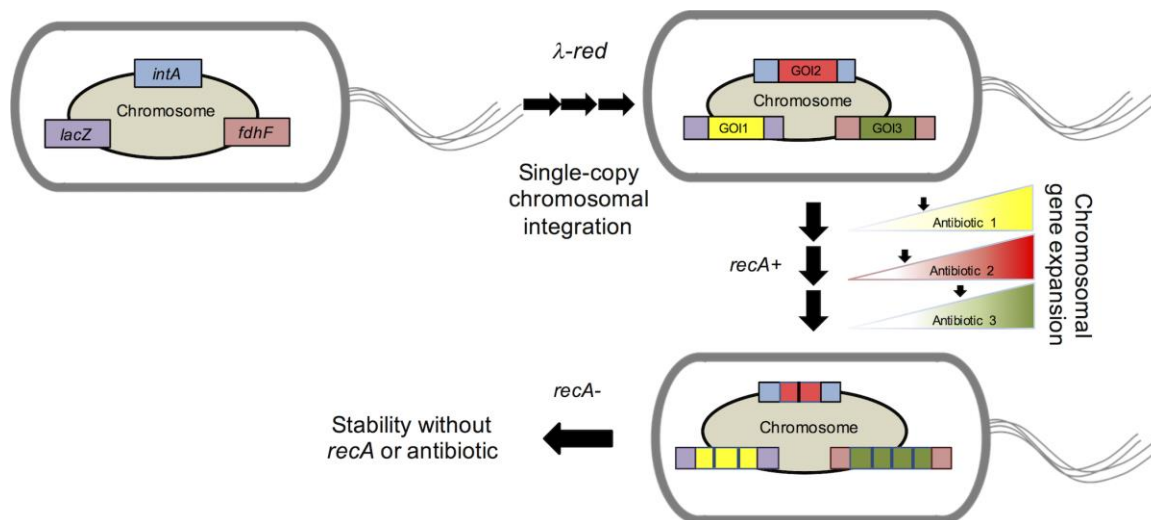
Figure 3.1. Parallel Integration and Chromosomal Expansion (PIACE) schematic.

Heat-curable helper plasmid pRedi2RecA (Supplemental Figure 6.1) is introduced to provide rhamnose-inducible *recA* homologous recombination functionality. Under *recA* expression conditions, the copy numbers of gene cassettes GOI1, GOI2, and GOI3 can be orthogonally and tunably expanded (via the previously established CIChE mechanism) by sequential antibiotic (gentamicin, chloramphenicol, and spectinomycin, respectively) titration. pKD46 and pRedi2RecA plasmids are heat cured post expansion to promote chromosomal gene cassette copy number stability in the absence of continued antibiotic selection pressure.

3.4.2 Proof of Concept Chromosomal Integration and Expansion of *rfp* and *gfp*

For PIACE proof of concept, *rfp* and *gfp* were selected as genes of interest (for slots GOI2 and GOI3 in Figure 1). As per the PIACE strategy, suicide vectors pPIACE_Cm_RFP and pPIACE_Sp_GFP (Figure 3.1) were sequentially integrated at *intA* and *fdhF* loci, respectively, and integration at these loci was confirmed by colony PCR. This resulting strain (1R1G_pRedi2RecA_DH1, Supplemental Table 6.1), containing one copy each of *rfp* and *gfp*, was then exposed (in the presence of *recA*) to stepwise parallel increases in chloramphenicol and spectinomycin titers to an 8×8 matrix of final antibiotic concentrations.

3.4.2.1 Fluorescence measurement of RFP and GFP

For each of these 64 resulting strains (1R1G−8R8G), Figures 3.2 A and B show the resulting RFP and GFP relative fluorescence units (microplate reader), and Figures 3.2 C and D show the *rfp* and *gfp* copy numbers (qPCR). For the most part, RFP and GFP relative fluorescence units along with *rfp* and *gfp* copy

numbers increase monotonically (and orthogonally) with chloramphenicol and spectinomycin titers, respectively.
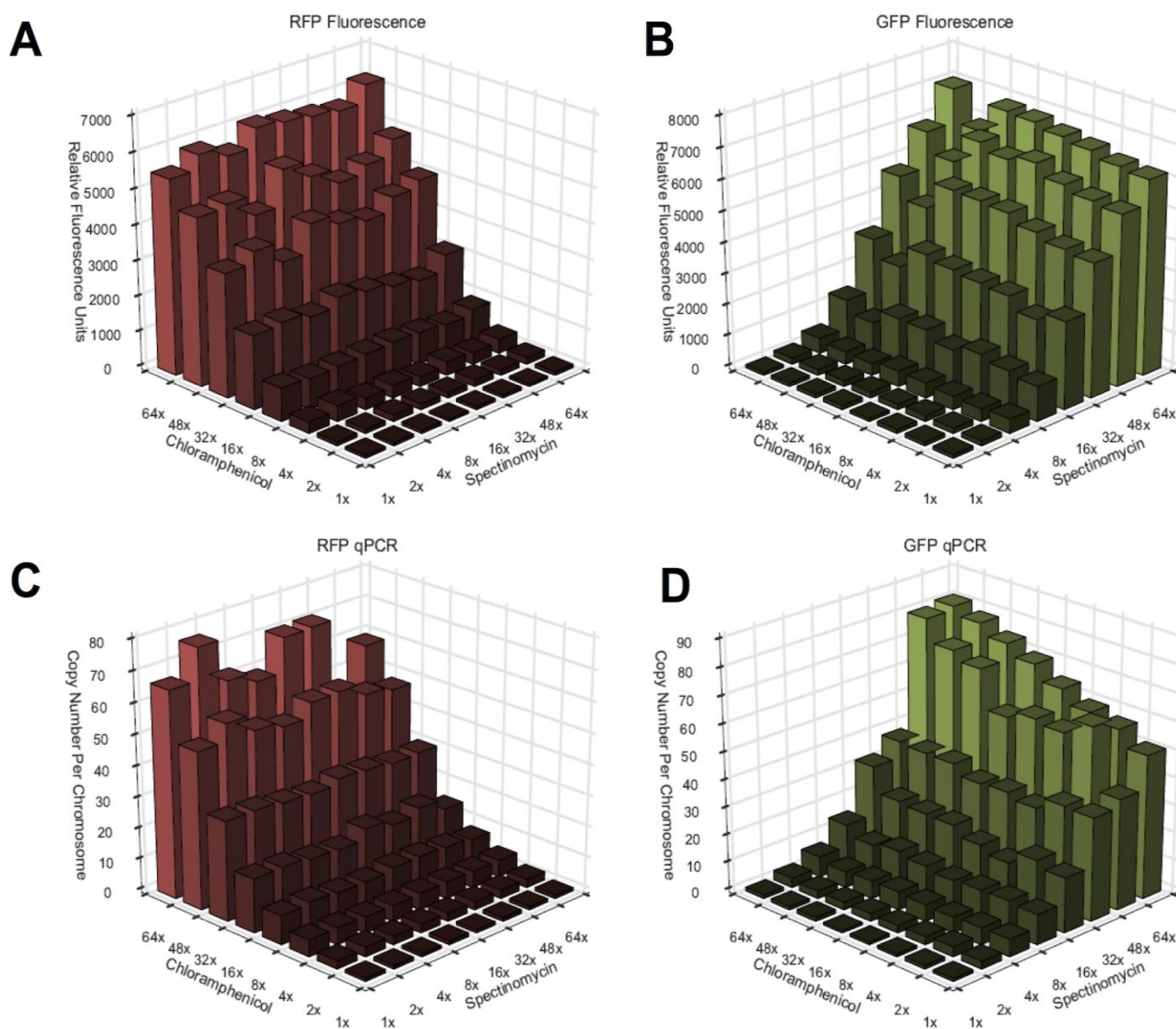

Figure 3.2. Pilot experiment expanding *rfp* (CmR) and *gfp* (SpecR) gene cassettes at two loci (*intA*, *fdhF*). An 8 x 8 matrix of chloramphenicol and spectinomycin titration conditions result in different extents of *rfp* and *gfp* expansion, which were quantitatively assessed by fluorescence and qPCR

measurements. (**A**) RFP fluorescence. (**B**) GFP fluorescence. (**C**) *rfp* qPCR. (**D**) *gfp* qPCR. Data points are the average of two biological replicates.

Up to 66 and 80 copies of *rfp* and *gfp*, respectively, were obtained in strain 8R8G. Two individual colonies were picked post expansion for each expansion condition, and the presented data in Figure 3.2 consist of the average (of the two colonies) values for fluorescence and copy numbers. A given colony's RFP or GFP copy number varied on average by roughly 15% or less from the average value, about 4−5 copies. A given colony's fluorescence values varied on average by roughly 5% from the average value, about a 140 (RFP) or 110 (GFP) relative units difference between the two colonies.

3.4.2.2 Stability of *rfp* and *gfp* Copy Numbers Post Expansion

We then assayed the stability of a subset of these *rfp* and *gfp* gene expansions post *recA* removal (via heat curing pRedi2RecA, as in Figure 3.3) in the absence of antibiotics. We began with 10 mL starting cultures (with antibiotics) of three strains (low-copy *rfp/gfp* (1R1G−1X/ 1X), medium-copy *rfp/gfp* (4R4G−8X/8X), and high-copy *rfp/gfp* (8R8G−64X/64X)) and then performed three cycles of 1000-fold (serial) dilution into fresh LB media (lacking antibiotics) followed by growth overnight at 37 °C. For each of the three initial strains, RFP and GFP relative fluorescence units along with *rfp* and *gfp* copy numbers were measured for the starting cultures and after each cycle of dilution and outgrowth (Figure 3.3). Even the least stable (highest *rfp/gfp* copy) initial strain (8R8G−64X/64X) retained more than 71 and 75% of its original copies of *rfp* and *gfp*, respectively, despite a billion-fold (volumetric) outgrowth in the absence of antibiotics.
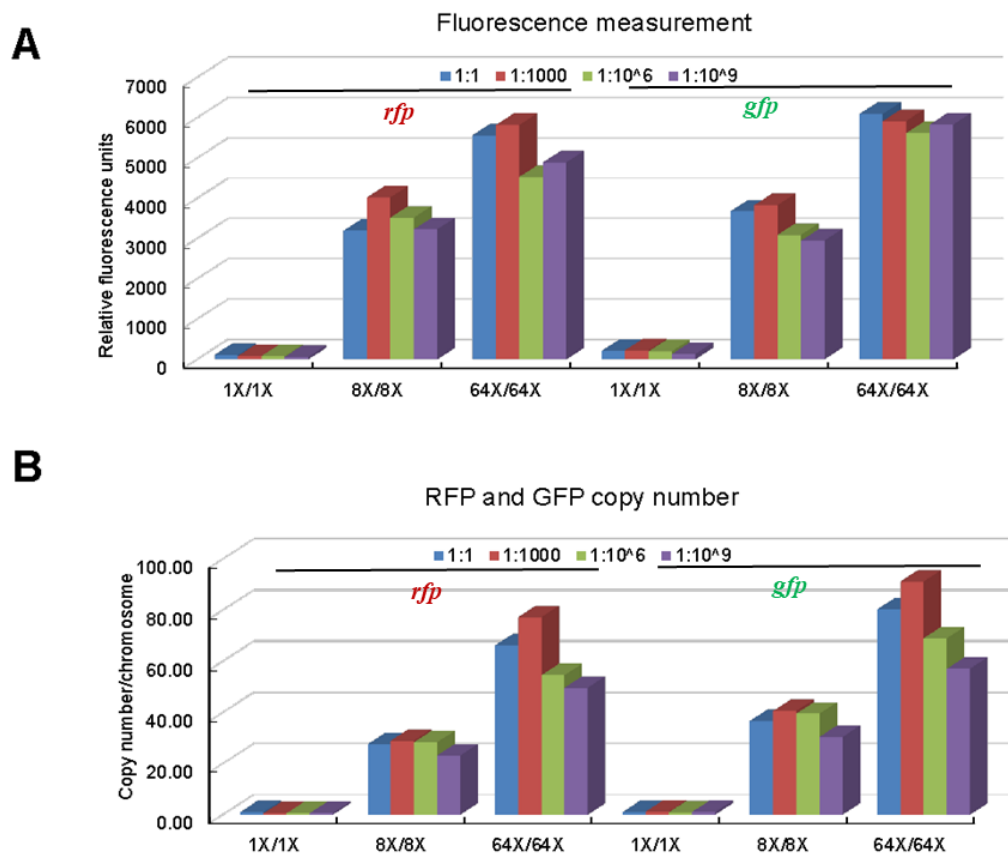
Figure 3.3. Stability of *rfp* and *gfp* expansions post *recA* removal and culture propagation in the absence of antibiotic selection pressure. 10 mL cultures of three starting strains (1X/1X, 8X/8X, 64X/64X) were repetitively (three cycles total) diluted 1000-fold in fresh LB media (lacking antibiotics) and grown overnight at 37°C. *rfp* and *gfp* copy numbers were quantitatively assessed by (**A**) fluorescence and (**B**) qPCR measurements. Data points are the average of two biological replicates.

3.4.3 Isopentenol Pathway Integration and Expansion

Following the *rfp/gfp* proof of concept experiments, we applied PIACE to the isopentenol metabolic pathway. Isopentenol pathway was divided into three

segments (Figure 3A), namely the enzymatic steps carried out by AtoB, MvaS, and MvaA (GOI1); MK and PMK (GOI2); and PMD and NudB (GOI3). We compared isopentenol titers and gene copy numbers for three isopentenol pathway expression configurations (Figure 3.4): (1) a two-plasmid system, (2) a hybrid plasmid/chromosomal system, and (3) a chromosomal system.

3.4.3.1 Two-plasmid system

The two-plasmid system is a previously published strain (JBEI-7003) [150] that contains two plasmids, namely (1) pMevTsa-PMKco-MKco, which has a p15A origin and contains GOI1 and GOI2 under the inducible expression control of a lacUV5 promoter, and (2) pTrc99A-NudB-PMD, which has a pBR322 origin and contains GOI3 under the inducible expression control of a pTrc promoter.

3.4.3.2 Hybrid plasmid/chromosomal system

The foundation of the hybrid plasmid/chromosomal system is an *E. coli* DH1 strain (pMevTop_Kan_DH1) that contains plasmid pMev_AtoB-mvaS-mvaA (Figure 6.1), which has a p15A origin and contains GOI1 under the control of a lacUV5 promoter. Into this strain, pMevTop_- Kan_DH1, GOI2, and GOI3 are sequentially integrated into the *intA* and *fdhF* chromosomal loci (via suicide vectors pPIACE_Cm_PMK-MK and pPIACE_Sp_NudB-PMD, Figure 6.1), respectively yielding (post heat-curing pKD46 and introducing pRedi2RecA) the hybrid plasmid/chromosomal system base strain 1p1N_pRedi2RecA_DH1, which is then competent (under *recA* expression) for orthogonal GOI2 and GOI3 copy number expansions. It could be noted that we could have alternatively selected GOI2 or GOI3 to retain in the plasmid instead of GOI1. In a manner analogous to that described above for a matrix of *rfp*/*gfp* expansions, we sequentially expanded GOI2 and then GOI3 in a 7×7 matrix
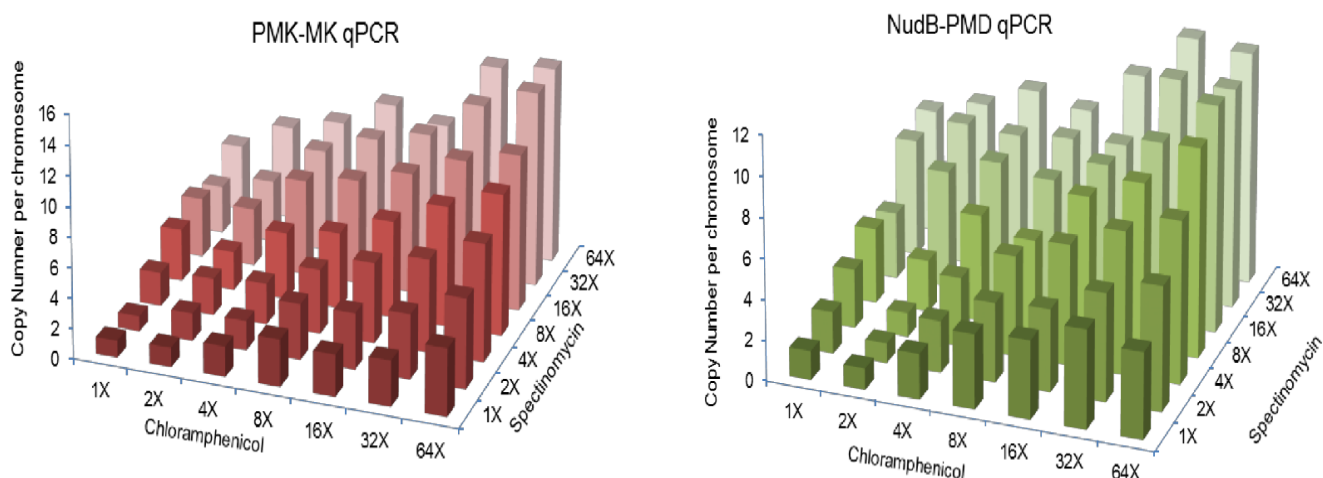
of chloramphenicol and spectinomycin titrations. These resulting 49 strains were then heat-cured of pRedi2RecA (resulting in strains 1P1N through 7P7N).
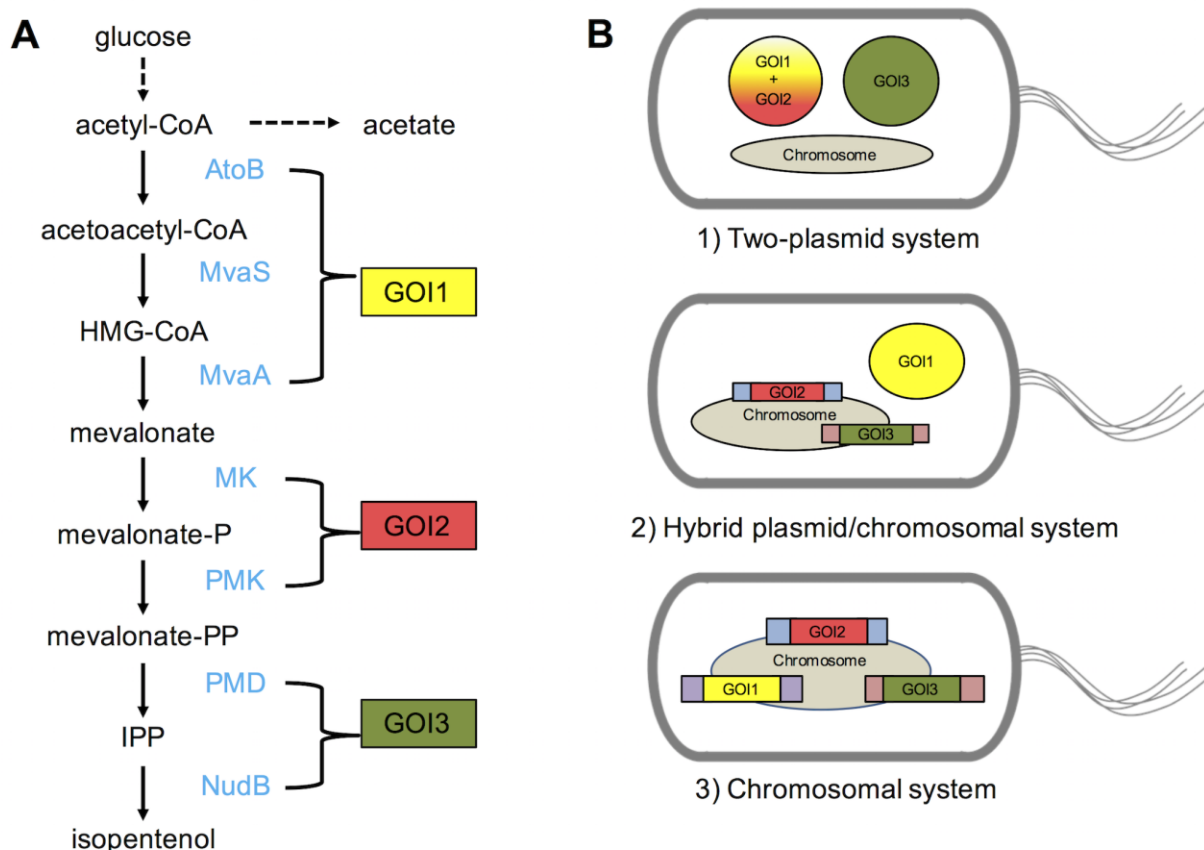
Figure 3.4. Schematic of three configurations evaluated for expressing a metabolic pathway to isopentenol. (**A**) An implementation of a glucose to isopentenol pathway that includes the enzymes AtoB, MvaS, MvaA (herein collectively referred to GOI1); MK, PMK (GOI2); and PMD, NudB (GOI3). (**B**) Schematic of three expression configurations: 1) two-plasmid system, with GOI1+GO2 and GOI3 split between p15A- and pBR322- derivative medium-copy plasmids, respectively; 2) hybrid plasmid/chromosomal integration system, with GOI1 residing in a p15A-derivative medium copy plasmid, and GOI2 and GOI3 integrated into the *intA* and *fdhF* loci, respectively; and 3) chromosomal integration system, with GOI1, GOI2, and GOI3 integrated into the *intA*, *fdhF*, and *lacZ* loci, respectively.

3.4.3.3 Chromosomal system

The chromosomal system is *E. coli* DH1, into which GOI1, GOI2, and GOI3 are sequentially integrated into the *lacZ*, *intA*, and *fdhF* chromosomal loci (via suicide vectors pPIACE_- Gent_AtoB-mvaS-mvaA, pPIACE_Cm_PMK-MK, and pPIACE_Sp_NudB-PMD; Figure 6.1), respectively yielding (post heat-curing pKD46 and introducing pRedi2RecA) the chromosomal system base strain 1A1P1N_pRedi2RecA_DH1, which is then competent (under *recA* expression) for orthogonal GOI1, GOI2, and GOI3 copy number expansions. In a manner similar to that described immediately above for the hybrid plasmid/ chromosomal system, we sequentially expanded GOI1, and then GOI2, and finally GOI3 in a 5 × 5 × 5 matrix of gentamicin, chloramphenicol, and spectinomycin titrations. A subset of 15 of these 125 total strains was selected for further characterization (post heat-curing pRedi2RecA, resulting in strains 1A1P1N through 5A5P5N).

Figure 3.5. For the hybrid plasmid/chromosomal configuration, qPCR measurements of *PMK-MK* and *NudB-PMD* gene copy numbers for a 7 x 7

matrix of spectinomycin and chloramphenicol titration conditions. Data points are the average of two biological replicates.

## 3.4.4 Isopentenol titers and genes of interest copy number

The isopentenol titers were examined by the three pathway expression configurations. They produced different isopentenol titers (Figure 3.6) and consisted of different copy numbers of GOI1, GOI2, and GOI3 (Figure 3.6B). All of them did not appear to differ in their cellular growth phenotypes (all achieving similar ODs at the stationary phase time point in which isopentenol production was assessed). The maximum isopentenol titers achieved for the three expression configurations were: (1) two-plasmid system (strain JBEI-7003), 708 mg/L; (2) hybrid plasmid/chromosomal system, 344 mg/L (strain 7P7N); and (3) chromosomal system, 105 mg/L (strain 5A5P5N). The maximum copy numbers we achieved for GOI1, GOI2, and GOI3 across the three different configurations were: (1) two-plasmid system: 15 GOI1, 15 GOI2, and 20 GOI3 (strain JBEI-7003) [150]; (2) hybrid plasmid/chromosomal system: 15 GOI1, 13 GOI2, and 12 GOI3 (strain 7P7N) (Figure 3.5); and (3) chromosomal system: 5 GOI1, 8 GOI2, and 8 GOI3 (strain 5A5P5N). While error bars (standard deviation of three replicate colonies) are presented in Figure 3.6B, for the hybrid plasmid/chromosomal systems (except for strain 7P7N, all those strains for which data are shown in Figures 3.6A), two individual colonies were picked post expansion for each expansion condition, and the presented data consist of the average (of the two colonies) values for copy number and isopentenol production. In terms of copy number variation for the hybrid plasmid/chromosomal systems, a given colony's GOI1−3 copy numbers varied on average by roughly 15% or less from the average value, about 1−2 copy difference between the two colonies for GOI1−3. For the hybrid plasmid/chromosomal system, a given colony's isopentenol production
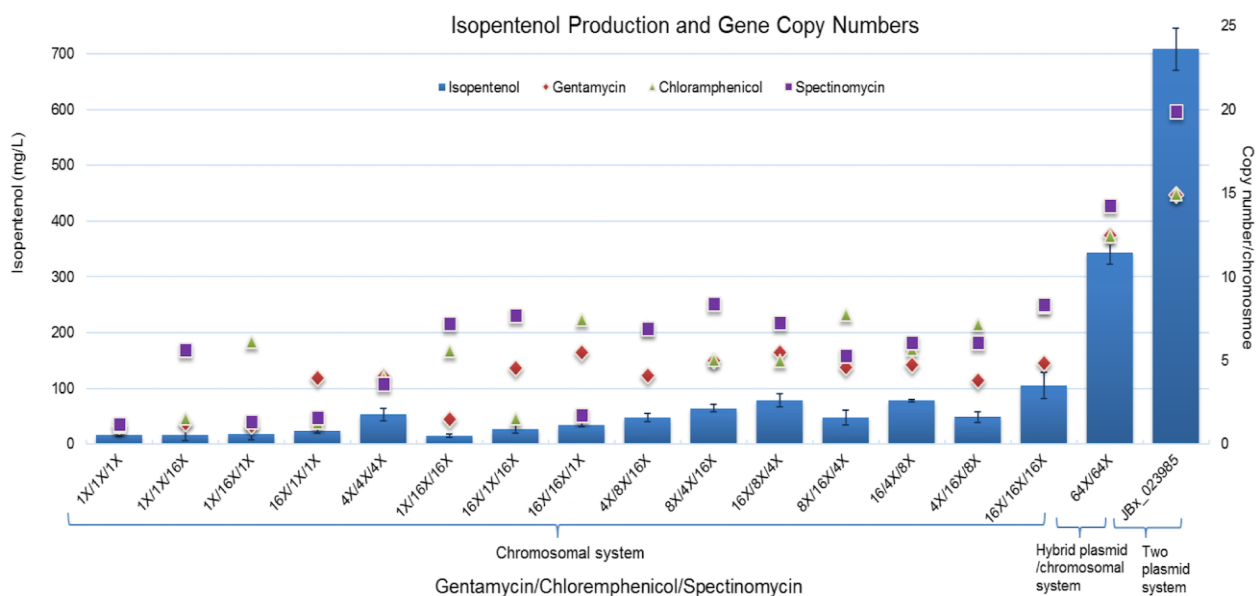
comparably varied on average by roughly 15% from the average value, about a 30 mg/L difference between the two colonies.
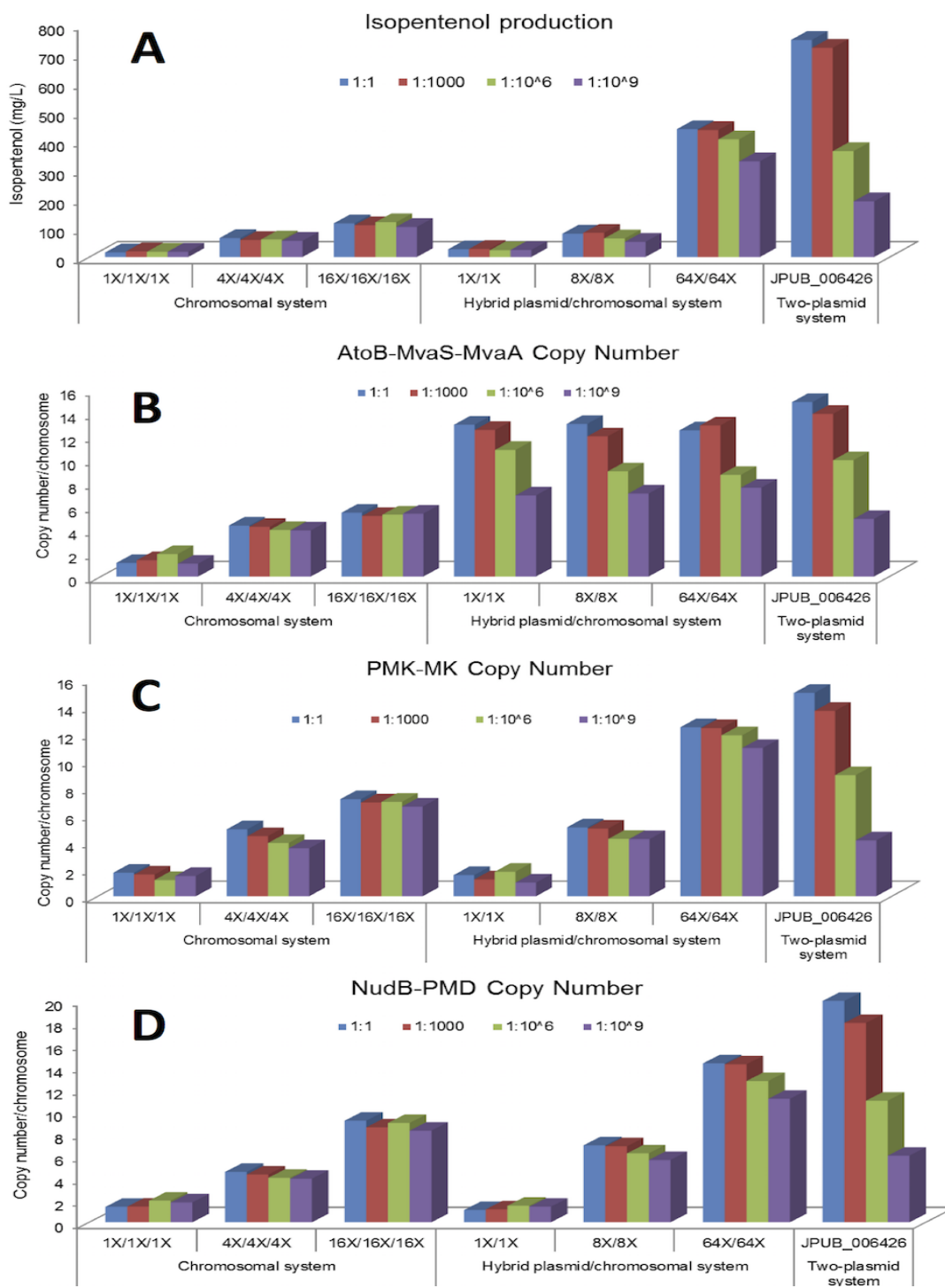
A
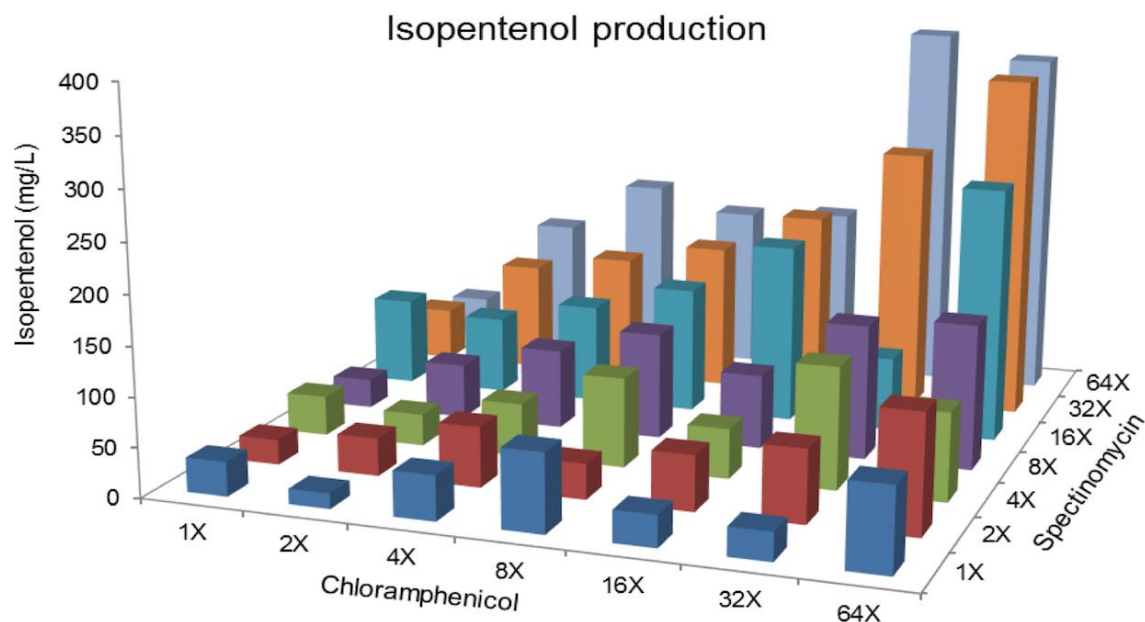
B

Figure 3.6. Isopentenol production and gene copy numbers for the three expression configurations. (**A**) For the hybrid plasmid/chromosomal configuration, isopentenol production for a 7 x 7 matrix of chloramphenicol and spectinomycin titration conditions. Data points are average of two biological replicates (**B**) For the chromosomal configuration, 15 strains subsampled from a 5 x 5 x 5 matrix of gentamicin, chloramphenicol, and spectinomycin titration conditions compared with the 64X/64X hybrid plasmid/chromosomal and two-plasmid configurations. Copy numbers of GOI1 (AtoB, MvaS, MvaA; GentR), GOI2 (PMK, MK; CmR), and GOI3 (PMD, NudB; SpecR) were assessed by qPCR. Error bars indicate the standard deviations of three biological replicates.

3.4.5 Stability of Isopentenol Titers and GOI1−3 Copy Numbers across Expression Configurations

A stability assay was performed to test the stability of isopentenol titers and GOI1−3 copy numbers across the two-plasmid (strain JBEI-7003), hybrid plasmid/chromosomal (strains 1P1N, 4P4N, and 7P7N), and chromosomal (strains 1A1P1N, 3A3P3N, and 5A5P5N) pathway expression configurations. Analogous to that performed above to assess the stability of *rfp/gfp* strains, these isopentenol-producing strains were repetitively diluted in fresh LB media without antibiotics 1000-fold and grown overnight for three cycles total. The resulting progression of isopentenol titers and GOI1−3 copy numbers are shown in Figure 3.7. The relative isopentenol titer and GOI1−3 copy number stability of the chromosomal expression configuration outperforms the hybrid plasmid/ chromosomal and two-plasmid systems, although the starting isopentenol titer and GOI1−3 copy numbers are lowest for the chromosomal system.

Figure 3.7. Comparison of isopentenol production and gene copy number stability across three variations of the chromosomal configuration, three variations of the hybrid plasmid/chromosomal configuration, and the two-plasmid configuration. 10 mL cultures were repetitively (three cycles total) diluted 1000-fold in fresh LB media (lacking antibiotics) and grown overnight at 37 ˚C. (**A**) Isopentenol production. (**B**) GOI1 (AtoB, MvaS, MvaA), (**C**) GOI2 (PMK, MK), and (**D**) GOI3 (PMD, NudB) copy numbers, as assessed by qPCR. Data points are average of two biological replicates.

3.5 Discussion

PIACE (Figure 3.1) was developed to overcome several of CIChE's limitations, namely (1) low integration efficiencies for multigene cassettes, (2) that only a single locus per strain was used for gene integration, and (3) the requirement to delete *recA* post gene cassette copy number expansion. To increase the integration efficiency of multigene cassettes of interest rather than introduce a gene cassette through the transformation of linear DNA, we opted to use supercoiled suicide plasmid vectors. These suicide vectors contain a R6K replicon, which can only replicate in strains expressing the π protein [171, 172] (such as *E. coli* PIR1 but not *E. coli* DH1). This approach significantly increased integration efficiencies and our ability to integrate multi-gene cassettes (data not shown). While the original Tyo et al. CIChE publication [111] does speculate that other integration methods would also suffice, our comparison here was made with respect to the integration method actually used in Tyo et al. Three different loci (*lacZ*, *intA*, and *fdhF*) in *E. coli* DH1 were selected for integrating GOI1, GOI2, and GOI3, respectively, all into the same strain. Dividing a metabolic pathway across multiple loci enables each portion (e.g., GOI1, GOI2, and GOI3) to have its copy number expanded independently of the others, providing more fine-grained control for optimizing pathway performance, as compared with integration at a single locus, in which the copy

numbers of all portions of the pathway must be coarsely expanded together in lockstep.

Rather than starting with a *recA* positive strain of *E. coli* and having to delete *recA* post gene expansion, we opted to start with a *recA* negative strain (*E. coli* DH1) and transiently induce *recA* expression from a heat-sensitive plasmid (pSC101 origin of replication) during the expansion process and then heat-cure this helper plasmid post gene expansion. This methodological distinction from CIChE is especially important when generating libraries of copy number variants because introducing the helper plasmid (and for *recA* positive *E. coli* backgrounds, deleting *recA*) once to the base strain prior to expansion and then heat-curing the plasmid from each derivative strain variant post expansion is far faster and requires less effort than deleting *recA* from each individual expanded strain variant. Phage transduction could possibly be an equally quick and easy method for removing *recA* (from the genome) post CIChE expansion, although this would necessitate the use of an additional selectable marker with which to indicate and select for positive transductants of the *recA* deletion.

The PIACE method continues to offer the best features of CIChE, such as enhanced pathway stability (even in the absence of selection pressure) and titratable pathway copy numbers. As reported in the CIChE publication [111] and supported here in Figure 3.6, chromosomally integrated pathways, in comparison with plasmid-borne metabolic pathways (especially for those pathways invoking high metabolic demands in the absence of antibiotics), retain pathway copy numbers better, which results in more sustained levels of target metabolite production over time. While it may be easier to initially obtain high gene copy numbers through the use of multicopy plasmid systems, it has been shown that a six-copy chromosomal gene integration produced more L-serine than the same gene harbored in medium copy plasmid [170]. Besides copy number stability, the ordered segregation (following DNA replication) of

multicopy chromosomal integrations makes it more difficult for pathway-attenuating mutations to take over a cellular population, compared with the random segregation that is characteristic of some multicopy plasmid systems.

These stability advantages of chromosomally integrated vs plasmid-borne pathways are of key importance in the context of industrial metabolite production, in which cost (and regulatory) drivers lead to large-scale (many cellular generations) and antibiotic-free fermentations. Beyond stability, the ability to fine-tune copy number (through selective pressure titration) offers advantages over some other gene integration methods, such as CIGMC [170], that result in variable (colony-to-colony) copy numbers. While methods like PIACE and CIChE require more steps than CIGMC to integrate multiple gene copies into the chromosome, they do not require additional molecular biology (beyond the creation of the base strain) or colony screening to generate strain variant libraries with monotonically increasing gene copy numbers.

There are at least three major opportunities for further improving PIACE, namely (1) detune selectable marker expression to boost maximal achievable copy numbers, (2) develop alternative selectable markers, and (3) use a recombinase to remove superfluous (e.g., vector backbone components) sequences prior to the expansion process. As shown in Figure 6.1, statistical modeling suggests that increasing GOI1−3 copy numbers would result in higher isopentenol titers (despite the potential concern that the coupled expansion of superfluous vector backbone component sequences would adversely affect performance). (As an aside, we expect that, as higher isopentenol titers are achieved, the interplay of GOI1−3 copy numbers with host metabolism will become more complex, and the relationship will still be accessible through this statistical technique. As shown in Figure 3.6, however, we were unable to achieve more than 10− 15 copies of GOI1−3. This is in contrast with *gfp/rfp* that could be expanded to about 80 copies (Figure 3.2).

Because there are roughly the same maximal copy numbers of each GOI (15 GOI1, 12 GOI2, and 13 GOI3) in the maximally expanded hybrid expression configuration strain 7P7N, this places some doubt that having selected GOI2 or GOI3 rather than GOI1 as the plasmid borne GOI would have enabled access to a significantly more diverse space of gene copy numbers. To attain higher copy numbers for larger gene cassettes (e.g., GOI1−3), we speculate that it may be necessary/possible to detune (i.e., reduce the strength of) the promoters/RBSs governing the expression of the selectable marker (e.g., the antibiotic resistance gene) to decrease the effective amount of selectable activity per copy and thus increase the total number of gene copies required to withstand a given amount of selective pressure. Second, our PIACE developments to date have exclusively used antibiotic resistance genes for selectable markers. It would be of great utility to develop alternative titratable selectable markers, such as auxotrophic markers, or perhaps even screenable fluorescent markers, so that the resulting strains do not bear antibiotic resistance genes, which may prove problematic in industrial contexts from an economic or regulatory standpoint.

 It would additionally be possible to use fabI/triclosan as one of the selectable marker/pressure pairs [173]; however, Because PIACE requires multiple orthogonal markers, using *fabI* would only partially address the replacement of antibiotic resistance markers. We speculate that selectively eliminating antibiotic resistance genes post expansion (e.g., flipping them out via a recombinase) could be possible but that this would need to be done in such a manner as not to eliminate/regress the expanded gene copies of the pathway of interest at the same time.

3.6 Conclusion and outlook

In this study, we developed an approach for the PIACE of metabolic pathways (Figure 1), which extends the CIChE method in several keyways. Following proof-of concept experiments with *gfp*/*rfp* (Figure 3.2), we applied PIACE to the integration of genes encoding the mevalonate metabolic pathway to isopentenol, divided across three gene cassettes, into three loci within the *E. coli* DH1 chromosome (Figure 3.4). Following gene cassette copy number expansion (under *recA* expression and tuned antibiotic concentration conditions), we achieved a three-dimensional grid sampling of GOI1, GOI2, and GOI3 cassette copy number variants, assessed these variants for isopentenol production (Figure 3.6), and applied statistical modeling to make predictions as to which additional gene cassette copy number variants should be pursued in the future to further improve isopentenol titers (Supplemental Figure 6.7). Finally, we assessed the stabilities of isopentenol production and gene cassette copy numbers in the absence of antibiotic pressure, for the fully chromosomal configuration in comparison with hybrid chromosomal/plasmid and two plasmid configurations (Figure 3.7).

Beyond the expression of *gfp/rfp* or the isopentenol metabolic pathway, PIACE could be applied to other metabolic or regulatory pathways of interest and in organisms besides *E. coli*. Whereas applying PIACE to another pathway is straightforward (e.g., swap out GOI1−3 contents), implementing PIACE in another organism may prove challenging. An important PIACE prerequisite (shared in common with CIChE) is switchable (e.g., conditionally inducible) homologous recombination activity, given the reversible nature of its mechanism for expanding gene copy number. Recombinase activity should be conditionally switched on to enable selectable gene duplication/expansion but should then (post expansion) be switched off to prevent regressive copy number

elimination/reduction (at least in the absence of selective pressure). It would be a technical challenge, then, to implement PIACE in organisms without endogenous (or heterologously expressed) recombinase activity or (in the absence of selective pressure) in organisms with chronic endogenous recombinase activity that cannot be switched off. The latter issue may not pose too great an obstacle in the case of persistent auxotrophic selection pressure, which could be cost- and regulatory-effective (in contrast with antibiotic selection pressure) even at industrial scales. We look forward to the application of PIACE (or an improved derivative thereof) to other pathways and to its implementation in organisms beyond *E. coli*.

Chapter 4

Engineering of Fatty Acid Biosynthesis for Inducible and orthogonal pathway expression in *E. coli*

4.1 Abstract

Engineering heterologous fatty acid metabolic pathways (e.g., for the production of biofuels and other bioproducts) in *E. coli* (and in other organisms) is complicated and made much more difficult by concurrent and highly regulated endogenous fatty acid metabolism (and catabolism). Since fatty acid metabolism is essential for cellular growth, and the genes responsible for it are distributed across the genome, it is not possible to simply remove it wholesale; nor is it straightforward how to repress it in a controllable fashion. Our proposed approach is to remove the genes responsible for endogenous fatty acid metabolism from the genome via CRISPR Cas9-like approaches and place them instead on a replicating artificial chromosome that can be controllably degraded (via CRISPR Cas9). The heterologous fatty acid pathway (either plasmid-borne or chromosomally integrated) could then be induced orthogonal to and independent from the endogenous metabolism.

4.2 Introduction

The fatty acid synthesis pathway, which is a vital component of the primary metabolism in bacteria, is severely regulated [67, 69]. Although this pathway's engineering leads to the production of many valuable chemicals and fuels such as fatty alcohols, methyl ketones, and fatty acids, there are many drawbacks to overcome. First, deregulation of the fatty acid biosynthesis pathway in *E. coli* is necessary for the industry-level production of these desirable chemicals. Second, decoupling the pathway from cell growth and self-regulation, and last, deviating the consumption of fatty acids for other molecules is crucial to

improving yields of the commodity chemicals [174, 175]. This deregulation is tightly tied with cell growth, and due to this reason, product yield improvement efforts have been limited to engineering the heterologous fatty acid pathways. FadR is known to act as a repressor of beta-oxidation genes and activator of the *fabA* and *fabB* genes in the carbon chain elongation [176]. On the other hand, FabR acts as a repressor for *fabA* and *fabB* genes of the biosynthesis pathway. This results in changes in the membrane's physical properties and the production level of unsaturated fatty acids. Accumulation of long-chain acyl-ACP acts as inhibitory to several genes involved in the fatty acid biosynthesis, limiting the flux through the fatty acid biosynthesis pathway [76, 77, 175]. It is still unclear about the severity of the detrimental effects pathway regulatory genes may exhibit upon over or under expression.

End products of type II FAS in prokaryotes are acyl-ACP and acyl-CoA, which are precursors for various fatty alcohols, free fatty acids, and methyl ketones. The β-oxidation pathway can degrade these precursors, but deletion of the *fadE* (encoding for acyl-CoA oxidase) gene prevents the loss of the precursors [6, 177]. Four key enzymes play an important role in converting the acyl-CoAs and acyl-ACPs to fatty alcohols. (1) Thioesterase cleaves the thioester bond linking chain to ACP and releases free fatty acids (FFAs). (2) Acyl-CoA synthase is converting free fatty acids to acyl-CoA. (3) Acyl-CoA or acyl-ACP reductases are reducing the acyl-ACP and acyl-CoA to fatty alcohols. (4) Aldehyde reductase is reducing the fatty aldehyde to fatty alcohols [178, 179]. Optimization of these enzymes can achieve significant titer improvements in the production of fatty alcohols [84, 180, 181]. There have been numerous reports of engineered heterologous pathways comprising these enzymes from different species, but the native fatty acid biosynthesis regulation has rarely been addressed.
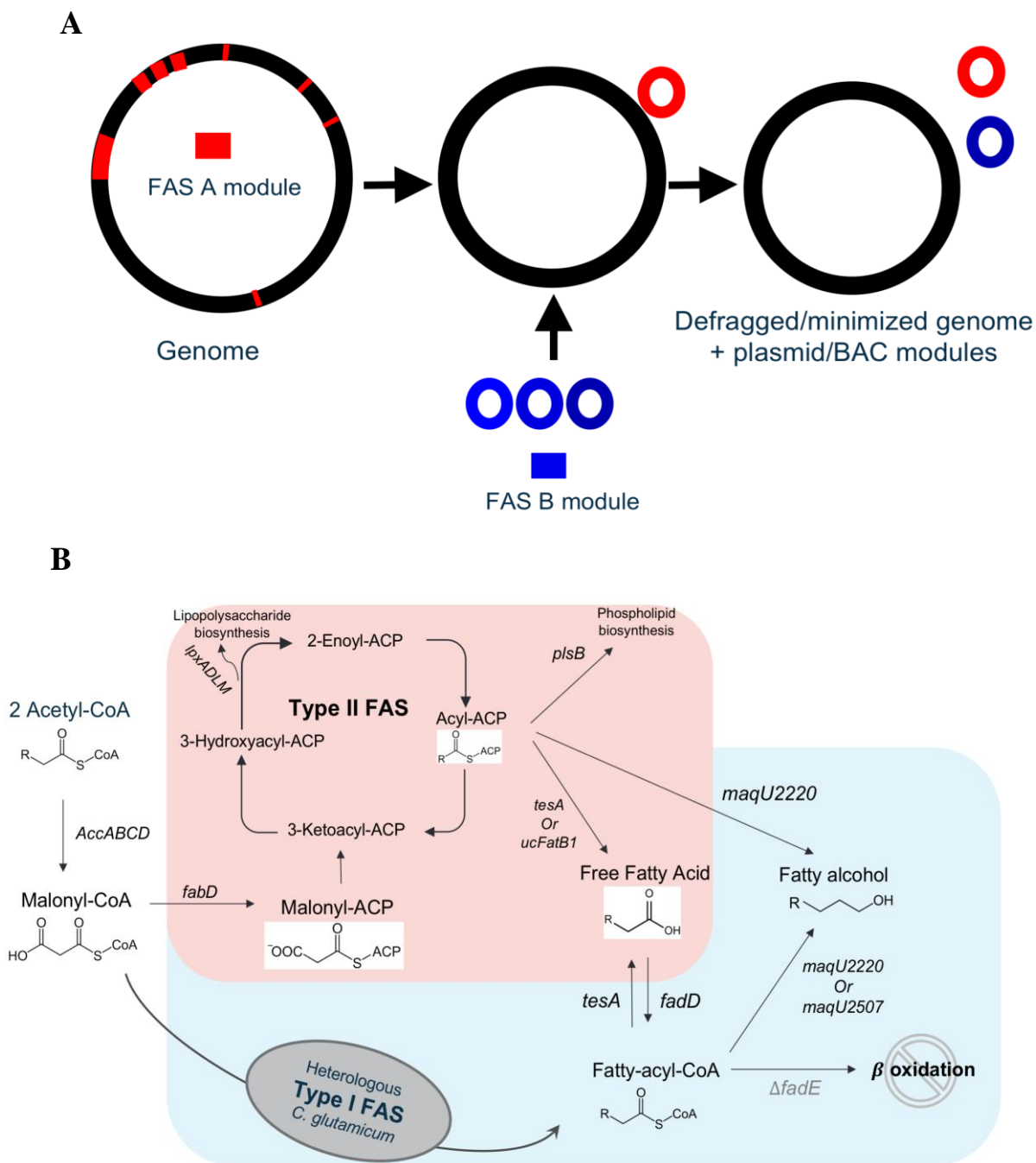
Figure 4.1. (A) Fatty Acid Biosynthesis Inducible and Orthogonal overview. Sporadically distributed native fatty acid biosynthesis pathway genes are

removed to create a defragged genome in *E. coli*. Two independently regulated and orthogonally controllable fatty acid biosynthesis pathways are introduced in the form of FAS-A (red circle) and FAS-B (blue circle) – FAS-A to carry out native functions essential for cell growth and FAS-B to express heterologous pathway and catalyze production of the desired biofuel precursor. **(B)** FAS-A (red) represents the endogenous fatty acid pathway producing acyl-ACP (feedback inhibits the native type II FAS), and phospholipids and lipopolysaccharide metabolites (other biomolecules). FAS-B (blue) represents the heterologous pathway with type I FAS from *C. glutamicum* which produces acyl-CoAs. CoA-ACP reducatase Maqu2220 reduces CoA intermediates directly to fatty alcohols.

A deeper understanding of the fatty acid biosynthesis regulation is critical in engineering the fatty acid biosynthesis metabolic pathway dedicated towards optimized fatty alcohol production. Native fatty acid pathway genes are distributed throughout the genome, making it very difficult to repress them or remove them altogether. Here we report a fatty acid biosynthesis orthogonal and independent (FABIO) system. We remove the genes responsible for the endogenous fatty acid metabolism from the genome via CRISPR Cas9 and place them on a replicating bacterial artificial chromosome (BAC-A) or FAS-A that can be controlled for degradation. Then, we introduce a heterologous pathway in the form of an inducible plasmid (FAS-B) targeted towards yielding high titers of acyl-CoAs and eventually fatty alcohols. FAS-A and FAS-B modules are controllable in a manner that they are orthogonal and independent to each other (Figure 4.1).

Genes that were selected for placing on the FAS-A module and removal from the genome are growth-essential genes that belonged mainly to the type II FAS for fatty acid biosynthesis. Our strategy includes creating a defragged genome free from the regulatory proteins, lipopolysaccharide biosynthesis proteins,

phospholipid biosynthesis proteins, and thioesterases to comprehend and control the fatty acid regulation and improve the fatty alcohol titers. These growth-essential genes are complemented by the FAS-A module (red), and the heterologous pathway (green) in the form of the FAS-B module is designed to produce fatty alcohols as shown in Figure 4.1. Properties of the products derived from the fatty acid precursors are dependent on the chain length, unsaturation, and branching. Native fatty acids typically maximize at C16 and C18 in *E. coli*. It is reported that thioesterase UcFatB1 from *Umbellularia californica* has shown a preference towards C-12 free fatty acids from acyl-ACPs. In the FAS-B module, thioesterases and reductases can be selected specific to the product and carbon chain length of interest.

A resulting defragged genome hosts two modules FAS-A and FAS-B, allowing them to perform their functions independently, free from native regulation and without contributing their metabolic fluxes towards byproducts such as phospholipids and lipopolysaccharides. Beyond understanding the fatty acid regulation, this study produced findings confirming that repression of the FAS-A system forces glucose conversion into desired products and showed significant potential towards improving fatty alcohol titers. Another clear demonstration is that a highly efficient FAS-B system is needed to convert the acyl-CoAs into fatty alcohols entirely. Finally, engineered FABIO systems have potential to produce different carbon chain-length fatty acid pathway derived precursors by simply plugging in the chain-length specific heterologous thioesterases and reductases.

4.3 Materials and methods

All the commonly used materials used in this study are described in section 2.1.

4.3.1 Plasmids and strains construction

Cloning methods for constructing plasmids and strains used in this study are described in section 2.3. *E. coli* DH10B (Supplemental Table 6.5) was used as the cloning host for the plasmids pKD46Cas9_sgRNA(p15A) and FAS-B construct. *E. coli* PIR1 was used as the cloning host for all the suicide vectors constructed for the fixing templates (Supplemental Table 6.3). *E. coli* MG1655 Δ*fadE* (JBEI 3111) was used as the starting strain for all the modifications. List of all plasmids and strains used in chapter is provided in Table 4.5. DNA oligos and templates, PCR conditions, and the DNA assembly method type used to construct each plasmid are listed in Table 6.4. DNA fragments were amplified following PCR composition and touchdown thermocycling conditions described in Tables 2.3.1 and 2.3.2. Following PCR amplification, residual (methylated) DNA template in each PCR reaction was *DpnI*-digested and purified using a Qiagen gel extraction kit or Nimbus Size Selection, as specified. Gibson assembly methods were used for constructing the plasmids as described in table 2.6.1. All suicide vectors have a R6K origin of replication, which can be propagated only in a PIR1 strain background, as well as a 1 kb fixing templates with 500 homology arms targeting their respective sites for chromosomal integration. Supplemental Table 6.3 provides a list of 5' and 3' fixing template sequences.

4.3.2 CRISPR-Cas9 mediated genome editing and strain development

CRISPR-Cas9 approach is described in detail in section 6.5.1. Fatty acid biosynthesis type II genes listed in Table 4.1 are grouped into 9 gene groups and clusters based on their size and location in the genome (Table 4.4). synCRISPRs were designed as described in Alonso-Gutierezz et al[182]. synCRISPRs (Supplemental Table 6.2) in the pMCC vector were ordered from Genscript. pMCC_synCRISPR expresses the guide RNA to target specific sequences in the *E. coli* genome to be cut by Cas9. Suicide vectors containing the fixing templates were constructed using the Gibson assembly as described in section 2.6.1. CRISPR-Cas9 in combination with homologous recombination was performed in freshly prepared *E. coli* MG1655 Δ*fadE* (JBEI 3111[177]) transformed with plasmid pKD46Cas9_sgRNA(p15A) electrocompetent cells for all the parallel gene knockouts, as described in section 2.3.5.

Firstly, *E. coli* MG1655 Δ*fadE* (JBEI 3111) strain was transformed with bacterial artificial chromosome FAS-A resulting in strain FS1. FS1, then, was transformed with heat-sensitive plasmid pKD46Cas9_sgRNA(p15A) resulting in strain FS1_HP. FS1_HP was used as the background strain for all the gene knockouts. FS1_HP was modified into FS2 after gene knockouts and pKD46Cas9_sgRNA(p15A) plasmid removal. JBEI 3111 was transformed with FAS-B to make strain FS3. Strain FS was formed by transforming FAS-B into strain FS2. Strain FS was tested with many different conditions to assess FABIO strategy.

Transformations were otherwise performed in chemically competent cells. Deletion of each gene group/cluster was confirmed by colony PCR and Miseq sequencing (of its colony PCR product, using the corresponding colony PCR primers). Colony PCR composition and thermocycler conditions are described in table 2.3.3 and 2.3.4. The complete list of DNA oligos used for colony PCR is provided in Table 4.2. In the sequential knockout approach, after each gene group/cluster deletion, modified strain was used for more deletions using a similar protocol. pMCC_synCRISPRs were cured by inducing the positive colonies with 500 μM IPTG in stationary phase for 24 hours. Thereafter, 5 μL of each strain was plated on LB agar with ampicillin, and single colonies were isolated, and replica streaked on LB agar with ampicillin and LB agar with chloramphenicol, with no growth on LB with chloramphenicol indicating the loss of pMCC_synCRISPRs.

4.3.3 Cell culture and F-OH measurement by GC-MS

Production strains were grown in LB medium containing the appropriate antibiotics overnight at 30 ^{0}C, and then adapted into M9-MOPs minimal medium containing 2% glucose. Cultures were induced by addition of 1 mM IPTG and incubated with shaking at 30C (F- OH strains)) for an additional 40 hours and the fatty alcohol contents were measured periodically. Detailed procedure is explained in section 2.4.1. Five standards containing C12, C14, C16, C16-1, C18, and C18-1 fatty alcohols were prepared and used for F-OH measurements as described in section 2.4.2.

Table 4.1. List of fatty acid biosynthesis type II genes, and their location and size in MG1655. Highlighted genes (in green, orange, and yellow) are gene clusters, co-located in the genome.

Fatty acid biosynthesis (Type II) genes	locus tag in MG1655	length in MG1655 (bp)
accA	b0185	960
accB	b3255	471
accC	b3256	1350
accD	b2316	915
fabD	b1092	930
fabB	b2323	1221
fabF	b1095	1242
fabH	b1091	954
fabG	b1093	735
fabA	b0954	519
fabZ	b0180	456
fabI	b1288	789
acpP	b1094	237
acpS	b2563	378
acpT	b3475	585
Lipopolysaccharide biosynthesis (including Lipid A)		
lpxA	b0181	789
lpxD	b0179	1026
lpxL	b1054	921
lpxM	b1855	972
Phospholipid biosynthesis		
plsB	b4041	2424
plsX	b1090	1071
plsC	b3018	738

Thioesterases		
tesA	b0494	627
tesB	b0452	861
fadM	b0443	399
Regulatory proteins		
fabR	b3963	648
fadR	b1187	720
total	27	22938

Table 4.2. List of colony PCR oligos and product size

gene/s	screening primer fwd name	screening primer fwd sequence	screening primer rev name	screening primer rev sequence	Pro duct size
recA	recA_screening_Fwd	cagccaggcagtcgcattc	recA_screening_Rev	gagtaaaaatggctatcgacgaa	200
accA	accA_screening_Fwd	gttagccgtcaggatgagaaa	accA_screening_Rev	ttcagaactttgcgaattacgc	200
accB+accC	accBaccC_screening_Fwd	atacagctgtcggggcatac	accBaccC_screening_Rev	gttggttttgataccgtcgatg	200
lpxD+fabZ+lpxA	lpxDfabZlpxA_screening_Fwd	aactacacggtgatggcgata	lpxDfabZlpxA_screening_Rev	aatcgttaatggacgctgttcag	200
plsB	plsB_screening_Fwd	catccggctttcttctgggt	plsB_screening_Rev	cttcgttgatccatgcagcg	551
acpS	acpS_screening_Fwd	ccactcgtttcaatggggga	acpS_screening_Rev	acaacgtgaaagccattgcc	583
fadM	fadM_screening_Fwd	cctgtgccggaatgtctcat	fadM_screening_Rev	gcaattccccttccagagct	575
lpxL	lpxL_screening_Fwd	atcagggggatttgctgtgg	lpxL_screening_Rev	cgttggaaatgcggttgtgt	540
tesA	tesA_screening_Fwd	aacacgccggtaaatcccat	tesA_screening_Rev	tgacttacccgatcccgact	508
fabI	fabI_screening_Fwd	ggcagaacgtcatactccgt	fabI_screening_Rev	ttgcacaatacggtcccctc	543

lpxM	lpxM_screening_Fwd	tatacggctggattcgccc	lpxM_screening_Rev	ttcagtgggtgacggtgaag	535
fabR	fabR_screening_Fwd	cagacatcgtgatgggcgta	fabR_screening_Rev	ccatatggagtgcggcgtat	577
accD	accD_screening_Fwd	cacaaatggcgctggtttca	accD_screening_Rev	gctgctgatcgtcgggatta	591
fabA	fabA_screening_Fwd	aaggtcacaatggctgggag	fabA_screening_Rev	atccccacagctaacgttcg	582
fadR	fadR_screening_Fwd	gagtttgtaccagtcggggg	fadR_screening_Rev	aacagcaaaacgaagacggc	519
fabB	fabB_screening_Fwd	cttttgcgttagtgccggag	fabB_screening_Rev	cgggaaacaggtgtaccctc	528
tesB	tesB_screening_Fwd	tcgccattacgatcgtctgg	tesB_screening_Rev	tcgccggaagagtggttaac	516
acpT	acpT_screening_Fwd	agggatgctacttcggtgc	acpT_screening_Rev	ttagtggcccgacattcacc	545
plsX+fabH+fabD+fabG+acpP+fabF	plsX-fabF_screening_Fwd	aagcggacatggtctttggt	plsX-fabF_screening_Rev	caagctgaggccagaagtca	552

Table 4.3 List of plasmids and strains used in this chapter

	Name	Description	Reference
Plasmid	pE5C.cgFAS1A	colE1 ori, Cmr, PLacUV5-fas-IA (C. glutamicum)	[84]
Plasmid	E5C-'tesA.fadD	colE1 ori, Cmr, PLacUV5-'tesA.fadD	[84]
Plasmid	pB5K.Maqu2220	BBR1 ori, Kanr , PLacUV5-Maqu2220	[84]
plasmid	pKD46.Cas9.sgRNA(p15A)	r101 ori, Ampr, constitutive promoter Cas9, pBAD-lambda proteins, pTrc sgRNA(p15A)	This study

plasmid	FAS-A	bacterial artificial chromosome	This study
plasmid	FAS-B	BBR1 ori, Kanr, pT7-accD.accBC.accA, pLacUV5-FAS1(CG), pTAC-fadD.ucFatB1.Maqu2220, pBAD-sgRNA(FAS-A)	This study
plasmid	pMCC_synCRISPR	p15A ori, Cmr, synCRISPR	This study
Strain	MG1655	F- lambda- ilvG- rfb-50 rph-1	[183]
Strain	JBEI-3111	MG1655 ΔfadE	[177]
Strain	JBEI-5029	MG1655 ΔfadE, araD::acpS (C. glutamicum)	[84]
Strain	JBEI-8085	JBEI-5029 bearing E5C-CG.FAS1A and pB5K.Maqu2220	[84]
Strain	JBEI-2096	JBEI-3111 bearing pA5K.fadD.ucFatB1.Maqu2220	[181]
Strain	JBEI-9017	JBEI-5029 bearing pE5C.'tesA.fadD and pB5K.Maqu2220	[84]
Strain	FS	FS2 bearing FAS-A and FAS-B	This study
Strain	FS1	JBEI-3111 bearing FAS-A	This study
Strain	FS2	FS1 Δ(accA, accB, accC, accD, fabD, fabB, fabF, fabH, fabG, fabA, fabZ, fabI, acpP, acpS, acpT, lpxA, lpxD, lpxL, lpxM, plsB, plsX, plsC, tesA, tesB, fadM, fabR, fadR)	This study
Strain	FS3	JBEI-3111 bearing FAS-B	This study

Strain	FS1_HP	FS1 bearing plasmid pKD46.Cas9.sgRNA(p15A)	This study

4.4 Results and Discussion

4.4.1 FABIO Strategy

The FABIO strategy is described in Figure 4.1. A defragged genome and BAC modules are created to make fatty acid biosynthesis orthogonal and independent (FABIO). Firstly, the BAC-A module, a replicating and repressible bacterial artificial chromosome, is synthesized by placing selective native pathway genes under native constitutive promoters (Figure 4.2) with a dual origin of replication 1) CEN/ARS for replication machinery in yeast and 2) oriS for unidirectional replication in *E. coli* and ParA and ParB partitioning proteins.

For each selected gene, 5' and 3' protospacer sequences (target sites) adjacent to PAM sequences, are designed. Protospacer sequences in the BAC-A module were modified by codon optimization to be dissimilar to the CRISPR sites of target genes located on the genome to direct the double stranded break (DSB) inflicted by Cas9 for gene deletions towards the genome as opposed to the BAC-A. Consequently, BAC-A was transformed into an *E. coli* production strain, JBEI-3111 (Table 4.3), to supplement the cellular growth post deletion of essential genes from the genome. Selected native genes (groups and clusters, Table 4.4) were removed from the genome via the CRISPR-Cas9 approach (Figure 4.4) to generate a defragged genome free of native gene regulation. A

heterologous pathway for fatty alcohol production in a FAS-B module is introduced to work orthogonally and independently of FAS-A. Subsequent minimal strain can repress FAS-A and express FAS-B at will and eliminate the native gene regulatory effects.

4.4.2 Fatty Acid Synthesis (FAS) modules construction

4.4.2.1 FAS-A (or BAC-A) construction

Bacterial artificial chromosome (BAC) libraries play a pivotal role in genomics studies and are increasingly being used as the foundation for physical mapping, map-based cloning, and sequencing projects [184, 185]. The Copy number of these cloning vectors is rigorously maintained at one by the BAC replication system derived from the *E. coli* F plasmid. Although BAC vectors can maintain 100-200 kb of DNA pieces, BAC-A is about 40 kb in size, making it much easier to propagate and stably maintain in the recombination deficient *E. coli* strain. In this study, BAC hosts the native fatty acid biosynthesis genes and serves to complement them in their absence from the chromosome. Due to one copy of the BAC vector, the expression level of the pathway genes placed on the BAC mimicked the naturally available chromosomal genes.

For BAC-A, twenty-seven native genes collectively responsible for pathway regulation like ACC complex, *fadR* and *fabR*, native FAS type II, and formation of products like lipopolysaccharides biosynthesis phospholipids biosynthesis, were selected and placed on a replicating plasmid (Figure 4.2A). As shown in the highlighted sequence in Figure 4.2A, protospacer sequences were codon-optimized and changed from the genome sequence to construct

unique CRISPR sites that introduce the double stranded breaks in the genes in the genome. Besides carrying out cellular growth functions, it also produces acyl-ACPs (Figure 4.1) from the native FAS II system. Conditional repression after the growth phase allows uncoupling from cell growth and acyl-ACP regulation. The synthesis of BAC-A was carried out at the Joint Genome Institute.
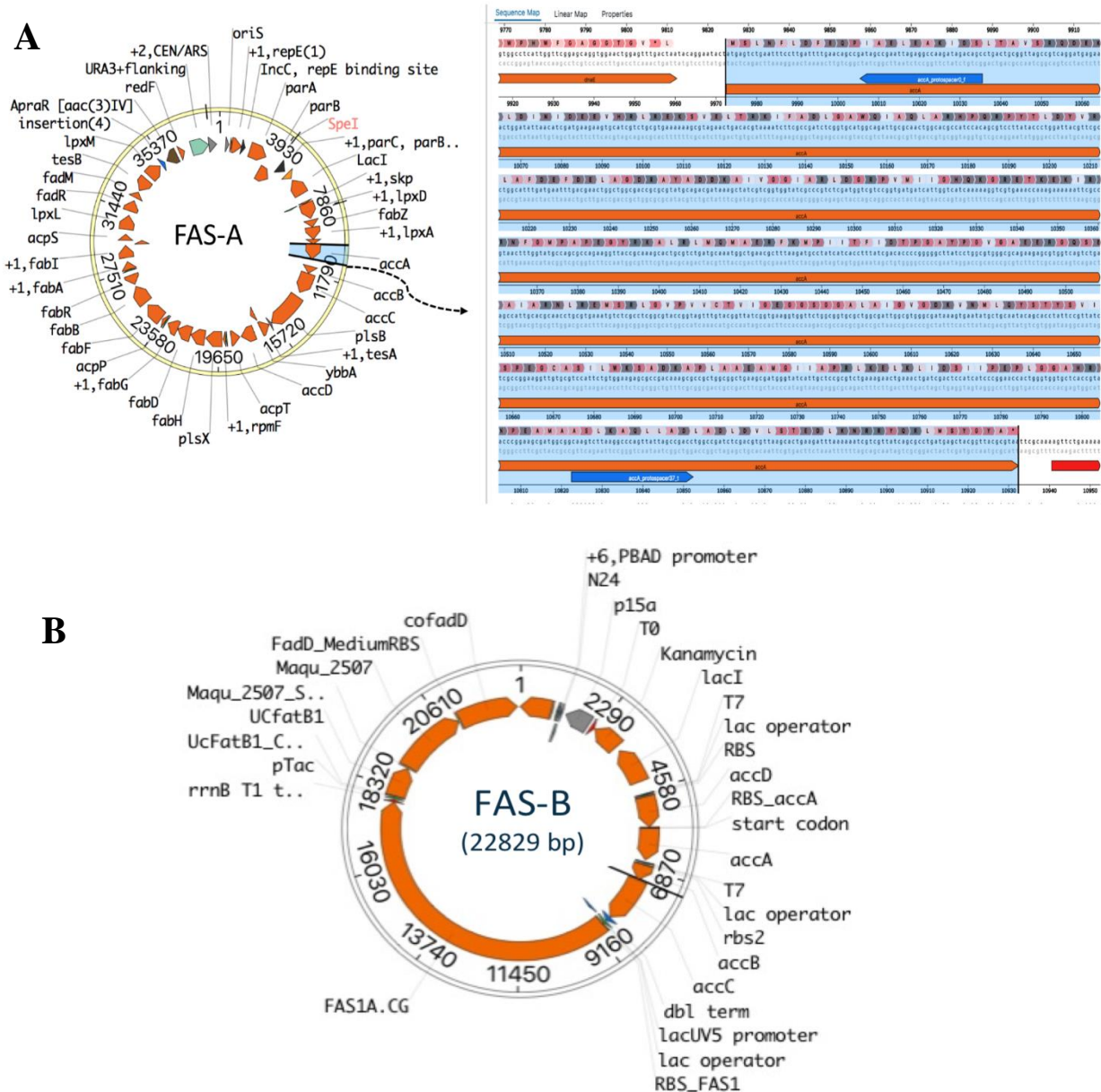
Figure 4.2 (A) Selected genes responsible for endogenous metabolism are placed under native constitutive promoters on a replicating artificial chromosome (also known as bacterial artificial chromosome or BAC), capable of repression, with dual origin of replication 1) CEN/ARS for replication machinery in yeast and 2) oriS for unidirectional replication in *E. coli* and ParA, and parB partitioning proteins. For each selected gene, target sites (protospacers) adjacent to PAM sequences are designed. They are modified (slightly) by codon optimization to make them dissimilar so that Cas9 can specifically target the native genes in the genome. Highlighted in blue on the right are the target sites in gene the *accA* from the BAC plasmid. (B) FAS-B module for the expression of the heterologous pathway for fatty alcohols production.

4.4.2.2 FAS-B construction

The primary purpose of FAS-B is to enable the fatty acid biosynthesis pathway to produce biofuel precursors uncoupled from cell growth and acyl-ACP regulation. FAS-B (Figure 4.2B) module has four main features: 1) repression of FAS-A via CRISPR/Cas9 induced by arabinose, 2) native ACC complex (accABCD) to produce malonyl CoA, 3) heterologous FAS I system from *C. glutamicum* to produce acyl-CoAs[84] (coenzyme A) from malonyl CoA induced by IPTG, and 4) genes encoding for thioesterase (*ucFatB1* (or BTE) from *Umbellularia californica*), acyl-CoA synthase (*fadD* from *E. coli*), and acyl-CoA/acyl-ACP reductase (*maqu_2220* from *Marinobacter aquaeolei* VT8)[181] for conversion of acyl-ACP and acyl-CoA to fatty alcohols induced by IPTG. Heterologous FAS I from *C. glutamicum* was selected based on the findings in publication Haushalter et al [84] and genes encoding for thioesterase (UcFatB1), acyl-CoA synthase (FadD), and acyl-CoA/acyl-ACP reductase (Maqu_2220) were selected based on the findings in publication Opgenorth et al [181]. When induced after the initial growth phase, heterologous type I FAS

makes tailored acyl-CoAs that are preferred precursors of biofuels and chemicals compared to regulation causing acyl-ACPs. The FAS-B system also enables more control over fatty acid properties (e.g., chain length, unsaturation, branching) to modulate the fuel properties of the fatty acid-derived products. Finally, FAS-B decouples the fatty acid biosynthesis pathway from cell growth, self-regulation, and consumption of fatty acids for other biomolecules such as phospholipids and lipo-polysaccharides.

4.4.3 Defragged genome construction by knocking out native genes using CRISPR Cas9

CRISPR Cas9 tools were developed to allow multiple changes concurrently and sequentially to optimize the complex metabolic pathway, such as the fatty acid biosynthesis pathway. Specifically, in this study, the task is to remove 27 genes (Table 4.2) from the chromosome that are co-present in the BAC-A module.

4.4.3.1 Parallel and sequential gene knock-out

The majority of the selected genes are distributed across the genome, with a few tandemly located in the genome. Based on their size and location in the genome and BAC-A, genes were gathered into nine groups (randomly distributed) and clusters (tandemly located) for efficient knock-out from the genome. As mentioned earlier, fatty acid metabolism genes pose tight regulation. Not only could their deletion lead to cell death, but over-expression of specific genes (due to already existing one copy of the genes in the form of BAC-A) may act lethally for the growth too.
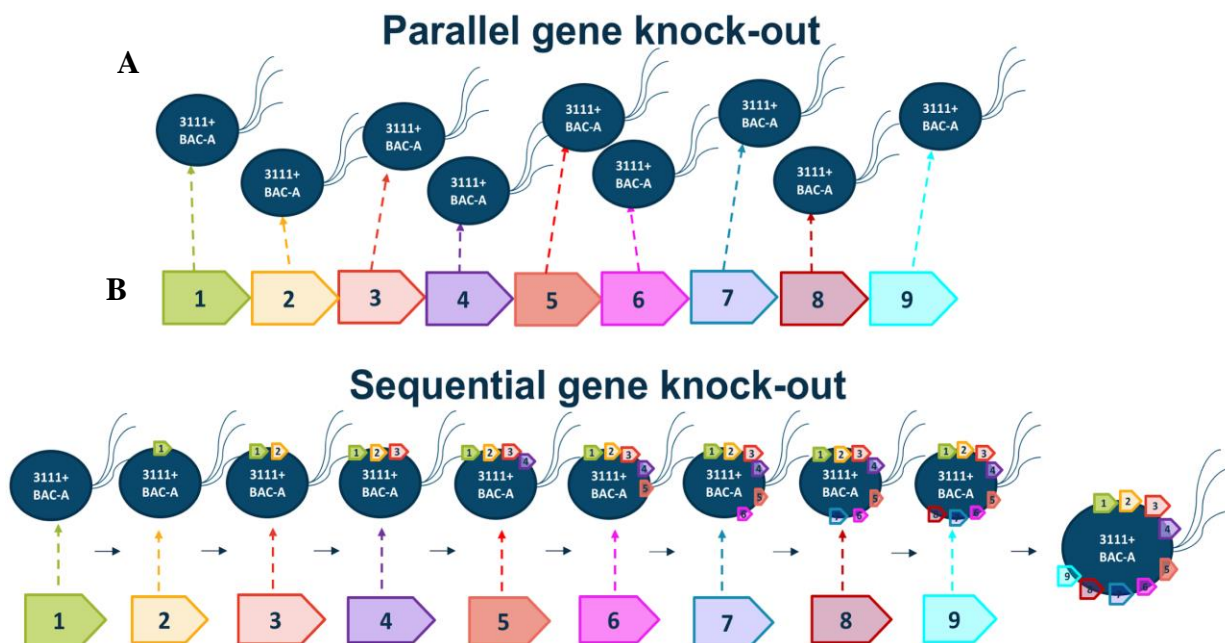
Figure 4.3. Parallel and sequential knockouts of FABIO gene groups and clusters**.** 27 FABIO genes were divided into 9 gene groups (randomly located in the genome) and gene clusters (co-located in the genome) for faster CRISPR-cas9 knockout rate. **A**. Parallel knockout approach to determine the order for knocking out the growth essential genes and **B**. Once the order is determined, genes are knocked-out following a sequential approach.

To assist with overcoming the complexity associated with their expression level and understand the viability of gene groups and clusters, parallel and sequential gene knock-out approaches (Figure 4.3) were designed. In the parallel gene knock-out approach, nine gene groups and clusters were knocked out from the genome in a parallel manner and their success rate was monitored. Based on

their success rate and potential toxicity problems by the expression of the pathway genes, order for sequential gene knock-out was carefully accomplished.

**Table 4.4. List of gene groups and gene clusters.**

No.	Gene groups	Deletion size
1	recA+accA	2022
2	plsB+acpS+fadM	3201
3	tesA+accD+acpT	2127
4	fabB+fabR+fabA	2388
5	fadR+tesB+lpxM	2553
6	fabI+lpxL	1710
Gene clusters		**Deletion size**
7	accB+accC	1821
8	lpxD+fabz+lpxA	2271
9	plsX+fabH+fabD+fabG+acpP+fabF	5169

4.4.4 Development of CRISPR/Cas9 for genome editing

CRISPR Cas9 system used in this work is summarized in Figure 4.4. It consists of three plasmids – pMCC-synCRISPRs, pKD46-cas9-sgRNA(p15A), and suicidevector_fixingtemplate. pMCC-synCRISPRs express the guide RNA to target specific sequences (same as target sites designed in BAC-A) in the *E. coli* genome to be cut by Cas9. Every synCRISPR (Supplemental Table 6.2) consists of *S. pyogenes* promoter (constitutive promoter for driving the transcription of gRNA (guide RNA), Repeat I, Spacer left (30 bp sequence that

guides Cas9 to cut the genome in the reverse strand i.e protospacer sequence preceded by a 'CNN' protospacer adjacent motif (PAM), Repeat I, Spacer right (30 bp sequence that guides Cas9 to cut in a protospacer in the 5'-3' strand (followed by a 'NGG' PAM), Repeat III, and *S. pyogenes* terminator sequence. Double cut induced by two spacers (left and right) ensured higher specificity and efficacy.
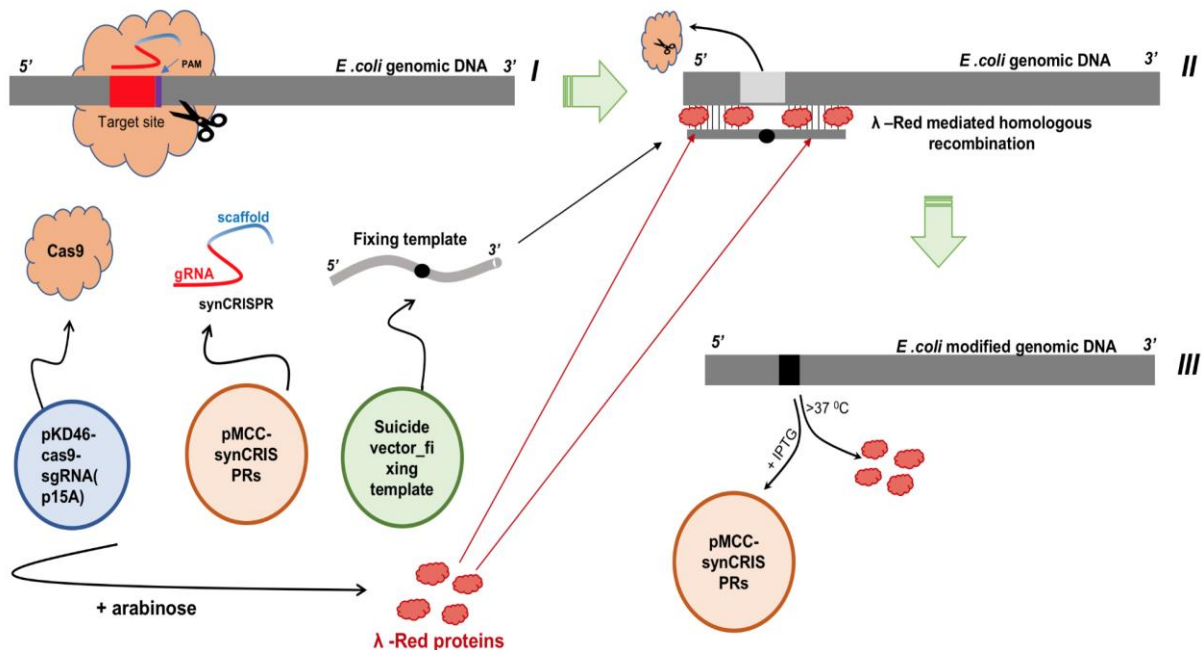
Figure 4.4. CRISPR-cas9 and λ –Red homologous recombination-mediated genomic DNA modification. Cas9 endonuclease expressed constitutively by heat-sensitive plasmid pKD46-cas9-sgRNA(p15A) and gRNA encoded by the pMCC-synCRISPR plasmid form a gRNA-Cas9 ribonucleoprotein complex which specifies the site for cleavage for Cas9, resulting in a double-stranded break in the *E. coli* genomic DNA. Subsequently, λ-Red proteins expressed by pKD46-cas9-sgRNA(p15A) upon arabinose induction along with fixing template (through in a suicide vector) repair the DNA double strand break via λ –Red mediated homologous recombination. Once genomic DNA is cut, a

deletion, insertion or modification in the genomic DNA can be incurred by the fixing template. After establishing a modification, plasmids pMCC-synCRISPR and pKD46-cas9-sgRNA(p15A) were removed by inducing the strain with IPTG to encode for gRNA targeting the ori p15A of the plasmid pMCC-synCRISPR and subsequently, by growing the strain at 37 ^{0}C or more promotes stability of the strain under stressful nutrient depleting conditions.

pKD46-cas9-sgRNA(p15A) consists of λ-red proteins under the control of arabinose inducible promoter, Cas9 protein under constitutive promoter to cut the target site in pMCC-synCRISPR, and guide RNA targeting p15A ori of pMCC-synCRISPR under pTrc promoter to facilitate removal of pMCC-synCRISPR post deletion of the corresponding gene. By keeping carbenicillin in the medium and the temperature at 30 °C, this plasmid was maintained during iterative rounds of modification. Suicidevector_fixingtemplate delivers a 1000 bp sequence of fixing template containing 500 bp with upstream and downstream homology of the target gene for deletion to repair the double-stranded break induced by Cas9 in the presence of λ-red proteins expressed by pKD46-cas9-sgRNA(p15A). These suicide vectors have R6K ori, which doesn't allow them to propagate in the *E. coli* background strain. Homology-directed repairs were achieved with the help of fixing templates containing homology arms. Plasmid maps for pKD46-cas9-sgRNA(p15A), pMCC-synCRISPR, and fixing template are shown in Supplementary Figure 6.3.

The first step was transforming the pKD46-cas9-sgRNA(p15A) plasmid into the background strain (strain FS1) resulting in a strain FS1_HP. This strain becomes the starting strain or strain I (Figure 4.4) for the following iterative genomic modifications. After electroporating the pMCC-synCRISPR and fixing template into the *E. coli* strain expressing Cas9 and λ-red, Cas9 protein cuts the target site guided by gRNA, and in the presence of λ-red proteins, strain

II gets repaired by fixing template. Received colonies are confirmed by colony PCR (Table 4.2) and strain II is subjected to removal of pMCC-synCRISPR by inducing the p15A guide RNA located in pKD46-cas9-sgRNA(p15A) with IPTG. pKD46-cas9-sgRNA(p15A) can be heat cured post genomic modifications. The iterative process for gene deletion and removal of auxiliary plasmids is summarized in Figure 4.5.
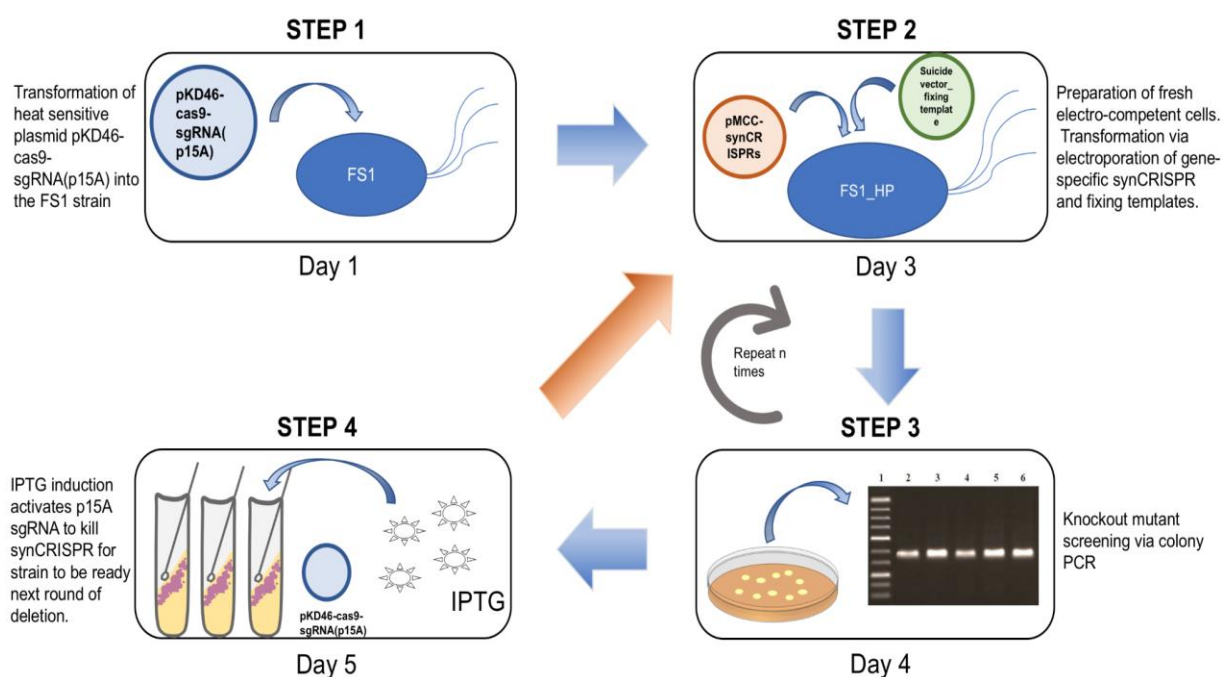
Figure 4.5. FABIO gene knockout workflow in *E. coli* using CRISPR-cas9 methodology consists of four steps. An engineered *E. coli* strain (JBEI-3111 or 3111) transformed with BAC-A plasmid, resulting in strain FS1, is the base strain for mutagenesis. On day 1, (Step 1) heat-sensitive plasmid pKD46-cas9-sgRNA(p15A) responsible for expressing Cas9 protein, λ-Red proteins, and guide RNA targeting ori p15A of pMCC-synCRISPR is transformed into base strain FS1 resulting in strain FS1_HP. Day 2 is growing the FS1_HP strain for preparing electro-competent cells on day 3 (step 2). Following that, gene-specific fixing templates and synCRISPR plasmids were immediately

transformed via electroporation and incubated overnight to carry out gene knock-out using CRISPR-cas9 and λ-Red homologous recombination. Day 4 (step 3) is screening the knockout mutant via colony PCR. Upon confirming, positive strain is grown in LB media containing 500 µM IPTG on day 5 (and step 4) to cure pMCC-synCRISPR by encoding gRNA targeting p15A ori of the synCRISPR plasmid. On day 6 (back to step 2), the strain is ready for another round of deletion of FABIO genes, and steps 2-4 are repeated n times.

We developed a successful CRISPR Cas9 model that allowed multiple simultaneous and sequential genome modifications with all the auxiliary plasmids removal post experimentation. Up to three guide RNAs for randomly located genes were designed in one pMCC-synCRISPR to allow multiple gene deletions simultaneously. We effectively performed removal of the largest DNA size of 5.2 kb (6 co-located genes) from the genome. Efficiency of the model was further improved by ensuring all the auxiliary plasmids were curable from the strain to avoid any unrequired metabolic burden during the product formation phase.

4.4.4.1 Strain development via sequential gene group deletions and its challenges

Strain development to create a defragged genome was a continuous and systematic process (Figure 4.3) that took multiple sequential iterations. Steps 2 to 4 were repeated for every genomic modification. Hundreds of colonies were screened via colony PCR to confirm the simultaneous gene deletions, and the final strain (strain FS) was confirmed by sequencing amplicon via Miseq sequencing.

One precise observation from the iteration process was that the deletion of some gene groups was more challenging than others. Although, in the parallel approach, all the gene groups and clusters resulted in successful deletions with high efficiency (>80%, data not shown), sequential knock-out suffered from reduced efficiency due to continuous *E. coli*'s evolution, and excessive burden due to protein and metabolic flux imbalance on cell machinery caused by incessant deletion of native genes. Even though a fixing template was provided for Homology Directed Repair (HDR) supported by homologous recombination, it exhibited drop in recombination efficiency in sequential knockouts. As is widely reported, recombination suffers drastically as more and larger DNA fragments are removed or inserted. In other words, the resulting population of bacterial colonies contained a combination of wild type alleles, evolved *E. coli* alleles, and HDR-repaired alleles. It required a cost and labor-intensive process to confirm the presence of the desired genome edit.

4.4.5 Stability and verification of the defragged *E. coli* genome

It was crucial to check the stability of the defragged genome and the ability of BAC-A to complement cell growth. Based on the FABIO strategy, a modified strain with a defragged genome shouldn't display any growth in the absence of the BAC-A module. The BAC-A module was deleted from the strain FS by inducing the culture in the late growth phase with 100mM arabinose. Deletion of BAC-A was confirmed by plating the induced culture on LB agar plate supplemented with Apramycin. No growth indicated that BAC-A has been removed. Arabinose-induced culture was also plated on LB agar plate supplemented with ampR to test if strain FS is capable of growing in the absence of growth complementary to BAC-A. Almost no growth was observed

for strain FS after it lacked BAC-A. This experiment confirmed that 1) defragged genome lacks the growth essential genes, 2) BAC-A complementation works, and 3) BAC-A is repressible.

The sequence of genomic DNA extracted from the arabinose induced strain was verified via Miseq sequencing. It was essential to remove the BAC-A from the strain for sequencing analysis to avoid any misalignments to the homology sequences i.e., native genes in BAC-A. Strain FS was also sequence confirmed to contain FAS-A and FAS-B and defragged genome, which was used further for fatty alcohol production experiments.

4.4.6 Fatty alcohol production with a defragged genome and BAC modules

The FABIO strain consisting of the defragged genome and FAS modules was tested under five different FAS-A repression and FAS-B expression conditions for producing fatty alcohols in the C12-C18 range (Figure 4.6). Condition 4, when FAS-A is repressed in the stationary phase, and FAS-B is expressed in the early log phase of cell growth, yielding the maximum total fatty alcohols. Figure 4.6 confirms the feasibility of the FABIO model with 2.5-fold increased production (165 mg/L of total alcohols) compared to an engineered non-FABIO strain (3111+FAS-B strain). It provides convincing evidence that the FAS-A (native pathway genes) repression facilitates decoupling the fatty alcohol production from cell growth, other biomolecules and acyl-ACPs.
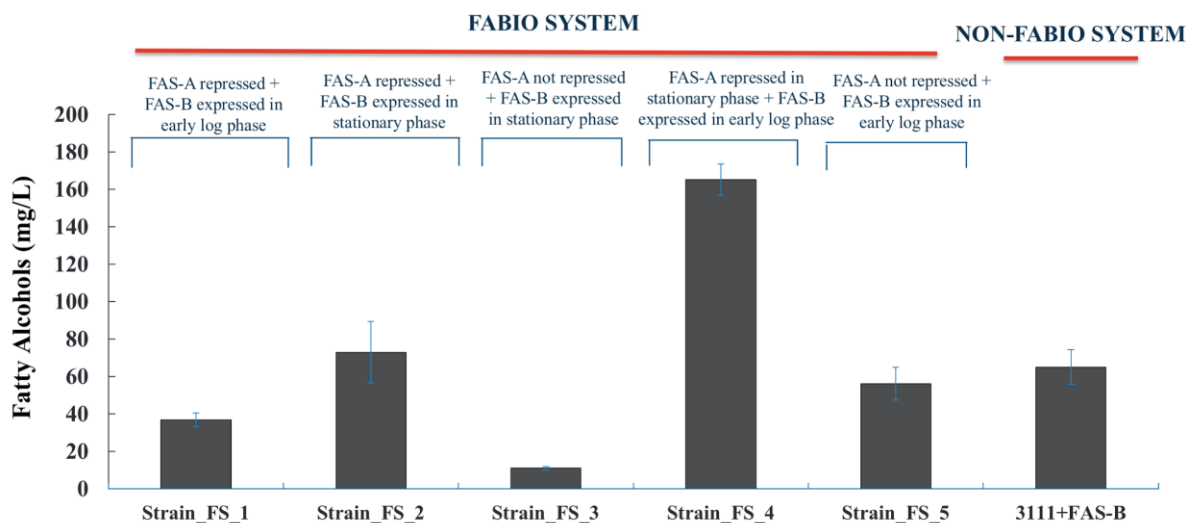
Figure 4.6. Comparison of FABIO and Non-FABIO systems. 5 different conditions of repression and expression during of FAS systems were tested in FABIO system (consisting of a defragged genome, FAS-A, and FAS-B) 1)FAS-A is repressed and FAS-B is expressed in early log phase 2) FAS-A is repressed and FAS-B is expressed in stationary phase 3) FAS-A is not repressed and FAS-B is expressed in stationary phase 4) FAS-A is repressed in stationary phase and FAS-B is expressed in early log phase 5) FAS-A is not repressed and FAS-B is expressed in early log phase. Non-FABIO system is a JBEI 3111 strain transformed with FAS-B (absence of FAS-A and defragged genome). Error bars indicate the standard deviation of three biological replicates.

Similar total alcohols yields in condition 5 (FAS-A is not repressed, and FAS-B is expressed in early log phase) and non-FABIO system, strain FS3 (3111+FAS-B), confirm that FAS-A effectively replaces the functionalities of fatty acid native biosynthesis based on the growth assays and total alcohols production. Lower titers were observed in other tested conditions indicating that further optimization on different conditions may result in different alcohol titers and provide us a deeper understanding of these orthogonal modules.

We compared the FABIO strain with three other engineered strains that have been reported before (Figure 4.7) - JBEI 8085 JBEI 8085 (JBEI-5029 bearing pE5C-CG.FAS1A and pB5K.Maqu2220), JBEI 9017 (JBEI-5029 bearing pE5C.'tesA.fadD and pB5K.Maqu2220), and JBEI 2096 (JBEI 3111 bearing pA5K.UcfatB1.Maqu2220.fadD) (Table 4.3). Unfortunately, total alcohol production levels for the FABIO system are significantly lower (4.6-5.5 times) than the other engineered strains. In Figure 4.6, lower production of alcohols by the engineered non-FABIO strain (FS3) indicates that FAS-B is underperforming and could be the reason for the significantly low titers of the FABIO strain when compared with other strains (figure 4.7).
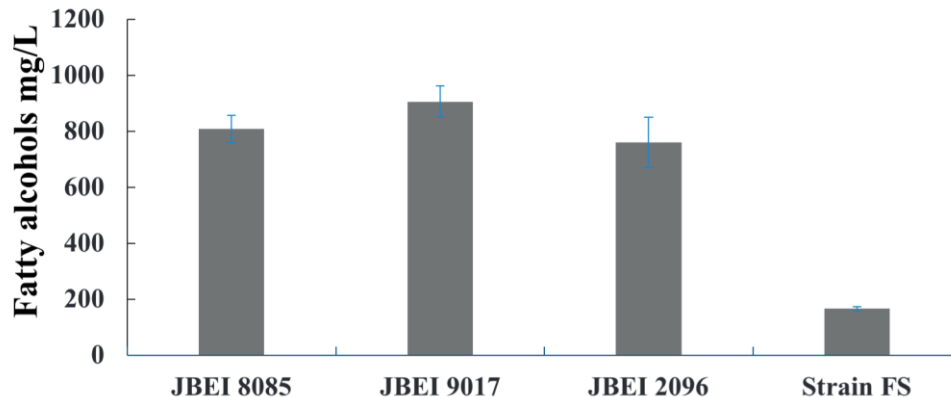

Figure 4.7. FABIO strain (FS) compared to other engineered JBEI strains for fatty alcohol production. Other JBEI strains included are JBEI 8085 (JBEI-5029 bearing pE5C-CG.FAS1A and pB5K.Maqu2220), JBEI 9017 (JBEI-5029 bearing pE5C.'tesA.fadD and pB5K.Maqu2220), and JBEI 2096 (JBEI 3111

bearing pA5K.UcfatB1.Maqu2220.fadD) (Table 4.3). Error bars indicate the standard deviation of three biological replicates.

These results highlight the potential for further optimization of the FAS-B module. One possible solution is splitting the FAS-B module into two plasmids to distribute the heterologous protein expression burden similar to JBEI strains 8085 and 9017. JBEI 8085, 9017, and 2096 strains illustrate the drastic impact of gene expression strength, balanced heterologous pathway distribution, and codon optimization on F-OH production. Using a similar expression system to these strains for FAS-B may improve the total alcohol titers drastically which can further be improved (~2.5 times reported here) when combined with the FAS-A module in the FABIO system. It is clear from the data points in Figure 4.6 that 2.5-fold improvement in condition 5 for the strain FS is due to conversion of acyl CoA to F-OH instead of acyl-ACP to F-OH which further suggests the decoupling of native pathway from acyl-ACP regulation. To our knowledge, FABIO is one of its kind models with which we successfully attempted to improve the native fatty acid biosynthesis deregulation.
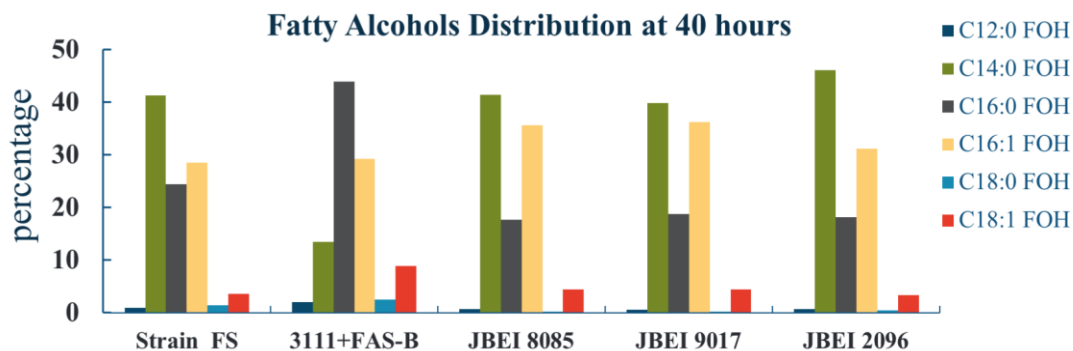
Figure 4.8. The chain length distributions of the five strains are indicated. While the total fatty alcohol titers varied in the five strains, the proportions of different chain lengths remained similar.

When we compared the chain length distribution of total alcohols in the five strains, we found all the strains exhibited consistent distribution of different F-OHs, though their titers showed variations. C14-OH and C16-OH are contributing more than 90% towards the total alcohols. It is unambiguous to evaluate heterologous FAS1 from *C. glutamicum* for F-OH production as the titers are overall very low. This needs further investigation to ensure all the heterologous pathway components are working individually and in coherence with each other.

4.5 Discussion

In this chapter, we successfully implemented independent and orthogonal FAS modules in *E. coli* for stable and improved F-OH production. Our main goal was to establish a system where F-OH can be decoupled from the cell growth, acyl-ACPs regulation, and production of other biomolecules. Two modules FAS-A (complemented cell growth) and FAS-B (heterologous pathway expression) were functioning as intended in the defragged genome created via strain development methodologies. We developed a CRISPR-Cas9 method compatible with the design-build-test-learn (DBTL) cycle for multiplex genome editing. We targeted simultaneous and consequential chromosomal gene deletions along with the removal of all auxiliary plasmids through this method. There are only a few published reports on the successful integration of entire metabolic pathways and in our knowledge none on continuous and

multiple modifications in a concurrent fashion related to complex pathways. Our development of the CRISPR-Cas9 tool is a good step in that direction. Though we successfully screened clones with positive gene edits via high throughput and automated colony PCR and Miseq sequencing, the process became more strenuous with continuous deletions and confirmation. A more systematic and efficient screening tool is desirable to confirm the multiplex genome editing to generate industrially relevant strains.

We addressed the fatty acid pathway regulation by strategically selecting the native genes for genome removal with three main objectives - no metabolic flux towards formation of lipopolysaccharides and phospholipids, no glucose consumption towards cell growth, and removal of acyl-ACPs - thus creating a defragged genome (Figure 4.1) along with FAS modules. FAS-A is encoding for the genes knocked out from the genome thus supplementing the cell growth and after its suppression, aiding the survival of strain with adequately expressed growth elements. With FAS-B, we are diminishing the self-regulation by acyl-ACP by utilizing heterologous type I FAS (IPTG inducible) which produces acyl-CoA as a precursor for fatty acid derived products. To convert the acyl-ACPs produced by native pathway (FAS-A) expressed by FAS-A during cell growth phase and acyl-CoAs produced by heterologous pathway (FAS-B) into F-OHs, *maqU2220* (CoA-ACP reductase) along with *ucFatB1* (thioesterase) and *fadD* (acyl-CoA synthase) genes were encoded by FAS-B module under IPTG inducible promoter pTac. Orthogonality of FAS modules is achieved by their expression/repression independently e.g repression of FAS-A is accomplished by CRISPR-Cas9 mediated FAS-A targeted guide RNA expressed by FAS-B (arabinose inducible) and expression of FAS-B module is accomplished by IPTG induction.

As described in the Results section, an engineered strain with FABIO strategy produced 165 mg/L of F-OHs (Figure 4.6) and showed 2.5-fold improvement over an engineered non-FABIO strain. Even though there's significant total alcohol production improvement with the FABIO system, titers are almost 5-fold lower than other established and previously reported strains (Figure 4.7). Similar F-OH titers by 3111+FAS-B and strain_FS_5 (under condition 5) in Figure 4.6 indicate that the overall performance of FAS-B is underwhelming, and the heterologous pathway expressed by FAS-B needs to be further optimized. These data points, along with F-OH titers shown by strain_FS_4 (Figure 4.6, condition 4), also confirm that repression of FAS-A is (1) possible when arabinose is added, (2) improving the F-OH titers, (3) decoupling the production of the desired chemical from cell growth, acyl-ACPs, and other biomolecules, and (4) complimenting the cell growth in a defragged genome.

Other engineered strains (Figure 4.7) is expressing and distributing the burden of the heterologous pathway on multiple high/medium copy plasmids which is resulting in comparatively more F-OH production. Pathway components of the FAS-B module need to be tested further or replaced for better gene expression directed towards higher F-OH titers. Although they have worked very well individually or when expressed on different plasmids in other reports [84, 181], this is the first report used in coherence under one replication system. Uneven pathway distribution and protein expression may have led to its weak performance. The data points in Figure 4.6 suggest that with improved FAS-B expression (similar to other JBEI strains), F-OH titers could be significantly improved in the FABIO system.

Type I FAS from *C. glutamicum* in JBEI 8085 strain [84] didn't demonstrate clear assessment on its performance as compared to native FAS type II. It was

suggested that heterologous type I FAS has potential to perform better in the absence of native type II FAS, but we didn't see any evidence of this in the present report. It is important to evaluate other types of FAS from different species as it is a key component of FAS-B module. Previous report [181] on genes Maqu_2220, UcFatB1, FadD indicated the complexity of predicting protein levels tied with RBS sequences, different gene orthologs, and machine learning. Even with DBTL infrastructure and synthetic biology advancements, there are still many challenges related to reliably modulate protein expression, unknown toxicity levels of pathway genes, and high metabolic burden on single plasmid effects to overcome. FAS-B can be made more versatile by swapping Maqu_2220 with Maqu_2507 from *Marinobacter aquaeolei* to make C_{12} F-OH, which are higher value compounds. By swapping Maqu_2220 with FadM, FadB, and Mlut_1700 in the heterologous pathway, methyl ketones can be produced.

Overall, this study highlights the challenges of selecting between contrasting approaches such as native FAS II vs heterologous FAS I, and multiple plasmids vs single plasmid pathway expression, which are going to be key parameters for optimizing the FABIO strains towards high F-OH titers. We expect that exponential increases in our capacity to synthesize DNA and perform strain development will make the DBTL cycles inexpensive and fast. This approach of deregulating the metabolic pathways can be applied systematically to any molecule, pathway, and host in the near future.

4.6 Conclusion and Outlook

We are reporting a FABIO system designed to deregulate the native fatty acid biosynthesis pathway from cell growth, acyl-ACP, and phospholipids, and lipopolysaccharides by introducing two modules FAS-A and FAS-B. These two modules perform orthogonal and independent of each other. A defragged genome was successfully created by developing and deploying the multiplex CRISPR-Cas9 method and transformed with FAS-A and FAS-B modules. Following, fatty alcohols titers were measured by testing different expression-repression conditions of a FABIO strain and reported to be 2.5 times higher than an engineered non-FABIO strain consisting of the same FAS-B module. These results are only a proof-of-concept study, and thus further improvement of modules, product titers, and validation of the system remain to be elucidated.

We exclusively focused on creating two independent modules and validation of the FABIO system. We didn't, for example, explore alternative promoters, RBSs, enzyme orthologs, operon structures, or terminators. All these approaches will benefit the FABIO approach further and could be explored in the future. Beyond the C14 and C16 fatty alcohol production, we envision using the FABIO strategy for various acyl-ACP/acyl-CoA-derived products such as methyl ketones, free fatty acids, or even different carbon-chain fatty alcohols by employing target compound-specific thioesterase and reductases. Regulation of native fatty acid biosynthesis pathway in other organisms beyond *E. coli* can also be addressed by applying the FABIO strategy.

Chapter 5

Concluding remarks

The present doctoral thesis has covered the optimization of metabolic pathways in *E. coli* for high commodity chemical titers. Genome editing tools are developed for editing large and complex metabolic pathways with the objective of fine-tuned and orthogonal expression of pathway genes. We demonstrate the power of synthetic biology tools like machine learning, DBTL infrastructure, and automated cloning and screening procedures by exploring them to accelerate the strain engineering efforts and thereby facilitating the entire developmental process. Development of the parallel integration and chromosomal expansion technique in *E. coli* opens the door for a myriad of genetic-level metabolic developmental possibilities in other microbes. A modular approach like FABIO is successfully implemented in *E. coli* to deregulate the native pathways like fatty acid biosynthesis. With the potential to be applicable in other prokaryotes and eukaryotes, FABIO can be further developed to improve fatty alcohol titers. Finally, adopting these genetic development approaches in strain engineering for controllable and stable protein expression is expected to drive the microbial processes towards commercialization.

Chapter 6

Appendix

6.1 List of Figures

6.2 List of Tables

6.3 List of Abbreviations

ACP	fatty acyl-carrier protein
acyl CoA	acyl-Coenzyme A
Amp	ampicillin
CAD	computer aided design
CAGR	compound annual growth rate
CIChE	chemically inducible chromosomal evolution
CIGMC	chromosomal integration of genes with multiple copies
CRIM	conditional-replication integration and modular
CRISPR	clustered regularly interspaced short palindromic repeats
DBTL	design-build-test-learn
dH$_2$O	deionized water
DMAPP	dimethylallyl diphosphate
DSB	double stranded break
E. coli	*Escherichia coli*
Em	emission
Ex	excitation
FABIO	fatty acid biosynthesis inducible and orthogonal
FAD	fatty acid degradation
FAS	fatty acid synthesis
FAS-A	fatty acid synthesis A
FAS-B	fatty acid synthesis B
FFA	free fatty acids
FLP-FRT	flippase-flippase recognition targets
FOH	fatty alcohols
G3P	glyceraldehyde 3-phosphate
GC-FID	gas chromatography flame ionization detector
GCMS	gas chromatography mass spectrometry

GFP	green fluorescent protein
GOI	gene of interest
HDR	homology directed repair
HMGR	HMG-CoA reductase
HMGS	HMG-CoA synthase
IPP	isopentenyl diphosphate
Kan	kanamycin
KIKO	knockin/knockout
LB	Luria broth
MEP	methylerythritol phosphate
MK	methyl ketones
MVA	mevalonate pathway
NADH	nicotinamide adenine dinucleotide
NHEJ	non-homologous end joints
PCR	polymerase chain reaction
PIACE	parallel integration and chromosomal expansion
qPCR	quantitative PCR
RFP	red fluorescent protein

6.4 List of Symbols

μL	microliter
L	liter
μg	microgram
μM	micromolar
μm	micromole
bp	base pairs
mL	milliliter
mg	milligram
mM	millimolar
mAmp	milliampere
h	hour
min	minute
sec	seconds
nm	nanometer
kDa	kilo-Dalton
Km	Michaelis-Menten kinetic constant
kV	kilovolt
rpm	rounds per minute
Vmax	Maximum reaction velocity (Michaelis-Menten kinetic constant)
v/v	volume per volume
^{0}C	celsius

6.5 Supplementary information

6.5.1 Machine Learning for the Prediction of Isopentenol Titers from Gene Copy Numbers

Because there are no available mechanistic models to predict how GOI1−3 copy numbers will affect isopentenol titers, we opted for a statistical approach. The performance of three (namely support vector regressor, random forest regressor, and polynomial regressor) machine learning models was assessed for the prediction of isopentenol titers, given GOI1−3 copy numbers. Each model's fit, cross validation, and error residuals are shown in Supplemental Figure 6.1 and Supplemental Figure 6.2. To evaluate the extent of putative model overfitting, one data point (strain 7P7N) was withheld during model development. The root-mean-square residual prediction errors of the models (lower is better) are 80.5, 58.9, and 34.1 mg/L isopentenol for the support vector regressor, random forest regressor, and polynomial regressor models, respectively.

While these models are qualitatively similar (each suggests that to increase isopentenol titers, GOI1−3 copy numbers should all be increased beyond that present in the training set), the polynomial regressor (Supplemental Figure 6.1) significantly outperforms the others (Supplemental Figure 6.2) in terms of its prediction accuracy. For strain 7P7N (the data point withheld during model development), which achieved 344 mg/L isopentenol, the support vector regressor and random forest regressor do not perform very well, predicting 211.5 and 276.0 mg/L of production, respectively. In comparison, the polynomial regressor has a more accurate rate prediction of 292.4 mg/L isopentenol.
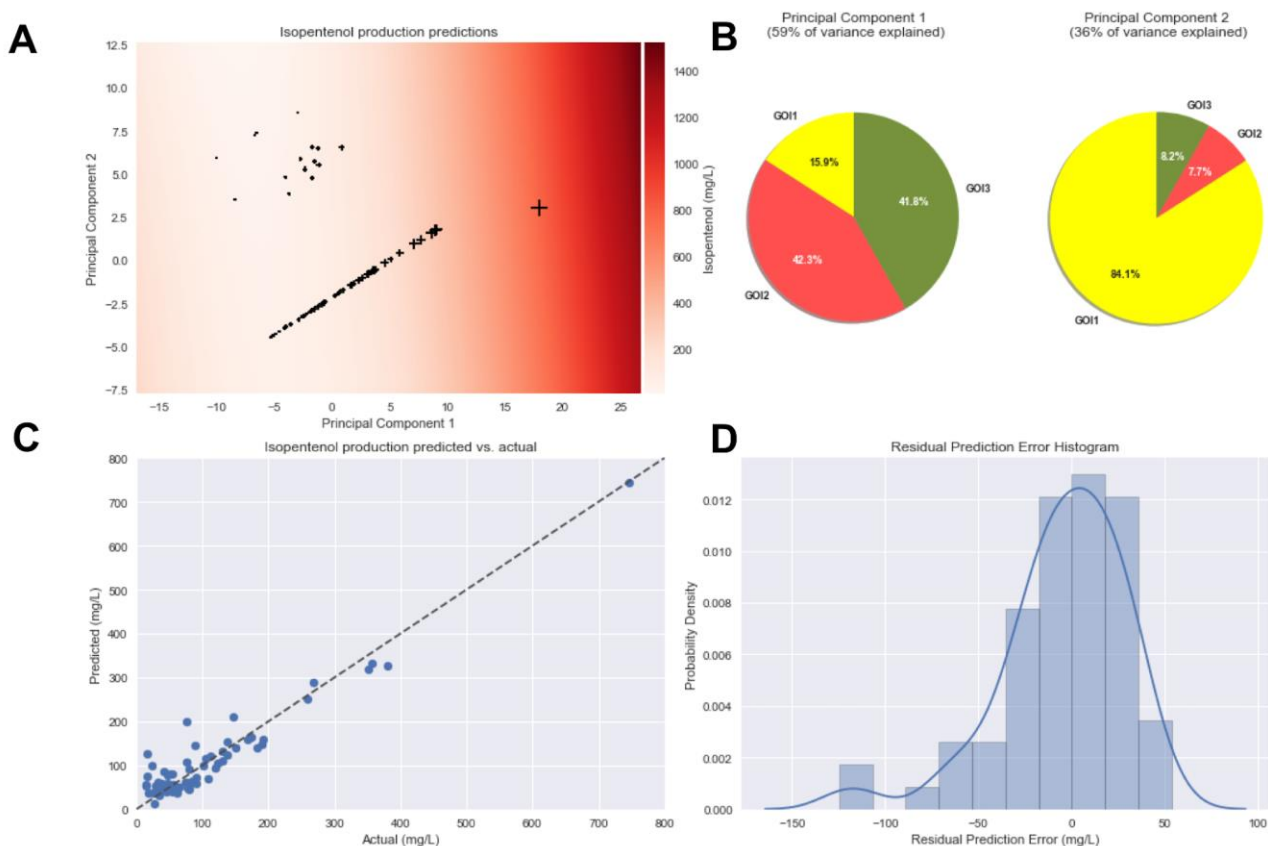
Figure 6.1. Polynomical regressor modeling of isopentenol production data. (**A**) Predicted isopentenol production (darker red indicates higher isopentenol titer) plotted against first two principal components of gene copy number measurements. Cross-hatch marks ("+") correspond to isopentenol/copy number measurements, with larger cross-hatch marks indicating higher isopentenol titers. The gradient towards the darkest red region at right corresponds to increasing copy numbers of all isopentenol pathway genes. (**B**) Decomposition of first two principal components of gene copy number measurements. (**C**) Cross-validated model predictions plotted against actual measurements. Dashed line coincides with flawless predictions. (**D**) The residuals of the prediction errors in (C) illustrated in a histogram. Within one standard deviation, all points have less than 34.1 mg/L isopentenol prediction error.

6.5.2 CRISPR-Cas9 selection protocol

In chapter 4, we used CRISPR-Cas9 to target gene groups/clusters Table 4.5 for deletion from JBEI 3111. Cells transformed with plasmid pKD46-Cas9_sgRNA(p15A) co-expressing Cas9 under a constitutive promoter and λ-red under P$_{BAD}$ (induced with 10 mM arabinose) were grown in LB with ampicillin (Figure 4.4) and made electrocompetent as explained above. We use ~500 bp homology on both arms of the fixing template and electroporated it at the same time as the sgRNA coding pMCC_synCRISPR plasmid using freshly prepared electro-competent cells. Then, cells were plated on LB-agar containing the antibiotics to select for both the pKD46-Cas9_sgRNA(p15A) (Ampicillin) and the plasmid expressing the pMCC_synCRISPR (Chloramphenicol).

Different synCRISPRs in the pMCC vector were ordered from Genscript code for sgRNAs as in *Streptococcus pyogenes* native system. The DNA contained the following from 5' to 3' end: i) SMR20 (20 bp homology with pMCC and pMCK backbone), ii) P$_{Spyog}$ (a constitutive promoter from *Streptococcus pyogenes* driving the transcription of the crRNA), iii) Repeat 1 (*S. pyogenes* repeat sequence (36 bp)), iv) Spacer left (30 bp sequence that guides Cas9 to cut the genome in the reverse strand (i.e. protospacer sequence preceded by a 'CCN' protospacer adjacent motif (PAM)), v) Repeat 2, vi) Spacer right (guides Cas9 to cut in a protospacer in the 5'-3' strand (followed by a 'NGG' PAM)), vii) Repeat 3, viii) *S. pyogenes* terminator sequence, ix) SMR21 (20 bp homology with pMCC backbone (Table 4.2).

Each of this fixing template contains the following: i) R6K ori, ii) 500 bp homology upstream the left spacer cut, iii) 500 bp homology downstream of the right spacer cut. These fixing templates were constructed by amplifying with specific primers, gel purifying and assembling using Gibson assembly protocol. Fixing template were then co-electroporated with its respective

pMCC_synCRISPR coding sgRNA. 50 ng of pMCC vector and 100 ng of purified fixing template was co-transformed onto electro-competent cells expressing the λ-red protein. The fixing template is designed to recombine outside of the protospacer sequences in the *E. coli* genome and therefore provide means for the cells to escape the Cas9 double strand break. Colonies generated from the cells who survived the DSB were further screened to confirm the deletion.

6.5.2 Supplementary figures used in Chapter 3 and 4.

Figure 6.2. Plasmid maps for the plasmids used in chapter 3.
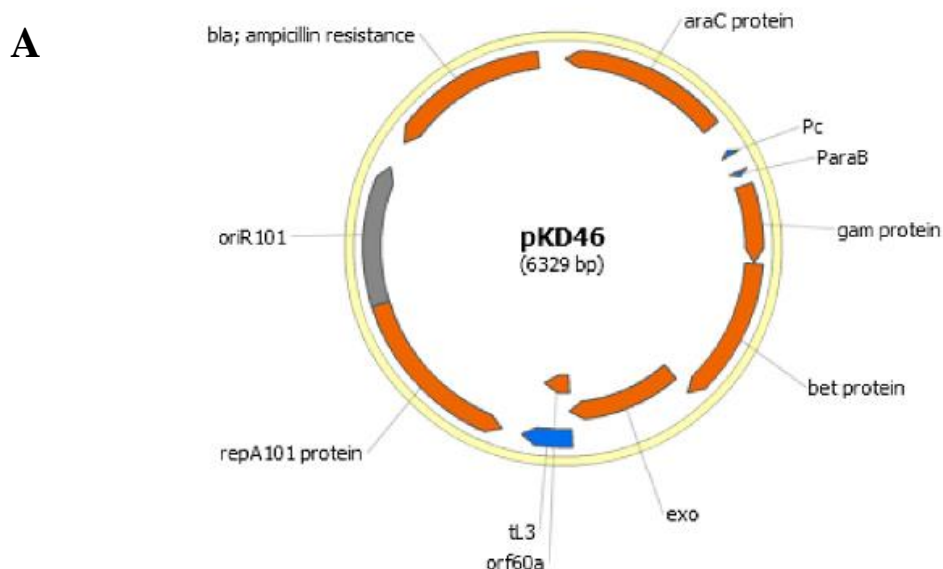
A) Auxiliary heat-sensitive pKD46 plasmid, used to facilitate homologous recombination
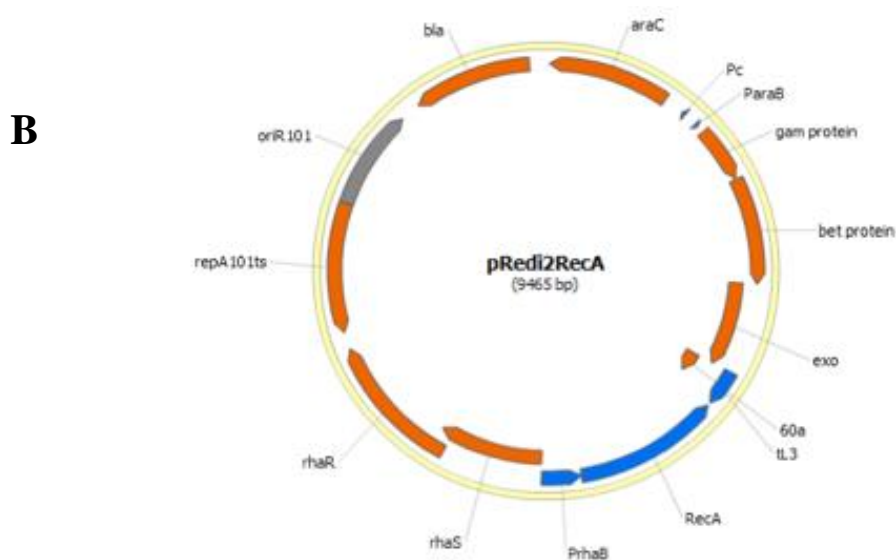
(**B**) Auxiliary heat- sensitive pRedi2RecA plasmid, used to facilitate chromosomal expansion.
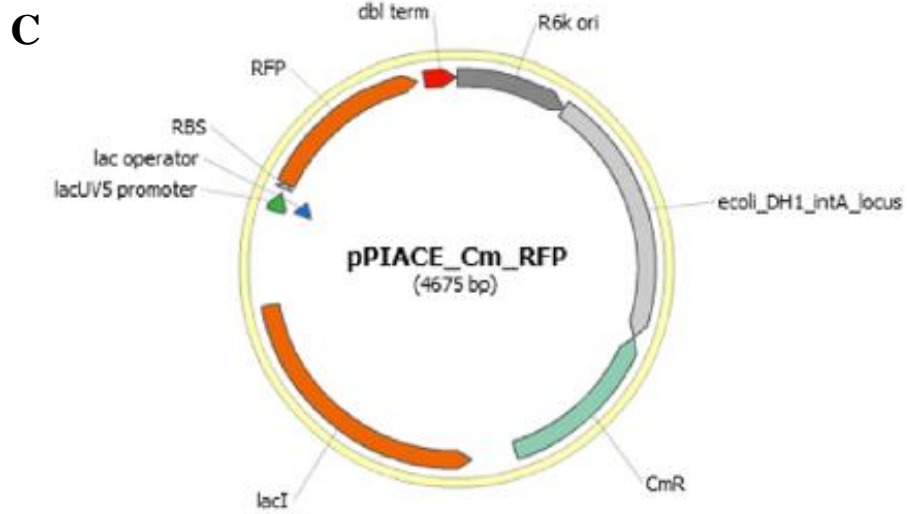
(**C**) pPIACE_Cm_RFP suicide vector, used for the chromosomal integration of *rfp* and the chloramphenicol resistance cassette at the *intA* locus in the *E. coli* DH1 genome.
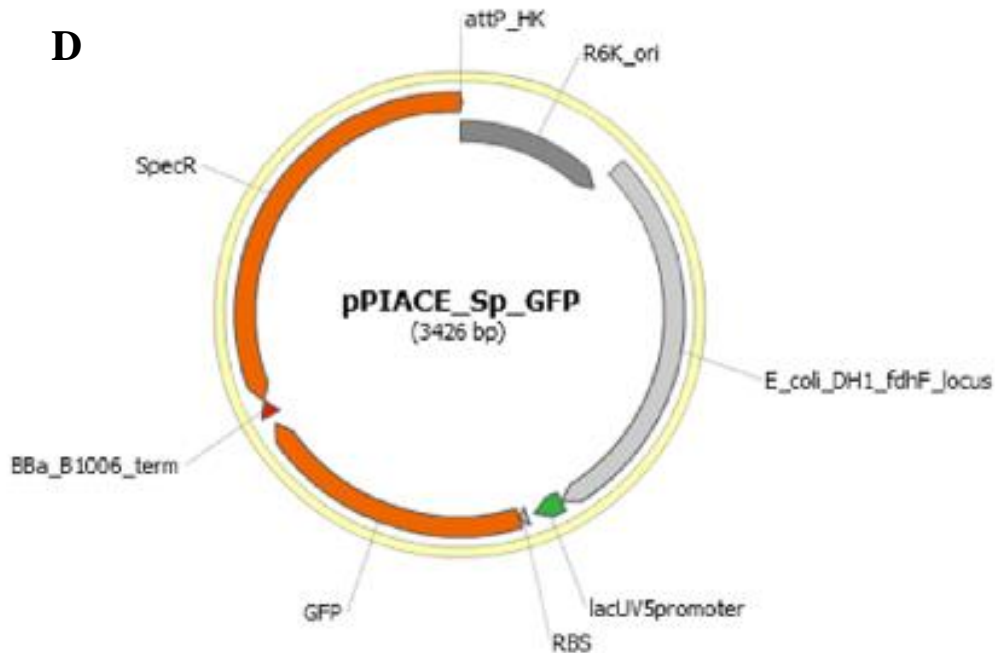
(**D**) pPIACE_Sp_GFP suicide vector, used for the chromosomal integration of *gfp* and the spectinomycin resistance cassette at the *fdhF* locus in the *E. coli* DH1 genome
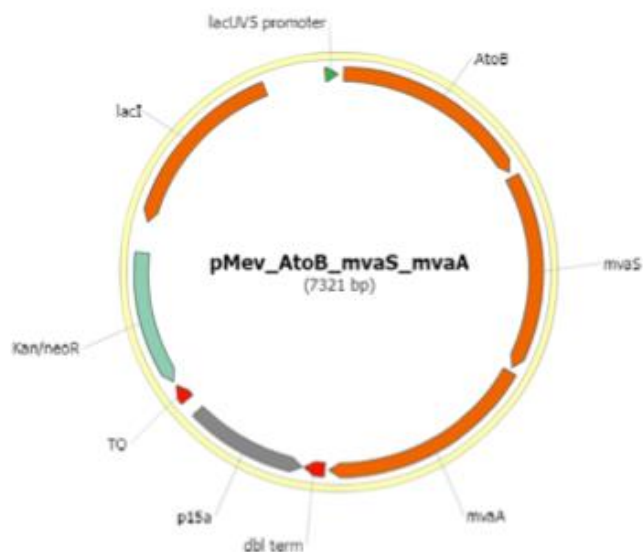
(**E**) pMevTop_Kan plasmid containing *AtoB*, *mvaS*, *mvaA* genes, used as a component of the isopentenol hybrid plasmid/chromosomal system.
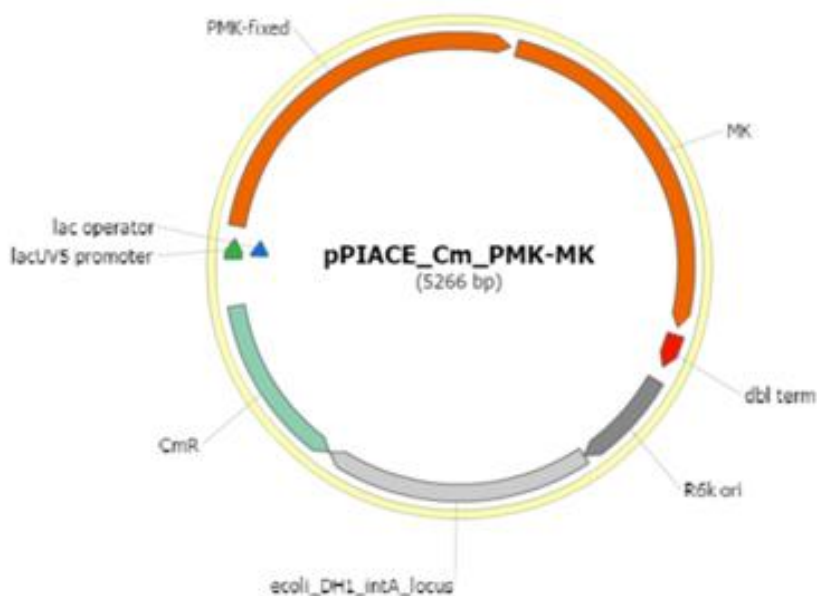
(**F**) pPMK_MK_CmR suicide vector component of the isopentenol hybrid plasmid/chromosomal and chromosomal systems, used for the chromosomal integration of the *PMK* and *MK* pathway genes at the *intA* locus in the *E. coli* DH1 genome.
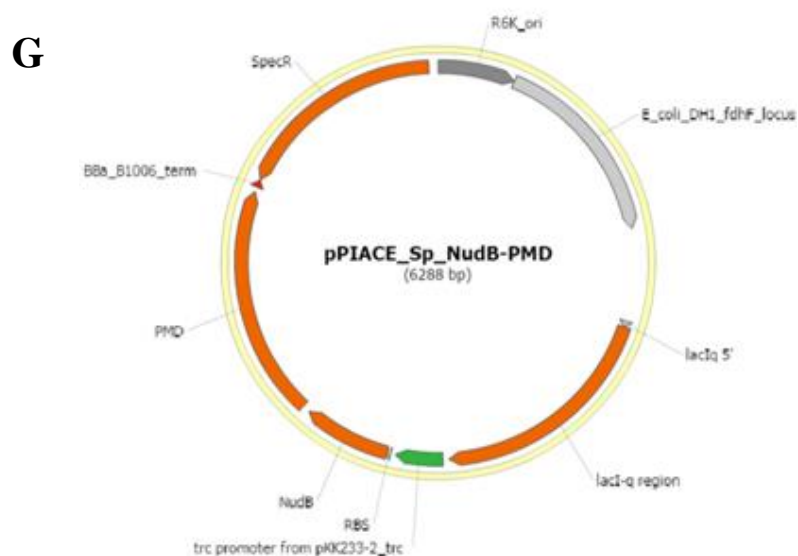
(**G**) pNudB_PMD_SpecR suicide vector component of the isopentenol hybrid plasmid/chromosomal and chromosomal systems, used for the chromosomal integration of the *NudB* and *PMD* pathway genes at the *fdhF* locus in the *E. coli* DH1 genome.
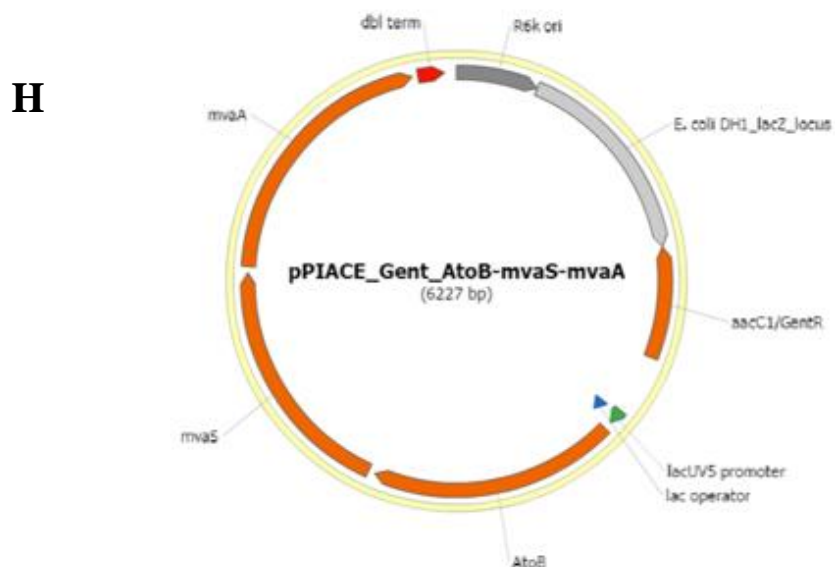
(**H**) pAtoB_mvaS_mvaA_GentR suicide vector component of the isopentenol chromosomal system, used for the chromosomal integration of the *AtoB*, *mvaS*, and *mvaA* pathway genes at the *lacZ* locus in the *E. coli* DH1 genome.

Figure 6.3. Support vector (A, C, E) and random forest (B, D, F) regressor modeling of isopentenol production data for chapter 3.
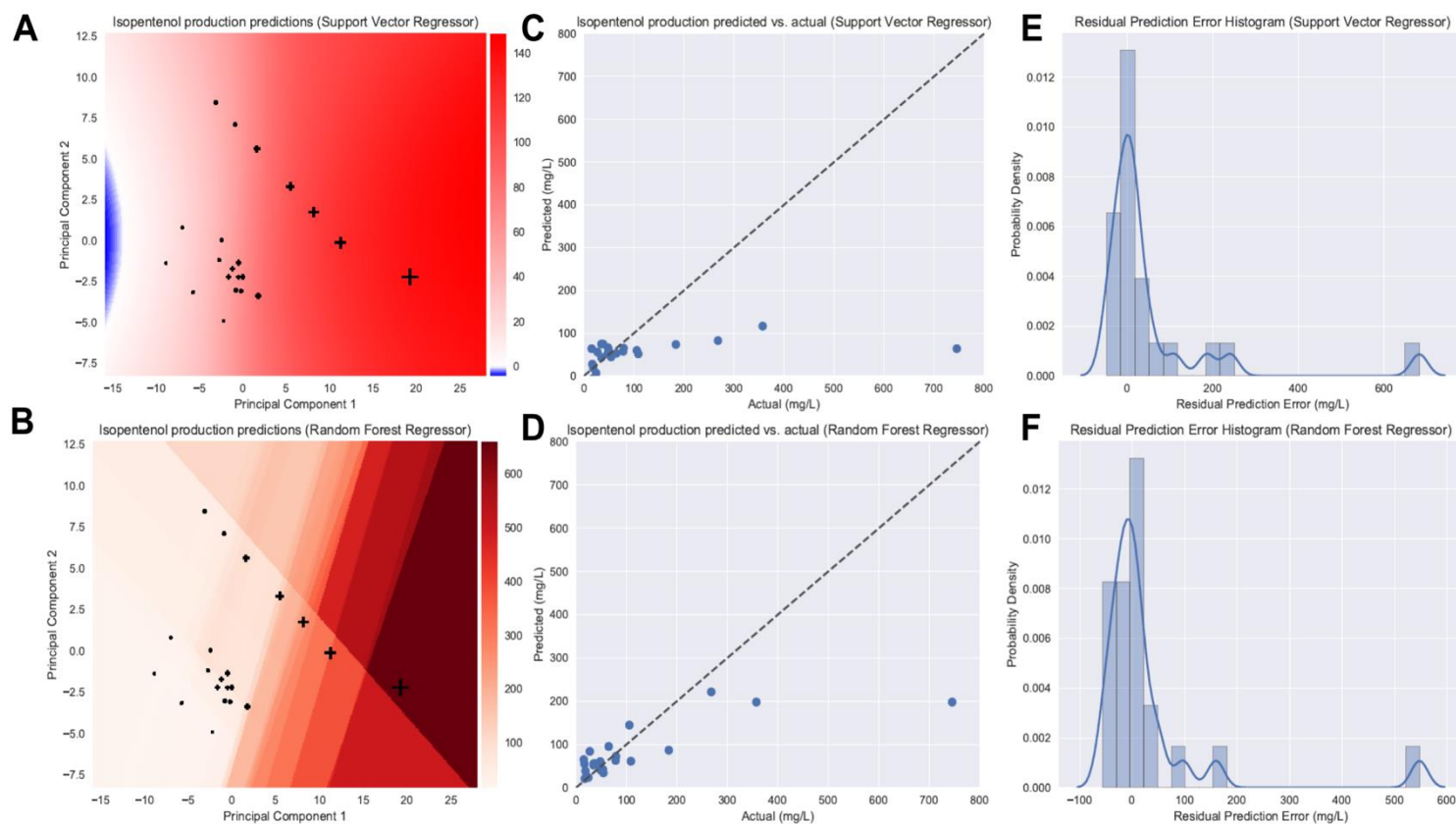
(A, B) Predicted isopentenol production (darker red indicates higher isopentenol titer) plotted against first two principal components of gene copy number measurements. Cross-hatch marks ("+") correspond to isopentenol/copy number measurements, with larger cross-hatch marks indicating higher isopentenol titers. The support vector and random forest regressor models show qualitative agreement despite differences in structure. (C, D) Cross-validated model predictions plotted against actual measurements. Dashed lines coincide with flawless predictions. (E, F) Histograms of prediction error residuals. The support vector and random forest regressors have error residual standard deviations of 142.2 and 116.0 mg/L isopentenol production, respectively.

Figure 6.4. Plasmid maps for the plasmids used in chapter 4.
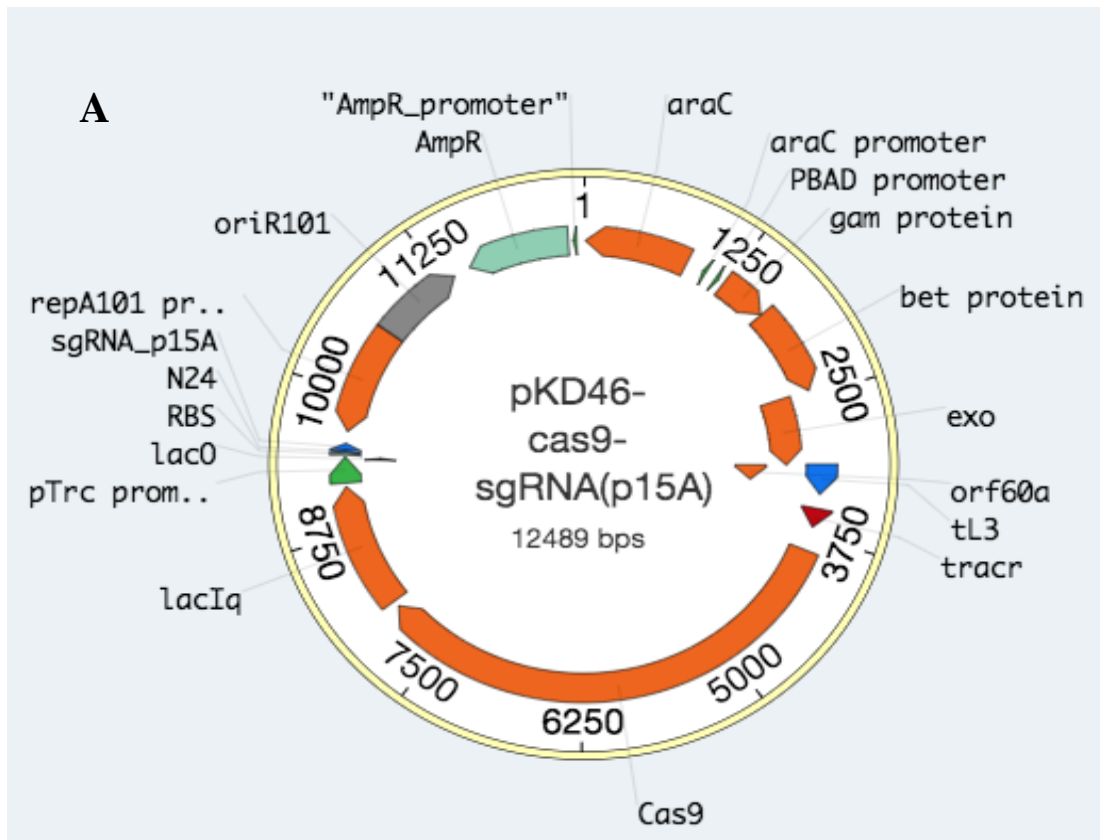
(A) Auxiliary heat-sensitive plasmid pKD46-Cas9-sgRNA(p15A) consists of λ-red proteins under the control of arabinose inducible promoter, Cas9 protein under constitutive promoter to cut the target site in pMCC-synCRISPR, and guide RNA targeting p15A ori of pMCC-synCRISPR under pTrc promoter to facilitate removal of pMCC-synCRISPR post deletion of the corresponding gene.

B
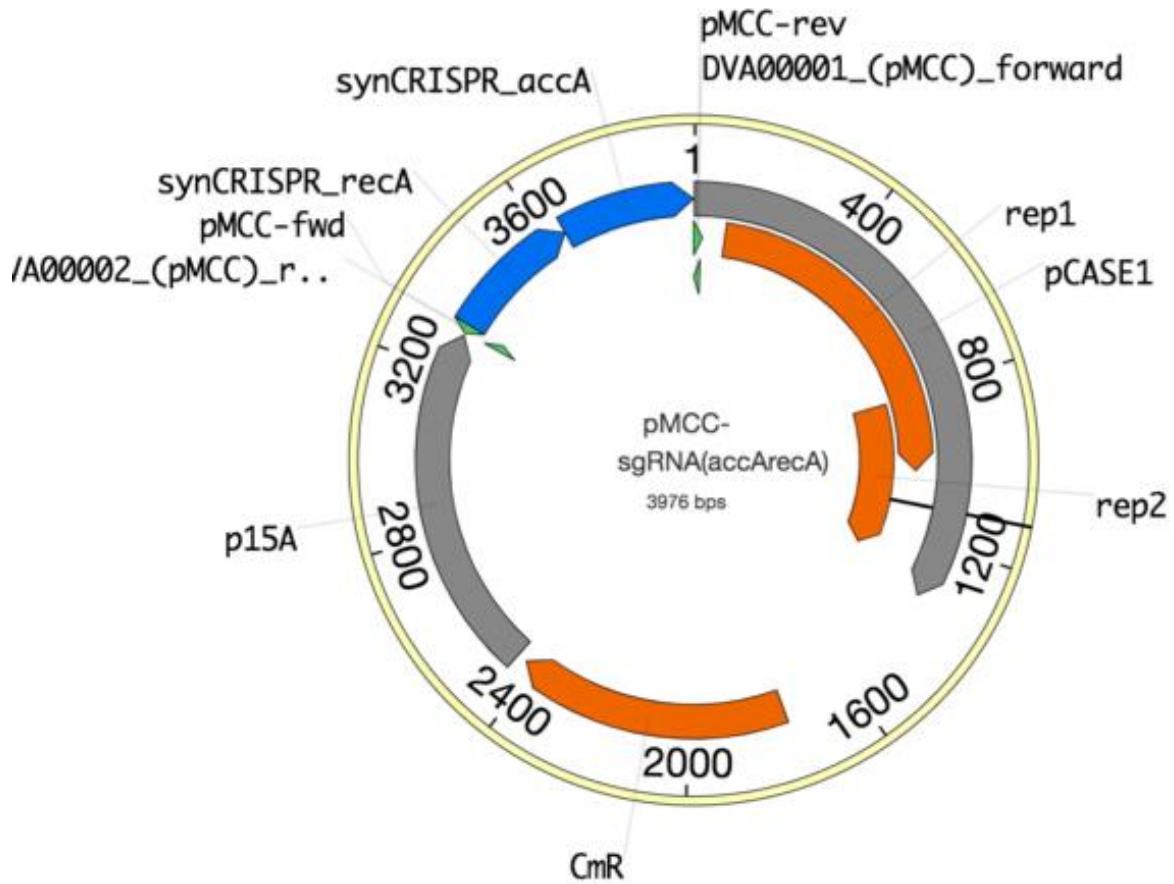
(B) pMCC-SynCRISPR delivers a synCRISPR that contains specific guide
RNA for double cut in the genome for gene deletion.

C
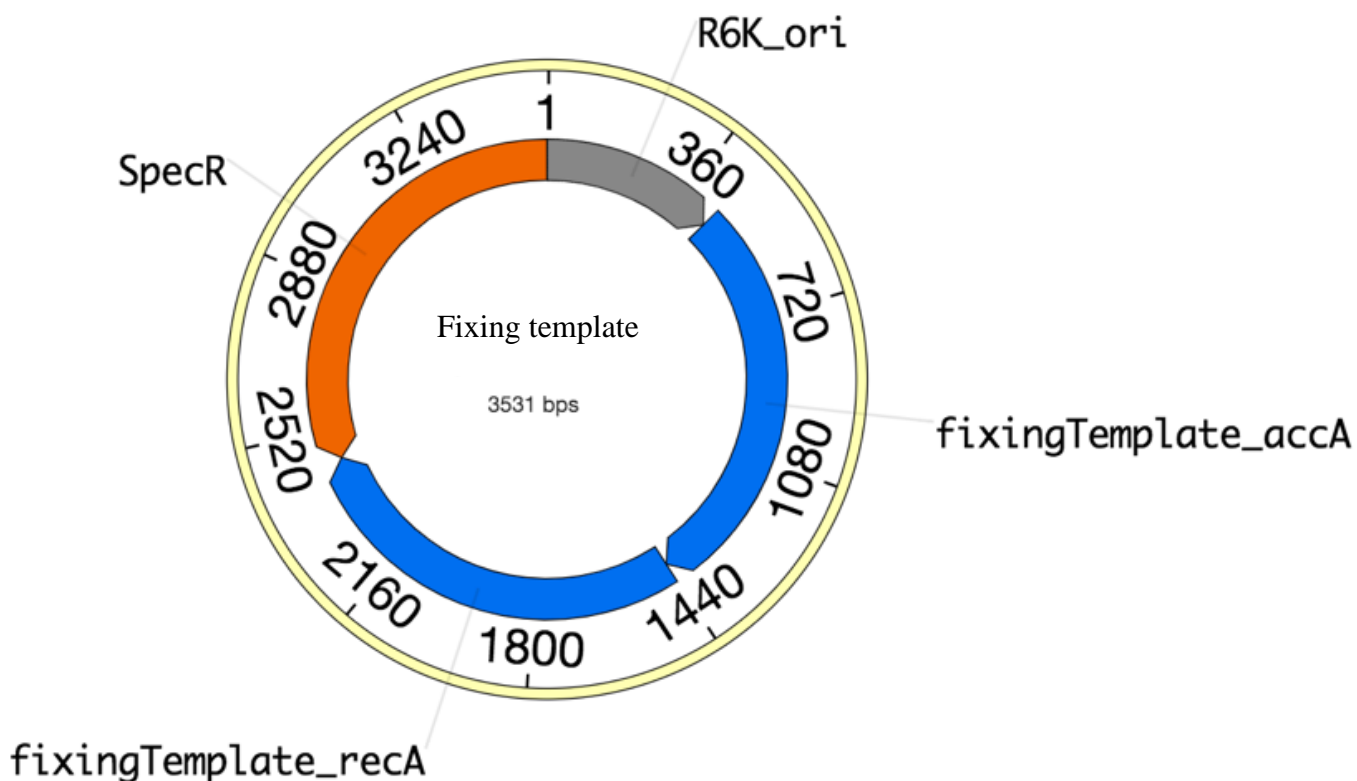
(C) Suicidevector_fixingtemplate delivers a 1000 bp sequence of fixing template containing 500 bp with upstream and downstream homology of the target gene for deletion to repair the double-stranded break induced by Cas9

6.5.3 Supplementary tables used in Chapter 3 and 4.

Table 6.1. List of Plasmids and Strains used in Chapter 3

Plasmid Name	JBEI ID	JBEI PUB ID	Description	Reference
MAGE - 01	JBx_000923	JPUB_010098	Template for amplifying Spectinomycin resistance gene	[186]
pAH143	JDK_p01032	JPUB_010091	Template for amplifying R6K origin of replication	[163]

pBbA5k-RFP	JBp_000015	JPUB_000048	Biobrick vector digested with *EcoRI* and *XhoI* to release *KanR* gene for constructing plasmid pMev_AtoB-mvaS-mvaA.	[187]
pBbE5c-RFP	JBp_000011	JPUB_000044	Template for amplifying chloramphenicol resistance gene and RFP for the plasmid pPIACE_Cm_RFP	[187]
pCM62-Prha-T7SL-RFP	JBx_013711	JPUB_010104	Template for amplifying Tet 5 prime and 3 prime DNA parts for the plasmid pCIChE_Spec_GFP	Unpublished work
pKD46	JBx_000276	JPUB_010097	Helper plasmid for λ-red homologous recombination	[188]
pKWG0012	JBx_012998	JPUB_010102	Template for amplifying genes *PMK* and *MK* for the plasmid pPIACE_Cm_PMK-MK	Unpublished work
pMCS-4	JBx_000254	JPUB_010093	Template for amplifying gentamycin resistance gene for the plasmid pPIACE_Gent_AtoB-mvaS-mvaA	[189]
pMTSA	JBx_001365	JPUB_010100	Template for amplifying genes *AtoB, mvaS,* and *mvaA* for plasmid pPIACE_Gent_AtoB-mvaS-mvaA and for constructing plasmid pMev_AtoB-mvaS-mvaA by digesting with EcoRI and XhoI	Unpublished work
pREDI	JBx_000256	JPUB_010095	Template for constructing plasmid pRedi2RecA	[190]
pTrc99A-NudB-PMD	JBx_025920	JPUB_004938	Template for amplifying genes *NudB* and *PMD* for the plasmid pPIACE_Sp_NudB-PMD	[150]
pCIChE_Spec_GFP	JBx_022134	JPUB_010108	Template for amplifying Spectinomycin resistance gene and *GFP* for the plasmid pPIACE_Sp_GFP	this study
pMev_AtoB-mvaS-mvaA	JBx_047317	JPUB_010342	Contains *AtoB, mvaS,* and *mvaA* genes from isopentenol pathway under lavUV5 promoter in a p15A ori vector for expressing top mevalonate pathway genes in hybrid plasmid/chromosomal system	this study
pPIACE_Cm_PMK-MK	JBx_030452	JPUB_010116	Contains *PMK and MK* genes from isopentenol pathway under lavUV5 promoter in a suicide vector with R6K origin of replication for homologous recombination at *intA* locus into DH1 *E. coli* cells	this study
pPIACE_Cm_RFP	JBx_026152	JPUB_010110	Contains *rfp* gene under lacUV5 promoter in a suicide vector with R6K origin of replication for homologous recombination at *intA* locus into DH1 *E. coli* cells	this study
pPIACE_Gent_AtoB-mvaS-mvaA	JBx_063771	JPUB_010117	Contains *AtoB, mvaS,* and *mvaA* genes from isopentenol pathway under lavUV5 promoter in a suicide vector with R6K origin of replication for homologous recombination at *lacZ* locus into DH1 *E. coli* cells	this study
pPIACE_Sp_GFP	JBx_026154	JPUB_010112	Contains *GFP* gene under lacUV5 promoter in a suicide vector with R6K origin of replication for homologous recombination at *fdhF* locus into DH1 *E. coli* cells	this study
pPIACE_Sp_NudB-PMD	JBx_030450	JPUB_010114	Contains *NudB* and *PMD* genes from isopentenol pathway under pTrc promoter in a suicide vector with R6K origin of replication for homologous recombination at *fdhF* locus into DH1 *E. coli* cells	this study
pRedi2RecA	JBx_020325	JPUB_010106	Contains λ-red proteins and RecA protein for homologous recombination and gene expansion	this study

Strain Name	JBEI ID	JBEI PUB ID	Description	Reference
DH1	-	-	*E. coli K-12; F-endA1 recA1 gyrA96 thi-1 glnV44 relA1 hsdR17(rK- mK+) λ−*	invitrogen
DH10B	-	-	*F-endA1 recA1 galE15 galK16 nupGrpsL lacX74Φ80lacZM15araD139 (ara,leu)7697mcrA (mrr-hsdRMS-mcrBC) λ−*	invitrogen
pMevTop_Kan_DH1	JBx_083367	JPUB_010246	*E. coli* DH1 strain containing plasmid providing top mevalonate pathway genes used for homologous recombination of GOI2 and GOI3 at *intA* and *fdhF* loci	this study
1R1G_pRedi2RecA_DH1	JBx_083368	JPUB_010247	*E. coli* DH1 strain containing *rfp* and *gfp* integrated at *intA* and *fdhF* loci and plasmid pRedi2RecA transformed for chromosomal gene expansion	this study
1P1N_pRedi2RecA_DH1	JBx_083369	JPUB_010248	*E. coli* DH1 strain containing *PMK-MK* and *NudB-PMD* gene cassettes integrated at *intA* and *fdhF* loci and plasmid pRedi2RecA transformed for chromosomal gene expansion	this study

1A1P1N_pRedi2RecA_DH1	JBx_083370	JPUB_010249	*E. coli* DH1 strain containing *AtoB-mvaS-mvaA*, *PMK-MK*, and *NudB-PMD* gene cassettes integrated at *lacZ*, *intA*, and *fdhF* loci and plasmid pRedi2RecA transformed for chromosomal gene expansion	this study
1R1G	JBx_078844	JPUB_010118	*E. coli* DH1 strain containing approximately one copy of each RFP and GFP in the chromosome at loci intA and fdhF respectively	this study
1R2G	JBx_078845	JPUB_010119	*E. coli* DH1 strain containing approximately one copy of RFP and 2 copies GFP in the chromosome at loci intA and fdhF respectively	this study
1R3G	JBx_078846	JPUB_010120	*E. coli* DH1 strain containing approximately one copy of RFP and five copies of GFP in the chromosome at loci intA and fdhF respectively	this study
1R4G	JBx_078847	JPUB_010121	*E. coli* DH1 strain containing approximately one copy of RFP and nine copies of GFP in the chromosome at loci intA and fdhF respectively	this study
1R5G	JBx_078848	JPUB_010122	*E. coli* DH1 strain containing approximately one copy of RFP and seventeen copies of GFP in the chromosome at loci intA and fdhF respectively	this study
1R6G	JBx_078849	JPUB_010123	*E. coli* DH1 strain containing approximately one copy of RFP and thirty-two copies GFP in the chromosome at loci intA and fdhF respectively	this study
1R7G	JBx_078850	JPUB_010124	*E. coli* DH1 strain containing approximately one copy of RFP and fifty copies of GFP in the chromosome at loci intA and fdhF respectively	this study
1R8G	JBx_078851	JPUB_010125	*E. coli* DH1 strain containing approximately one copy of RFP and sixty-six copies of GFP in the chromosome at loci intA and fdhF respectively	this study
2R1G	JBx_078852	JPUB_010126	*E. coli* DH1 strain containing approximately two copies of RFP and one copy of GFP in the chromosome at loci intA and fdhF respectively	this study
2R2G	JBx_078853	JPUB_010127	*E. coli* DH1 strain containing approximately two copies of each RFP and GFP in the chromosome at loci intA and fdhF respectively	this study
2R3G	JBx_078854	JPUB_010128	*E. coli* DH1 strain containing approximately two copies of RFP and five copies of GFP in the chromosome at loci intA and fdhF respectively	this study
2R4G	JBx_078855	JPUB_010129	*E. coli* DH1 strain containing approximately two copies of RFP and nine copies of GFP in the chromosome at loci intA and fdhF respectively	this study
2R5G	JBx_078856	JPUB_010130	*E. coli* DH1 strain containing approximately two copies of RFP and seventeen copies of GFP in the chromosome at loci intA and fdhF respectively	this study
2R6G	JBx_078857	JPUB_010131	*E. coli* DH1 strain containing approximately two copies of RFP and thirty-two copies of GFP in the chromosome at loci intA and fdhF respectively	this study
2R7G	JBx_078858	JPUB_010132	*E. coli* DH1 strain containing approximately two copies of RFP and fifty-six copies of GFP in the chromosome at loci intA and fdhF respectively	this study
2R8G	JBx_078859	JPUB_010133	*E. coli* DH1 strain containing approximately two copies of RFP and seventy-seven copies of GFP in the chromosome at loci intA and fdhF respectively	this study
3R1G	JBx_078860	JPUB_010134	*E. coli* DH1 strain containing approximately five copies of RFP and one copy of GFP in the chromosome at loci intA and fdhF respectively	this study
3R2G	JBx_078861	JPUB_010135	*E. coli* DH1 strain containing approximately five copies of RFP and two copies of GFP in the chromosome at loci intA and fdhF respectively	this study
3R3G	JBx_078862	JPUB_010136	*E. coli* DH1 strain containing approximately five copies of each RFP and GFP in the chromosome at loci intA and fdhF respectively	this study
3R4G	JBx_078863	JPUB_010137	*E. coli* DH1 strain containing approximately five copies of RFP and nine copies GFP in the chromosome at loci intA and fdhF respectively	this study

3R5G	JBx_078864	JPUB_010138	*E. coli* DH1 strain containing approximately five copies of RFP and sixteen copies GFP in the chromosome at loci intA and fdhF respectively	this study
3R6G	JBx_078865	JPUB_010139	*E. coli* DH1 strain containing approximately five copies of RFP and thirty one copies GFP in the chromosome at loci intA and fdhF respectively	this study
3R7G	JBx_078866	JPUB_010140	*E. coli* DH1 strain containing approximately five copies of RFP and fifty one copies GFP in the chromosome at loci intA and fdhF respectively	this study
3R8G	JBx_078867	JPUB_010141	*E. coli* DH1 strain containing approximately five copies of RFP and sixty three copies GFP in the chromosome at loci intA and fdhF respectively	this study
4R1G	JBx_078868	JPUB_010142	*E. coli* DH1 strain containing approximately nine copies of RFP and one copy of GFP in the chromosome at loci intA and fdhF respectively	this study
4R2G	JBx_078869	JPUB_010143	*E. coli* DH1 strain containing approximately nine copies of RFP and two copies of GFP in the chromosome at loci intA and fdhF respectively	this study
4R3G	JBx_078870	JPUB_010144	*E. coli* DH1 strain containing approximately nine copies of RFP and five copies of GFP in the chromosome at loci intA and fdhF respectively	this study
4R4G	JBx_078871	JPUB_010145	*E. coli* DH1 strain containing approximately nine copies of RFP and ten copies of GFP in the chromosome at loci intA and fdhF respectively	this study
4R5G	JBx_078872	JPUB_010146	*E. coli* DH1 strain containing approximately nine copies of RFP and twenty copies of GFP in the chromosome at loci intA and fdhF respectively	this study
4R6G	JBx_078873	JPUB_010147	*E. coli* DH1 strain containing approximately nine copies of RFP and thirty-five copies of GFP in the chromosome at loci intA and fdhF respectively	this study
4R7G	JBx_078874	JPUB_010148	*E. coli* DH1 strain containing approximately nine copies of RFP and fifty-eight copies of GFP in the chromosome at loci intA and fdhF respectively	this study
4R8G	JBx_078875	JPUB_010149	*E. coli* DH1 strain containing approximately nine copies of RFP and sixty-five copies of GFP in the chromosome at loci intA and fdhF respectively	this study
5R1G	JBx_078876	JPUB_010150	*E. coli* DH1 strain containing approximately seventeen copies of RFP and one copy of GFP in the chromosome at loci intA and fdhF respectively	this study
5R2G	JBx_078877	JPUB_010151	*E. coli* DH1 strain containing approximately seventeen copies of RFP and two copies of GFP in the chromosome at loci intA and fdhF respectively	this study
5R3G	JBx_078878	JPUB_010152	*E. coli* DH1 strain containing approximately seventeen copies of RFP and five copies of GFP in the chromosome at loci intA and fdhF respectively	this study
5R4G	JBx_078879	JPUB_010153	*E. coli* DH1 strain containing approximately seventeen copies of RFP and ten copies of GFP in the chromosome at loci intA and fdhF respectively	this study
5R5G	JBx_078880	JPUB_010154	*E. coli* DH1 strain containing approximately seventeen copies of RFP and twenty copies of GFP in the chromosome at loci intA and fdhF respectively	this study
5R6G	JBx_078881	JPUB_010155	*E. coli* DH1 strain containing approximately seventeen copies of RFP and thirty-five copies of GFP in the chromosome at loci intA and fdhF respectively	this study
5R7G	JBx_078882	JPUB_010156	*E. coli* DH1 strain containing approximately seventeen copies of RFP and fifty-five copies of GFP in the chromosome at loci intA and fdhF respectively	this study
5R8G	JBx_078883	JPUB_010157	*E. coli* DH1 strain containing approximately seventeen copies of RFP and sixty-eight copies of GFP in the chromosome at loci intA and fdhF respectively	this study
6R1G	JBx_078884	JPUB_010158	*E. coli* DH1 strain containing approximately thirty-two copies of RFP and one copy of GFP in the chromosome at loci intA and fdhF respectively	this study

6R2G	JBx_078885	JPUB_010159	*E. coli* DH1 strain containing approximately thirty-two copies of RFP and two copies of GFP in the chromosome at loci intA and fdhF respectively	this study
6R3G	JBx_078886	JPUB_010160	*E. coli* DH1 strain containing approximately thirty-two copies of RFP and six copies of GFP in the chromosome at loci intA and fdhF respectively	this study
6R4G	JBx_078887	JPUB_010161	*E. coli* DH1 strain containing approximately thirty-two copies of RFP and thirteen copies of GFP in the chromosome at loci intA and fdhF respectively	this study
6R5G	JBx_078888	JPUB_010162	*E. coli* DH1 strain containing approximately thirty-two copies of RFP and twenty-three copies of GFP in the chromosome at loci intA and fdhF respectively	this study
6R6G	JBx_078889	JPUB_010163	*E. coli* DH1 strain containing approximately thirty-two copies of RFP and thirty-eight copies of GFP in the chromosome at loci intA and fdhF respectively	this study
6R7G	JBx_078890	JPUB_010164	*E. coli* DH1 strain containing approximately thirty-two copies of RFP and sixty-nine copies of GFP in the chromosome at loci intA and fdhF respectively	this study
6R8G	JBx_078891	JPUB_010165	*E. coli* DH1 strain containing approximately thirty-two copies of RFP and seventy-four copies of GFP in the chromosome at loci intA and fdhF respectively	this study
7R1G	JBx_078892	JPUB_010166	*E. coli* DH1 strain containing approximately fifty copies of RFP and one copy of GFP in the chromosome at loci intA and fdhF respectively	this study
7R2G	JBx_078893	JPUB_010167	*E. coli* DH1 strain containing approximately fifty copies of RFP and two copies of GFP in the chromosome at loci intA and fdhF respectively	this study
7R3G	JBx_078894	JPUB_010168	*E. coli* DH1 strain containing approximately fifty copies of RFP and five copies of GFP in the chromosome at loci intA and fdhF respectively	this study
7R4G	JBx_078895	JPUB_010169	*E. coli* DH1 strain containing approximately fifty copies of RFP and ten copies of GFP in the chromosome at loci intA and fdhF respectively	this study
7R5G	JBx_078896	JPUB_010170	*E. coli* DH1 strain containing approximately fifty copies of RFP and twenty-four copies of GFP in the chromosome at loci intA and fdhF respectively	this study
7R6G	JBx_078897	JPUB_010171	*E. coli* DH1 strain containing approximately fifty copies of RFP and thirty-eight copies of GFP in the chromosome at loci intA and fdhF respectively	this study
7R7G	JBx_078898	JPUB_010172	*E. coli* DH1 strain containing approximately fifty copies of RFP and seventy copies of GFP in the chromosome at loci intA and fdhF respectively	this study
7R8G	JBx_078899	JPUB_010173	*E. coli* DH1 strain containing approximately sixty-six copies of RFP and one copy of GFP in the chromosome at loci intA and fdhF respectively	this study
8R1G	JBx_078900	JPUB_010174	*E. coli* DH1 strain containing approximately sixty-six copies of RFP and one copy of GFP in the chromosome at loci intA and fdhF respectively	this study
8R2G	JBx_078901	JPUB_010175	*E. coli* DH1 strain containing approximately sixty-six copies of RFP and two copies of GFP in the chromosome at loci intA and fdhF respectively	this study
8R3G	JBx_078902	JPUB_010176	*E. coli* DH1 strain containing approximately sixty-six copies of RFP and seven copies of GFP in the chromosome at loci intA and fdhF respectively	this study
8R4G	JBx_078903	JPUB_010177	*E. coli* DH1 strain containing approximately sixty-six copies of RFP and fifteen copies of GFP in the chromosome at loci intA and fdhF respectively	this study
8R5G	JBx_078904	JPUB_010178	*E. coli* DH1 strain containing approximately sixty-six copies of RFP and thirty-three copies of GFP in the chromosome at loci intA and fdhF respectively	this study
8R6G	JBx_078905	JPUB_010179	*E. coli* DH1 strain containing approximately sixty-six copies of RFP and thirty-eight copies of GFP in the chromosome at loci intA and fdhF respectively	this study

8R7G	JBx_078906	JPUB_010180	*E. coli* DH1 strain containing approximately sixty-six copies of RFP and seventy copies of GFP in the chromosome at loci intA and fdhF respectively	this study
8R8G	JBx_078907	JPUB_010181	*E. coli* DH1 strain containing approximately sixty-six copies of RFP and eighty copies of GFP in the chromosome at loci intA and fdhF respectively	this study
1P1N	JBx_078908	JPUB_010182	*E. coli* DH1 strain containing approximately one copy of each PMK-MK and NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
1P2N	JBx_078909	JPUB_010183	*E. coli* DH1 strain containing approximately one copy of PMK-MK and two copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
1P3N	JBx_078910	JPUB_010184	*E. coli* DH1 strain containing approximately one copy of PMK-MK and three copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
1P4N	JBx_078911	JPUB_010185	*E. coli* DH1 strain containing approximately one copy PMK-MK and three copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
1P5N	JBx_078912	JPUB_010186	*E. coli* DH1 strain containing approximately one copy of PMK-MK and three copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
1P6N	JBx_078913	JPUB_010187	*E. coli* DH1 strain containing approximately one copy of PMK-MK and five copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
1P7N	JBx_078914	JPUB_010188	*E. coli* DH1 strain containing approximately one copy of PMK-MK and six copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
2P1N	JBx_078915	JPUB_010189	*E. coli* DH1 strain containing approximately two copies of PMK-MK and one copy of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
2P2N	JBx_078916	JPUB_010190	*E. coli* DH1 strain containing approximately two copies of each PMK-MK and NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
2P3N	JBx_078917	JPUB_010191	*E. coli* DH1 strain containing approximately two copies of PMK-MK and three copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
2P4N	JBx_078918	JPUB_010192	*E. coli* DH1 strain containing approximately two copies of PMK-MK and four copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
2P5N	JBx_078919	JPUB_010193	*E. coli* DH1 strain containing approximately two copies of PMK-MK and six copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
2P6N	JBx_078920	JPUB_010194	*E. coli* DH1 strain containing approximately two copies of PMK-MK and seven copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
2P7N	JBx_078921	JPUB_010195	*E. coli* DH1 strain containing approximately two copies of PMK-MK and seven copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
3P1N	JBx_078922	JPUB_010196	*E. coli* DH1 strain containing approximately four copies of PMK-MK and one copy of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
3P2N	JBx_078923	JPUB_010197	*E. coli* DH1 strain containing approximately four copies of PMK-MK and two copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
3P3N	JBx_078924	JPUB_010198	*E. coli* DH1 strain containing approximately four copies of PMK-MK and three copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study

3P4N	JBx_078925	JPUB_010199	*E. coli* DH1 strain containing approximately four copies of PMK-MK and five copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
3P5N	JBx_078926	JPUB_010200	*E. coli* DH1 strain containing approximately four copies of PMK-MK and seven copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
3P6N	JBx_078927	JPUB_010201	*E. coli* DH1 strain containing approximately four copies of PMK-MK and eight copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
3P7N	JBx_078928	JPUB_010202	*E. coli* DH1 strain containing approximately four copies of PMK-MK and ten copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
4P1N	JBx_078929	JPUB_010203	*E. coli* DH1 strain containing approximately six copies of PMK-MK and one copy of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
4P2N	JBx_078930	JPUB_010204	*E. coli* DH1 strain containing approximately six copies of PMK-MK and three copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
4P3N	JBx_078931	JPUB_010205	*E. coli* DH1 strain containing approximately six copies of PMK-MK and five copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
4P4N	JBx_078932	JPUB_010206	*E. coli* DH1 strain containing approximately six copies of PMK-MK and six copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
4P5N	JBx_078933	JPUB_010207	*E. coli* DH1 strain containing approximately six copies of PMK-MK and eight copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
4P6N	JBx_078934	JPUB_010208	*E. coli* DH1 strain containing approximately six copies of PMK-MK and nine copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
4P7N	JBx_078935	JPUB_010209	*E. coli* DH1 strain containing approximately six copies of PMK-MK and ten copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
5P1N	JBx_078936	JPUB_010210	*E. coli* DH1 strain containing approximately eight copies of PMK-MK and one copy of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
5P2N	JBx_078937	JPUB_010211	*E. coli* DH1 strain containing approximately eight copies of PMK-MK and two copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
5P3N	JBx_078938	JPUB_010212	*E. coli* DH1 strain containing approximately eight copies of PMK-MK and two copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
5P4N	JBx_078939	JPUB_010213	*E. coli* DH1 strain containing approximately eight copies of PMK-MK and five copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
5P5N	JBx_078940	JPUB_010214	*E. coli* DH1 strain containing approximately eight copies of PMK-MK and eight copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
5P6N	JBx_078941	JPUB_010215	*E. coli* DH1 strain containing approximately eight copies of PMK-MK and ten copies of NudB-PMD gene	this study

			cassettes in the chromosome at loci intA and fdhF respectively	
5P7N	JBx_078942	JPUB_010216	*E. coli* DH1 strain containing approximately eight copies of PMK-MK and eleven copy of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
6P1N	JBx_078943	JPUB_010217	*E. coli* DH1 strain containing approximately ten copies of PMK-MK and one copy of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
6P2N	JBx_078944	JPUB_010218	*E. coli* DH1 strain containing approximately ten copies of PMK-MK and two copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
6P3N	JBx_078945	JPUB_010219	*E. coli* DH1 strain containing approximately ten copies of PMK-MK and five copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
6P4N	JBx_078946	JPUB_010220	*E. coli* DH1 strain containing approximately ten copies of PMK-MK and seven copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
6P5N	JBx_078947	JPUB_010221	*E. coli* DH1 strain containing approximately ten copies of PMK-MK and ten copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
6P6N	JBx_078948	JPUB_010222	*E. coli* DH1 strain containing approximately ten copies of PMK-MK and eleven copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
6P7N	JBx_078949	JPUB_010223	*E. coli* DH1 strain containing approximately ten copies of PMK-MK and thirteen copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
7P1N	JBx_078950	JPUB_010224	*E. coli* DH1 strain containing approximately twelve copies of PMK-MK and one copy of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
7P2N	JBx_078951	JPUB_010225	*E. coli* DH1 strain containing approximately twelve copies of PMK-MK and two copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
7P3N	JBx_078952	JPUB_010226	*E. coli* DH1 strain containing approximately twelve copies of PMK-MK and five copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
7P4N	JBx_078953	JPUB_010227	*E. coli* DH1 strain containing approximately twelve copies of PMK-MK and two copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
7P5N	JBx_078954	JPUB_010228	*E. coli* DH1 strain containing approximately twelve copies of PMK-MK and eight copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
7P6N	JBx_078955	JPUB_010229	*E. coli* DH1 strain containing approximately twelve copies of PMK-MK and eleven copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
7P7N	JBx_078956	JPUB_010230	*E. coli* DH1 strain containing approximately twelve copies of PMK-MK and thirteen copies of NudB-PMD gene cassettes in the chromosome at loci intA and fdhF respectively	this study
1A1P1N	JBx_078957	JPUB_010231	*E. coli* DH1 strain containing approximately one copy of each AtoB-mvaS-mvaA, PMK-MK and NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study

1A1P5N	JBx_078958	JPUB_010232	*E. coli* DH1 strain containing approximately one copy of AtoB-mvaS-mvaA, one copy of PMK-MK and six copies of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study
1A5P1N	JBx_078959	JPUB_010233	*E. coli* DH1 strain containing approximately one copy of AtoB-mvaS-mvaA, six copies of PMK-MK and one copy of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study
5A1P1N	JBx_078960	JPUB_010234	*E. coli* DH1 strain containing approximately four copies of AtoB-mvaS-mvaA, one copy of PMK-MK and one copy of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study
3A3P3N	JBx_078961	JPUB_010235	*E. coli* DH1 strain containing approximately four copies of AtoB-mvaS-mvaA, six copies of PMK-MK and six copies of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study
1A5P5N	JBx_078962	JPUB_010236	*E. coli* DH1 strain containing approximately one copy of AtoB-mvaS-mvaA, six copies of PMK-MK and seven copies of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study
5A1P5N	JBx_078963	JPUB_010237	*E. coli* DH1 strain containing approximately four copies of AtoB-mvaS-mvaA, one copy of PMK-MK and seven copies of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study
5A5P1N	JBx_078964	JPUB_010238	*E. coli* DH1 strain containing approximately five copies of AtoB-mvaS-mvaA, seven copies of PMK-MK and one copy of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study
3A4P5N	JBx_078965	JPUB_010239	*E. coli* DH1 strain containing approximately four copies of AtoB-mvaS-mvaA, seven copies of PMK-MK and seven copies of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study
4A3P5N	JBx_078966	JPUB_010240	*E. coli* DH1 strain containing approximately five copies of AtoB-mvaS-mvaA, five copies of PMK-MK and eight copies of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study
5A4P3N	JBx_078967	JPUB_010241	*E. coli* DH1 strain containing approximately five copies of AtoB-mvaS-mvaA, five copies of PMK-MK and eight copies of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study
4A5P3N	JBx_078968	JPUB_010242	*E. coli* DH1 strain containing approximately five copies of AtoB-mvaS-mvaA, eight copies of PMK-MK and five copies of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study
5A3P4N	JBx_078969	JPUB_010243	*E. coli* DH1 strain containing approximately five copies of AtoB-mvaS-mvaA, six copies of PMK-MK and six copies of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study
3A5P4N	JBx_078970	JPUB_010244	*E. coli* DH1 strain containing approximately four copies of AtoB-mvaS-mvaA, seven copies of PMK-MK and six copies of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study
5A5P5N	JBx_078971	JPUB_010245	*E. coli* DH1 strain containing approximately five copies of AtoB-mvaS-mvaA, eight copies of PMK-MK and eight copies of NudB-PMD gene cassettes in the chromosome at loci lacZ, intA and fdhF respectively	this study

Table 6.2. List of synCRISPRs used in chapter 4

gene/s	synCRISPR	sgRNA1	sgRNA2
recA	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATA AATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCT GAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAAT AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCC AAAACtatcaacttctacggcgaactggttgacctGTTTTAGAGCTATGCTGT TTTGAATGGTCCCAAAACgtttccacatccatggaacggtcttcacccGTTTT AGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACAC TGAGACTTGTTGAGTT	tatcaacttctac ggcgaactggt tgacct	gtttccacatccat ggaacggtcttc accc
accA	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATA AATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCT GAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAAT AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCC AAAACccagttgcgcaatctgccatgcaccgagatGTTTTAGAGCTATGCTG TTTTGAATGGTCCCAAAACcatcgttgaaagcgcaactgctggcggatcGTT TTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCAC ACTGAGACTTGTTGAGTT	catcgttgaaa gcgcaactgct ggcggatc	catcgttgaaagc gcaactgctggc ggatc
accB+acc C	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATA AATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCT GAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAAT AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCC AAAACgttcggagatgcctgattcttcaaccagctGTTTTAGAGCTATGCTGT TTTGAATGGTCCCAAAACactatgactcaatgatcggtaagctgatttGTTTT AGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACAC TGAGACTTGTTGAGTT	gttcggagatg cctgattcttca accagct	actatgactcaat gatcggtaagct gattt
lpxD+fab Z+lpxA	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATA AATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCT GAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAAT AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCC AAAACtgtttgtgcagattgcatggacgcaacgcGTTTTAGAGCTATGCTGT TTTGAATGGTCCCAAAACtagcggtaaaacgctcgatgaagtgaaacGTTTT AGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACAC TGAGACTTGTTGAGTT	tgtttgtgcaga ttgcatggacg caacgc	tagcggtaaaac gctcgatgaagt gaaac
plsB	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATA AATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCT GAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAAT AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCC AAAACgaatagacttgcttttttaccaggatgcttaaGTTTTAGAGCTATGCTGT TTTGAATGGTCCCAAAACactgcgtgatgaagggtatatcagcgataGTTTT AGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACAC TGAGACTTGTTGAGTT	actgcgtgatg aagggtatatc agcgata	gaatagacttgct ttttaccaggatg cttaa
acpS	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATA AATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCT GAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAAT AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCC AAAACgtctgtccccactcctcgctatgacggcttGTTTTAGAGCTATGCTGT TTTGAATGGTCCCAAAACtacggctatggggcgaggcattaaaactggGTTT	gtctgtccccac tcctcgctatga cggctt	tacggctatggg gcgaggcattaa aactgg

	TAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGAGACTTGTTGAGTT		
fadM	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATAAATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCTGAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACgcgggcttcttcgagaaattcaaggtagcgGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACttaagccaggtcattacactggagccggaaGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGAGACTTGTTGAGTT	gcgggcttcttcgagaaattcaaggtagcg	ttaagccaggtcattacactggagccggaa
lpxL	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATAAATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCTGAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACtccacaaccgaggcggtagataaccgggtaAGAGCTATGCTGTTTTGAATGGTCCCAAAACgagtgttctccgccactggatgatgccgaaGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGAGACTTGTTGAGTT	tccacaaccgaggcggtagataaccgggta	gagtgttctccgccactggatgatgccgaa
tesA	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATAAATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCTGAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACaaactcgccaaagagtttgatgttccgctgAGAGCTATGCTGTTTTGAATGGTCCCAAAACtcaataacgtgtccgctgcggcggcacggaGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGAGACTTGTTGAGTT	aaactcgccaaagagtttgatgttccgctg	tcaataacgtgtccgctgcggcggcacgga
fabI	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATAAATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCTGAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACtaggcgatggatagtttgctggcaacaccgAGAGCTATGCTGTTTTGAATGGTCCCAAAACgtgctccgatctctctgccggtatctccggGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGAGACTTGTTGAGTT	taggcgatggatagtttgctggcaacaccg	gtgctccgatctctctgccggtatctccgg
lpxM	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATAAATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCTGAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACgcgcgggtggcgaaaggatttatcaaactcAGAGCTATGCTGTTTTGAATGGTCCCAAAACgaacaatacacctggatactaaaattgctgGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGAGACTTGTTGAGTT	gcgcgggtggcgaaaggatttatcaaactc	gaacaatacacctggatactaaaattgctg
fabR	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATAAATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCTGAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACctgcgaatgatttcgaaaggggcttattacAGAGCTATGCTGTTTTGAATGGTCCCAAAACaggctggcgaagctgcgttcagcacttaatGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGAGACTTGTTGAGTT	ctgcgaatgatttcgaaagggggcttattac	aggctggcgaagctgcgttcagcacttaat

accD	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATA AATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCT GAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAAT AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCC AAAACtcaggaatgctcgccttgcgggtgggagtaAGAGCTATGCTGTTTTG AATGGTCCCAAAACgaaactggcgagcattctggcgaagttgatGTTTTAGA GCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGA GACTTGTTGAGTT	tcaggaatgct cgccttgcggg tgggagta	gaaactggcgag cattctggcgaa gttgat
fabA	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATA AATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCT GAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAAT AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCC AAAACaacagttcaccgcgaccagaggcaagaaggAGAGCTATGCTGTTTT GAATGGTCCCAAAACtaaccgtcgtctgattatgggcctggcggaGTTTTAG AGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTG AGACTTGTTGAGTT	aacagttcacc gcgaccagag gcaagaagg	taaccgtcgtctg attatgggcctgg cgga
fadR	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATA AATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCT GAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAAT AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCC AAAACgatactttcaataatgtactcttccgcgaaAGAGCTATGCTGTTTTGA ATGGTCCCAAAACgatcaggtgtacgaaacagtgcgtcgctatGTTTTAGAG CTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGAG ACTTGTTGAGTT	gatactttcaat aatgtactcttc cgcgaa	gatcaggtgtac gaaacagtgcgt cgctat
fabB	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATA AATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCT GAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAAT AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCC AAAACtcacgcagagatgccaggacttcctgctggAGAGCTATGCTGTTTTG AATGGTCCCAAAACagctatctactctctgctgatgctggaacaGTTTTAGAG CTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGAG ACTTGTTGAGTT	tcacgcagaga tgccaggactt cctgctgg	agctatctactctc tgctgatgctgga aca
tesB	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATA AATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCT GAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAAT AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCC AAAACcggcgtccagcgcacgtggctttgtgcgcgAGAGCTATGCTGTTTTG AATGGTCCCAAAACggcgtaaacctaaatcttcactctggccgcGTTTTAGA GCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGA GACTTGTTGAGTT	cggcgtccagc gcacgtggcttt gtgcgcg	ggcgtaaaccta aatcttcactctg gccgc
acpT	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATA AATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCT GAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAAT AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCC AAAACttgtcagcttgaaaatttaagccttgcgatAGAGCTATGCTGTTTTGA ATGGTCCCAAAACgcgttcgcgtcgtggaccttgcggtgcttgGTTTTAGAG CTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGAG ACTTGTTGAGTT	ttgtcagcttga aaatttaagcct tgcgat	gcgttcgcgtcgt ggaccttgcggt gcttg
plsX+fab H+fabD+f	TATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATA AATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCT GAGACAAATAGTGCGATTACGAAATTTTTTAGACAAAAAT	gcgcgtcatgg gaacgacgcat gccacgtt	cacgaagcgcgt caggttagcgga atggaa

| abG+acpP+fabF | AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCC AAAACgcgcgtcatgggaacgacgcatgccacgttAGAGCTATGCTGTTTTG AATGGTCCCAAAACcacgaagcgcgtcaggttagcggaatggaaGTTTTAG AGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTG AGACTTGTTGAGTT | | |

Table 6.3. List of 5' and 3' 500 bp each fixing template arms used in chapter 4

gene/s	fixing template 5'	fixing template 3'
recA	CCATGTTCATGGCACCAGGCAATAACGCGC TCGTAATCTTCTGCCGTAGCATCAATCTCTT CTGGGCCATTTTTGCCCATAATCGGTGCCGC GAGTTTACGTCGCAGTTCTTGCTCACTGTGA TCGCGCACCGCCAGAATGCGTACCGCACGA TCCAACAGGCGAGCATATGCCGGGCGACGG GATGTTGATTCTGTCATGGCATATCCTTACA ACTTAAAAAAGCAAAAGGGCCGCAGATGCG ACCCTTGTGTATCAAACAAGACGATTAAAA ATCTTCGTTAGTTTCTGCTACGCCTTCGCTA TCATCTACAGAGAAATCCGGCGTTGAGTTC GGGTTGCTCAGCAGCAACTCACGTACTTTCT TCTCGATCTCTTTCGCGGTTTCCGGGTTATC TTTCAGCCAGGCAGTCGCATTCGCTTTACCC TGACCGATCTTCTCACCTTTGTAGCTGTACC ACGCGCCTGCTTTCTCGATCAGCTTCTCTTT TACGCCCAGG	cccaggcgcatgatggagcctttaccaaattgtttc tcaatctggcccagtgctgccgccaacgctttctgt ttgttttcgtcgatagccatttttactcctgtcatgcc gggtaataccggatagtcaatatgttctgttgaagc aattatactgtatgctcatacagtatcaagtgttttgt agaaattgttgccacaaggtctgcaatgcatacgc agtagcctgacgacgcaccgcatcacggtcgcc gctgaagcattcccgccgggtaatgccttcaccg cgggcagtggcaaaagcaaaccagacggtgcc gacaggcttctcttcactgccgccatccggcccg gcgataccactaatagacacggcataatcagcac gagccgctttcagtgcgcctatcgccatttccacc acgacgggttcactcaccgcgccatgctgcgcca gcgtctcttcgcgtacgccgatcatctgcgctttgg cttcgttac
accA	TGATGACTTCAGCGGTGGGCTTAAAATGAC CGCTCGCGAAGTGATGGATATTGACGAAGC CCGGGAAAAATATGCTCGCGGGCTTGCTAT CTCGCTGACGGACAGGCAAATTGATGACCA GCTTTTAAACCGACTCCGTCAGTCTCTGGAA CCCCACCGCTCTGGGACAATTCCAGTACAT CTCTACTATCAGAGGGCGGATGCACGCGCG CGGTTGCGTTTTGGCGCGACGTGGCGTGTCT CTCCGAGCGATCGTTTATTAAACGATCTCCG TGGCCTCATTGGTTCGGAGCAGGTGGAACT GGAGTTTGACTAATACAGGAATACTATGAG TCTGAATTTCCTTGATTTTGAACAGCCGATT GCAGAGCTGGAAGCGAAAATCGATTCTCTG ACTGCGGTTAGCCGTCAGGATGAGAAACTG GATATTAACATCGATGAAGAAGTGCATCGT CTGCGTGAAAAAAGCGTAGAACTGACACGT AAAATCTTCGCCGATC	atctggccgatctcgacgtgttaagcactgaagatt taaaaaatcgtcgttatcagcgcctgatgagctac ggttacgcgtaattcgcaaaagttctgaaaaaggg tcacttcggtggcccttttttatcgccacggtttgag caggctatgattaaggaaggattttccaggaggaa cacatgaacatcattgccattatgggaccgcatgg cgtcttttataaagatgagcccatcaaagaactgg agtcggcgctggtggcgcaaggctttcagattatc tggccacaaaacagcgttgatttgctgaaatttatc gagcataaccctcgaatttgcggcgtgattttttgac tgggatgagtacagtctcgatttatgtagcgatatc aatcagcttaatgaatatctcccgctttatgccttcat caacacccactcgacgatggatgtcagcgtgcag gatatgcggatggcgctctggtttttttgaatatgcg c

accB+ accC	CATCTTCGCAGGGCTGGCTTTTCAGCTTTTC ACCTTACGTTATAAGAAGTTCCGTCGATGAT GGCGCTAATTTCGTGAATTGTGCGGCTTGTT GCAAATTACACGGTGTTGAAGGTTATTTAC ATGTTAGCTGTTGATTATCTTCCCTGATAAG ACCAGTATTTAGCTGCCAATTGCTACGAAAT CGTTATAATGTGCGACCTCGTCCTCCCTGAC GCAGTTTTTGCGCTGCGGAAAAGGTGACAT TGGCGCAACGAAGGTATATTTTGTTTTTTGC CGGAGGATAGCAGCAGATCGCTGCACAATG TCCGTCAAGTCTAACATTGACACTCTGGGGC AAAATAGACCGGCGTCCCGGCCTGCTGGAA TTTATCGCTATGCATACAGCTGTCGGGGCAT ACGCTTTACAGACGGCGGTGAAACGCCTGT CACAATCACACTAAACAAAGAGTACGGAAC CCACTCATGGATATTCGTAAGATTAAAAAA CTGATCGAGCT	tttgctacggtgaaaaccgtgacgtggcgattgcc cgcatgaagaatgcgctgcaggagctgatcatcg acggtatcaaaaccaacgttgatctgcagatccgc atcatgaatgacgagaacttccagcatggtggca ctaacatccactatctggagaaaaaactcggtcttc aggaaaaataagactgctaaagcgtcaaaaggc cggattttccggccttttttattactgggggatcgaca acccccataaggtacaatcccccgctttcttcaccca tcagggacaaaaaatggacactcgttttgttcagg cccataaagaggcgcgctgggcgctggggctga cccttttgtatctggcagtttggttagtagccgctta cttatctggcgttgcccccggttttaccggctttccg cgctggtttgagatggcctgcatcctgacgccgct gctgtttattggactgtgctgggcgatggtgaaattt atctat
lpxD+f abZ+lp xA	TGCAATCGTCAACATGGGCAGCCTGTTCCA GCAGGTAGCGCAGAAAACCGGTGTTTCTAA CACGCTGGAAAATGAGTTCAAAGGCCGTGC CAGCGAACTGCAGCGTATGGAAACCGATCT GCAGGCTAAAATGAAAAAGCTGCAGTCCAT GAAAGCGGGCAGCGATCGCACTAAGCTGGA AAAAGACGTGATGGCTCAGCGCCAGACTTT TGCTCAGAAAGCGCAGGCTTTTGAGCAGGA TCGCGCACGTCGTTCCAACGAAGAACGCGG CAAACTGGTTACTCGTATCCAGACTGCTGTG AAATCCGTTGCCAACAGCCAGGATATCGAT CTGGTTGTTGATGCAAACGCCGTTGCTTACA ACAGCAGCGATGTAAAAGACATCACTGCCG ACGTACTGAAACAGGTTAAATAAGTAATGC CTTCAATTCGACTGGCTGATTTAGCGCAGCA GTTGGATGCAGAACTACACGGTGATGGCGA TATCGTCATCACCGGCG	aaccggaaattgctgaactggcggaaacatatcc ggaagtgaaagcctttaccgatttctttgcacgctc aacgcgcggtctgattcgttaatgactgaacagcg tccattaacgattgccctggtcgccggagaaaacct ccggcgatatcctgggggccggtttaatccgcgc tctgaaagaacatgtgcccaacgccccgctttgttg gtgttgccgggccacgaatgcaggctgaaggct gcgaagcctggtacgaaatggaagaactggcgg tgatgggcattgttgaagtgctcggtcgtctgcgtc gcttactgcatattcgtgccgatctgacaaagcgtt ttggcgaactgaagccagatgttttgttggtattga tgcgcctgacttcaatattactcttgaaggtaacctc aaaaagcagggtatcaaaaccattcattacgtcag tccgtcagtctgggcgtggcgacagaaacgtgttt tcaaaatag
plsB	TTATTGGTATTTTGCAGATTGGCGTACTGGC ACTGATGGCGATCATCGGTGAGTTAAATGG CTTAGGCTGGGGATATTACTGGTCAATTCTG GTGGCTGGCGCGCTGTTTGTTTATCAACAAA AACTGATTGCCAACCGCGAGCGTGAAGCCT GCTTTAAAGCATTTATGAATAATAACTATGT TGGTCTGGTACTATTTTTAGGGCTGGCAATG AGTTACTGGCATTTCTGATGATGTAAAAAA GCCGGATGATCATCCGGCTTTCTTCTGGGTT GCCTGATGCGCGGCGCTTCTCAGGCCTACA CAACACATCGCAATTTATTGAATTTGCAGAT TATGGAAGGCCGGATAAGGCGTTTTCGCCG CATCCGGCAATTCTCTCTGATTACCCTTCGC	taatggtaaattcagtaatttgtagtaaattcgtggc cagccggacataaacgatgtaaagcctctggttaa taatgcaaatgcgcggcaaggatagcagaaagtc atgggaaattctgtggtatccgctcatgtttcgcgc ggcgctacgcaaacccgaatcatcggatttaacg gtacactgatattgacgctcataatgtaaaaaggtt ctttcaatggccaataataccactggattcacccga attatcaaagctgctggctattcctggaaaggttta cgcgctgcatggatcaacgaagcggcattccgtc aggaaggcgtagcggtattgttggcggtggtcat cgcctgctggctggatgtggacgcgattacccgc gtgctgcttatcagctccgtgatgctggtgatgatt gtggaaatcctcaatagcgccatcgaagcagtgg

	CCTGCGTCGCACTCTCAATCGTCAAACGCA CGTCTGATGTAATCAACTCCGCCAGCAACT GATAAACCTTCATCGTTTCTGCCGGTTCGGC ATCGCCGCTAT	ttgaccgaattggctctgaataccatgagctttccg gacgcgca
acpS	CTTAAACAAGGAATGCAGCCATTCAGATTA GCCCTTACATCTCCCCAAAACTGAACGTGC GAGTTATTGAGGGTGCATGCTGCACTCCAC ACCAGAGCTTTGACGACACCACTCGTTTCA ATGGGGGAATTCTGTGGCATGGTGTAAAGC ACAGCAAAATCTTCAATAACGAAGCCAATT TTAATGTACTTACGAATTGGCGGTCACGTTA ATCTTTCCATCAATATTGCTTTCTTCGTAAA GGCTCGAGTTTTTATGCTAAAGATTGCAAGT TGCTTGTAAAAGATAAGTACACTGATCCAT AATCGCTGTTGTTGAGGGTGCATGCTGCAC AAAATTAAAGTTAAAAAGTAAAACCCCCGT TCCTTACCAGTTCGGGGGTTTTACTTTTTAA AGAGAACGGTATTATTTTTAACTTTCAATAA TTACCGTGGCACAAGCATAGTGCCGCTCAT CTGCCAGCGTTACATGCATATTTGCAACGCC CAGCTTTTCCGCCA	cttcgatgcgagcgatctccacaatatccgtgcct aaacctaatattgccattagccacgcgcttccagc atcagacgcttcatttctgccaccgcatctttcagtc cggtcatcactgcacgaccaataatggcatgacc gatattcagttcatgcatctcagggatggcggcaa tggctttcacgttgtgataggtcagaccgtgtccgg cgttaactttcagaccgaggcttgcggcaaaggtc gcggctttggcgatacgcgccagctcttgcgcct gttcggcgtcagttttggcatcagcatagcaaccg gtgtggatctcgataaacggtgcgccaacctctgc cgcagctttgatctgctcttcatcggcgtcaataaa cagagaaacctgaatcccggcatctgccagacgt ttgcaggcatcgcgcattttgtcacgctgccctgc gacatccaggccgccttcggttgttacttcctgacg cttttccg
fadM	ACCCTGTCCCTGGCCTGTGCCGGAATGTCTC ATAGCGCGCTGGCGGCAGCTTCTGTGGCGA AACCGACGGCGGTAGAAACCAAAGCGGAA GCTCCTGCAGCACAAAGTAAAGCAGCAGTA CCGGCGAAAGCCAGTGACGAAGAAGGCAC CCGGGTCAGCATTAATAATGCCAGCGCGGA AGAGCTAGCCCGCGCGATGAATGGCGTTGG CCTGAAGAAAGCGCAGGCGATTGTCAGTTA TCGCGAAGAGTACGGTCCGTTTAAAACTGT GGAGGATCTAAAGCAGGTGCCGGGGATGGG CAATTCGCTGGTGGAACGTAATCTGGCGGT ATTAACCCTGTAATTAATTTGCATAGTGGCA ATTTTTTGCCAGACTGAAGAGGTCATACCA GTTATGACCTCTGTACTTATAACAACAACGT AAGGTTATTGCGCTATGCAAACACAAATCA AAGTTCGTGGATATCATCTCGACGTTTACCA GCACGTCAACAACGCCCG	gaagggcaggtggtagcggatgcgcttattacgtt tgtttgtattgatcttaaaacgcagaaagcattagct ctggaagggaattgcgcgaaaagctggagcag atggttaagtaaacgtttgtggtgccggatgctca agccgcatccggcgacacccggaataattacctc aacccggtttctgcttcatcgctgccatcaccgtc ggtttatcggccagataatgattcaaaccgttggc gcgtaaattacatgccgcacaatgaccgcaaccg tcgcctttaaagccgttatagcaggtcaacgtttcg ttacggactaaatccagtttgccgtaataatctgcc agcgcccaggtttccgctttatcaatccacatcagc ggcgtttcaaaacgaatatctttcgccatgcccaaa ctgacggcatggtttagtgctttcacaaactcatcg cggcaatccgggtagccggagaaatccgtttcgc agacgc
lpxL	ATGGCCGAAAATAAAAATGTAATGCTGAAG ACAATACCGGACCACATATTCAGGGCGGAG TGACCGGTAACGCCAAGCTGCTCAACGTAG AGGGGTAAGAAGGGCATTACCAGACTGAAG GCGGCACCGGTAAGAAAACAGCCTAGCCAG GCGACGATCAGGTTTCGTTTCCAGTTTATAG GGGTGTCATTTTCACAGGGTGACATAGCAA TCCGCTGTTGGTGCGCCAGGCGCGGTGAAC ATAAGAAGAAAAGATAAGCACACTAATTAT	gtagggcaattgcacgactaaccaaagtacgcca atacccaaccaggttaaccaataacgcggatgaa gcagtgcggtggagaacttgggtagattcgtcata tcaatcctgtttttcaacctattcgggcaattgtatgt attgtcgcattttttcgcccgcaaccaaaatttgtgg ctgaagactgggcgaaattgccgcgcttgtaaata acaaataattttaatgcgcaaatgtagcgtaaaat gtgtggatgttaattatcgataattgctatatcatgcc gcggatttttactttcccatctcgcaggaaccgtac

	GCGCCCGACTTCCAGGGGGCGCAATCCAGA GAGCTTTTATCGCTAAATCAGGGGGATTTGC TGTGGTAATGCCGGATGCCATTCTGAAGCA TCCGGCATGGGAGATTTAATAGCGTGAAGG AACGCCTTCCGGGCGTGTTTTAAAGCGACG GTGTAACCACATATACTGCTCTGGTGCCATC ATGATGCATTTTTCGACCACTTTGTTCATCC ACGCGGCGGTAGTTTC	accatgccagtgttacacaaccgcatttccaacga cgcgctaaaagccaaaatgttggctgagagcgaa ccgcgaaccaccatttcgttttacaagtatttccaca tcgccgatcctaaggcgacccgtgacgctttatat cagctgtttaccgcgctgaatgttttttgggcgagtg
tesA	GTGCGCCGAAAAAGTTGGCGGAAAACTGCT GTTCCATCTGCGCACGGCTGATGGTGGAAA GGGGGCCATACATGCCGAATCCGGCATTGT TAAAGATCCCATACAGACAATTATCGGTCA GGGCGATCACCTCGTCGGCTGCGCGATCAA CACTTTCTGGTGAATCCAGATCGATCAACAC GCCGGTAAATCCCATGCTGTTCATGCGCTCA ACATCATCCGGTTTCCGGCAACCTGCCAGC ACATGAAAACCCTGGCGTTTTAATTCGAGC GCGCTTTCCAGGCCAATTCCACTGGAACAT CCGGTAATTAAGACCGATTTTTGCATAACTT TACCTGTCAGGATCTCCGTTGCTTTATGAGT CATGATTTACTAAAGGCTGCAACTGCTTCGC CATCCAGTCGGCAATAAACGGCTGGGCGTC GCGGTTGGGATGAATACCGTCATCCTGCAT CCATTGTGGCTTGAGGTAGACCTCTTCCATA AAAAAGGGCAGCAG	ggaaggttaacaggaccaggaacaggaagggc aaatgccagcggaaaacattgttgaagttcatcat cttaagaagtccgtcggtcaggggggagcatgaac tctccatcctcaccggagttgagctggttgtcaaac gtggcgagaccatcgcactggtgggcgagtcgg gatcgggtaagtcaaccttgctggcgatcctcgcc gggcttgatgacggcagcagtggcgaagtgagt ctggtgggacaaccgctacataatatggacgaag aagcgcgggcaaagttgcgcgcgaagcacgtc ggctttgttttcagtcatttatgttaattcctacccta acgcgctggaaaacgtcgagcttccggctctgct gcgcggtgagagtagcgcggaaagtcgtaacgg ggcgaaagcgttgctcgaacagttagggctgggt aaacgtctggatcatcttccggcacagctttccgg cggtgaacagcaacgagtgg
fabI	CCGCAAACCTGAGCAACCGTCAGGGCCGTC GACAAGCGTTCATAAGCGGTGAAGGTCTGT GAACGGGTCACGGTAAGCTGGTCAGGGAAC AAATCCACTGCAAGAACACGCGCAGGGTCC AGCCATGAAAGGAAGGGGGCCAGCGTTTCT TTTTCAAGGATCACGTCGAGTAGAAAGGGA TCGCTGACCTTTGCGGCTTCCGGGACCAGA GGCAGAACGTCATACTCCGTAGTACTGGAT TGCGTTTCATGATACCAGTGGCTTTTGCCGG TGATGTGGCGTTGTTCCATGGGCGCTCCTTG GTCGTAAAGGAAATCGTTATCCTGACGCAA GGCGGGAAGGGGAGAAAGACGGATCGGCG ATAACAAATATCAGAAAGGTATAACAGAGA TAACGGGCGGCAGAACGCCGCCCATCTTTA CCAACAGAACGATTATTTCAGTTCGAGTTCG TTCATTGCAGCAATGCTGAAACCGCCGTCA ACGTGGACCACTTCAccg	ccggttaccagaatgcgcttaccggaaagaaaac ccatagctttaatccttattgttgatgcttgttgtgcct gaaaatcaggcgattcgtcgttttagtaaacagtac gaacagataaacggttattataatcaacctggctgt gagtagctatagttgccaggtccgaccggagcag gctgcggcagggggggcgcttttcccctcaccct aaccctctccccagagggggcgaggggaccgtat tgtgcaaatattgttaccccagcaacaaacaggct catacagcccctaaccctttcatggcgatggctgg gacggttcagaccttgccgaatattctccagcacc gtctccatgtttttcaccacaacagctattcggctcg gtctgcccccctcgctctttcagggagagggtgacc ggcggttcagttcctgcagaatattctccagcacc gtctcctcattacaatcaatctcgttattccagaaac gcagca
lpxM	GGCGATCATCACCATGACCGAATCGGGTCG TACCGCGCTGATGACCTCCCGTATCAGCTCT GGTCTGCCAATTTTCGCCATGTCGCGCCATG AACGTACGCTGAACCTGACTGCTCTCTATCG TGGCGTTACGCCGGTGCACTTTGATAGCGCT	ctcaggaatgtattcgctattatttttttttcgtttccatg cttttccagtttcggataaggcaaaaatcaatctggt gatagtgtagcggcgcaacttgccccgcaccaaa taaaaaaagccggtactgactgcgtaccggctgc gaatggatgttaattaatcaaaccgtagctgcggc

	AATGACGGCGTAGCAGCTGCCAGCGAAGCG GTTAATCTGCTGCGCGATAAAGGTTACTTGA TGTCTGGTGACCTGGTGATTGTCACCCAGGG CGACGTGATGAGTACCGTGGGTTCTACTAA TACCACGCGTATTTTAACGGTAGAGTAAGT ACGTTGCCGGATGCGGCGAAAACGCCACAT CCGGCCTACAGTTCAATGATAGTTCAACAG ATTTCGAATATTCTGAAGCAAACTTGAACTT ATCATCAGGCGAAGGCCTCTCCTCGCGAGA GGCTTTTTATTTGATGGGATAAAGATCTTT GCGCTTATACGGCTGGATTTCGCCCGGTTTG CGAGTTTTCAG	acaatctctttggcctgcgccaggaattcgcgacg atcggagccggtcagcccttcggtacgcggcagt tttgccgtcagcgggtttacggcctgctggtttatc catacttcatagtgcagatgcggcccggttgaacg tccggtattaccggaaagcgcgatacggtcgcca cgtttcaccttctgtcccggtttcaccagaatcttgc gcaagtgcatataacgcgtggtgtagctgcgacc atgacgaatagccacataataacctgctgcgcca ctgcgtttggcaaccaccacttcaccgtcacccac tgaaag
fabR	TGCGACGCGCGCACCTTGCTTAACCAGGCC CATTGCAGCGCCTTCGCCGCCGGGGCCGGA ACCTATTACTATGGCATCGTAATCGTAGGA ATGTGGCATGGTAGGGCTTACCTGTTCTTAT ACATAAAAGCAACAGAATGGTAACATTTTA TCGCGGGTAAGCCAATTGATCCCCGTCATTT ATCTGGCTATATCCTGAGCGGCCTTTGCTTT GTCTGTTTCTTACTTTTGCCCTGACGTTTTAT TGGATTTTTATCGACGATACTCTCCGTTTAA GCGGCAGGTTTCCGCTGTACGTAAAAGAAC CGGCCAAAGAATTGCAGTAAATATGTTTTA TTGCGTTACCGTTCATTCACAATACTGGAGC AATCCAGTATGTTCATTCTCTGGTATAGTGC CAGCAGTACTTTTGGCAAGGATTCAGACAT CGTGATGGGCGTAAGAGCGCAACAAAAAG AAAAAACCCGCCGTTCGCTGGTGGAAGCCG CATTTAGCCAATT	tactggtatcgccgtgaacaagagaaaaccgcaa ttattccgggaaatgtgaaggacgagtaatgaaac aagcaaatcaagatagaggtacgctgctgctggc gttagttgctggcttatcgattaatggtactttcgca gcgctgtttagctccattgtgccattttctgtattccc gattatttccctggtgctgacggtttactgcctgcat caacgttatcttaatcgcactatgccggtaggcttg ccgggtctggcagctgcctgttttattctcggcgta ctgctgtacagcacggtagttcgtgcggaatatcc ggatatcggctctaacttcttcccggcagtactctc cgtcattatggtgttctggattggcgcgaagatgc gtaaccgtaagcaggaagttgctgagtaatcggta ttatgccggatgctggcgcatccggcatgggtttta cttcgcggtcagtaatacgccgcactccatatggt g
accD	GTGCCGATTCAATCTCCGCAAAAGAGGCGG TGTGGGCCGATTCCGGCAATTCCTGGCCCTG CACACGTACGCGCTCGGTATAACGCACCAG ATGAGGCGAACTGTAGACGCCCACTTTGTA CCCTGCCGCCATCAGAATCGACTCCAGCGT ACGGCAGGTGGTGCCTTTGCCATTCGTACCC GCAACGGTAAACACAAATGGCGCTGGTTTC AGGACGCCAAGACGCGCCGCGACCAGGCTC ACGCGCTCAAGGCCGAGATCGATAGTTTTA CTGTGCAGGTTTTCCAGATAAGAAAGCCAC GAAGCCAGAGGCGACGCGGCTTGAGGAGTG CGTTTGATAATCATGGTATCCGCTGATTCGT TACGGTGAGAATAGCAAAAGGGCAGAGCC AGTGGCCCTGCCCTTATCAGTTATCAGGCCT CAGGTTCCTGATCCGGTACCGGGGGTACCA CTACGCCTTCACGCGGCGCTTCAGGATTCG GCGCTGGCAGATTCATC	gtaatgttgcttttaattcgttcaatccagctcattag ggacctttctgtctgaacctggttcgatgccagtttt atctttggggacgcataatgccattttttgcccccaa cagaccatgaatgttgcacattaaaacataacagc ccgaaactttggataaaaaagtggtcgaaccgcg gagttactttttattttgcggcgcgtgccgcagcgc gtttgtgacggattatttcgatgacgcccggcaaa atagaaaccacaataatcccgacgatcagcagctt aaggttatcctgaaccatcgggattgtaccgaaga aatagcctgcgtaggtaaaaagcagtacccacaa cagtgcgccgatcacgttataagcggcgaaatga cggtacgacatgtggcccattcccgcaacaaacg gggcgaacgttctgacgatcggcacaaaacggg cgagaataatcgtttgccgccatgtttctcataaaa ctgatgg

fabA	TTCCTCTTCACATTTTTTGTACAACCGACAT GCCCGTGTAGCTCACAAATATGACAGTGGC GTGAATTTTGCGCATTGACGGCAGTTATGAT TCGCGGTATTGCTTAACTGTGATTGCACATT TAGTAATCACTGTTTTCTTTTCCACCAGAAA CCAGTATGAGGGAAACGAGGCATGAAGAG ACAAAAACGAGATCGCCTGGAACGGGCAC ATCAACGTGGTTATCAGGCCGGCATCGCCG GACGCTCAAAAGAAATGTGTCCCTATCAGA CGCTGAATCAAAGGTCACAATGGCTGGGAG GCTGGCGAGAAGCCATGGCGGACAGGGTAG TAATGGCCTGATTCTGTCTCTTTAAAAAGAA ACCTCCGCATTGCGGAGGTTTCGCCTTTTGA TACTCTGTCTGATTATAATCAGAAGGCAGA CGTATCCTGGAACAGACCGACTTTCAGGTC GCTGGCGGTATAGATCAGACGACCATCAAC CAGCACTTCGCCATCC	aggtcttcttttgtataggattcgcgtttatctaccat gttctctgtaagccttattttattgaagcacgcagga tagctaacacgtgtacgctgaacaagtccgatcag ttcggaataaaccagttcagccaacgtaatggcca tggaaaacggtgacgtccttcctgttgcgatgcttg cgcgatacgttcctggatggtttgcatcagcgtcgt ttggccttcgccatcccacaccagatttaataataa cggcagtgcgtcagtcacatcgtctactgcccag atggtgaatttgccttcttctaccgctttcaccagttc actgtgaagacttaaatggcgaacgttagctgtgg ggataatgacaccttgtttcccggttaactcacgtt gctggcaaatagcaaagaagccttcgattttctcat ttaaaccaccgaccggctgggcgcgaccgaact gatcgactgaacctgtgatagcgatactctgattc
fadR	CCGCGCTGGTCATGCCGATGAATACCGCTT CGATAGCCAACAGACCACCGGGGAGCAGC GGGTAGCATTTCAGGGCCATCGCCAGAGTG AAAATAAATTCCGCTACCAGCAACCAGCCC GCGACGAAAGGGCTGATGAGGAAAATTAAC GGGTTTACGATTAAGAAAATGATGAGGGCG AGTTTGTACCAGTCGGGGGACTGGCCCAAA AAATTGCGCCATAGCGCGCGGCCCCAGGAG ATCTCCATGATGGTTTCCCTTACCTTACAAA TAATCAATGATGTTTTTATGTTTAAACGCAA AGCTTAACGGTCAGGCAGGAGTGAGGCAAG TCTTGATAGTCAAGGGGAAAGAGATGCGGA AAATGAAGCCTTGATCCCTTTTTCTTCTTTTT GTCTGCTATCAGCGTAGTTAGCCCTCTGGTA TGATGAGTCCAACTTTGTTTTGCTGTGTTAT GGAAATCTCACTATGGTCATTAAGGCGCAA AGCCCGGCGGGTTTC	tatgggcatgagagtggcgagatttggcaccgga tgcagaaaaatctgccgggtgatttagccattcag gggcgataatcccttccgtttaaagagcaaacccc tcaaacgaggggtttttgttgtttttacagatttccc attcttggcgggcaacgttccagcaactcgatgct gccgtcttcgttttgctgttcgagcatcacatcaaat ccccacaggcgatgcacatgcttcaggacttcttt gcgcccccgatccagcggtgcgcgattatgtgga atataacgcagcgtcagcgaacggtcgccgcgc aaatccacgttccagatctgaatattcggctccaga ttacttaagttatattgcgacgataaccggttacgg atctcccgataaccttcttcattatgaatagcggaa atctccagataattatgccgatcgtcatccagcacg gtgaagaagcggaaatcacgcatcactttcggtg acagga
fabB	CAAAAGCGATGCGGAAGAGCGAAAAATTG GCAGAAGATCGTGTTTCGGTTTGGCTTACAG CAGTCATGCATTACTCCAGAATGCAGCGCA AGGCGAGGAGTATCCCCGTCTCATCTCTCTG GTTTCAGGGTTACGGTGCGTTGGCAGGATTT AACGCGTACGTCTTTTCAGAAGGAAATCGA CAAAGCGGGAAGTTTGCCTGGAACTGGCGG CGATTGTCAATGATGTGAAAAAGGGAACCA TCAGGTTCCCTTTTGCGTTAGTGCCGGAGGC CGCATTGGCGCGTAACGTCGGATGCGACGC TGGCGCGTCTACTCCGACCTACTGCGAATTA ATCTTTCAGCTTGCGCATTACCAGCGTGGCG TTGGTGCCGCCGAAGCCGAAGCTGTTAGAC	tggttattaccgatgctggaaacaatgcccaggcc agtaatcactgcacgtttcattcaatacctctgtaag tcgcacatagagtaagtttcgaatgcacaatagcg tacacttgtacgccgaacaagtccgatcagccattt aatagagaaatttgcgcagccttacacacatcgct aagatcgagccaccgcctgtaagacgagtaactt acgtgaaacactactccatacaacctgccaacctc gaatttaatgctgagggtacacctgtttcccgagat tttgacgatgtctattttttccaacgataacgggctgg aagagacgcgttatgtttttctgggaggcaaccaa ttagaggtacgctttcctgagcatccacatcctctgt ttgtggtagcagagagcggcttcggcaccggatt aaacttcctgacgctatggcaggcatttgatcagtt

	ATAACGGTGGTCAGTTCGCGATCGGTCGTTT CGGTCACGATGTTCAGACCCGCAGCCTGCT CGTCCAGCTCTTCAATGTTGATGCTCGGGGC GATAAAGCCGTGT	tcgcgaagcgcatccgcaagcgcaattacaacg cttac
tesB	GGCGCGTTGATTATCGGCGTGGTAGCTGGT CTGGCGGGCTTGTGGGGCGTTACCATGCTC AAACGCTTGCTGCGGGTGGATGATCCCTGC GATGTCTTCGGTGTGCACGGCGTTTGTGGCA TTGTCGGCTGTATCATGACCGGGATTTTTGC CGCCAGCTCGCTGGGCGGCGTGGGCTTCGC TGAAGGTGTGACGATGGGCCATCAGTTGCT GGTACAGCTGGAAAGCATCGCCATTACGAT CGTCTGGTCCGGTGTTGTGGCATTTATCGGC TACAAATTGGCGGATCTGACGGTTGGTCTG CGTGTACCGGAAGAGCAGGAGCGAGAAGG GCTGGATGTCAACAGCCACGGCGAGAATGC CTATAACGCGTAACAAGCACTGCAAAAAAC AGCCGGACGGTTTTCACCTCCGGCTATTTTT TTAATTGTGATTACGCATCACCCCTTCCTGA ACGGTCGAGGCAACCAGTACGCCGTCTTGG GTATAAAACTCACCGC	cgcgaaagagtccttcctcaatttttccagatttaa caatgtcagtaaattttttagcgcctgactcatataa ctctccagtaacaaagctgccgcagcaagccaaa gtgagttgagtataacgcaaatttgctactggtccg atgggtgcaatggtctgaattacgggctaattaca ggcagaaatgcgtgatgtgtgccacacttgttgac gttactattttgttaaccactcttccggcgaggaaa gttagcccgctggtgcattgataataaggagaaat gaatgaaactcgtgcacatggccagtggtttagcg gttgcgattgcgttggcggcttgcgcagataaaag cgcggatattcagacgccagcaccggctgcaaat acgtctatttcagcaacacaacaaccagctatcca gcaaccgaatgtctccggtaccgtctggatccgtc agaaagtcgcactgccgcctgatgctgtgctgac cgtgaca
acpT	GGGCAGTACTGGCGTTTTTGCTCAACTACGT GCCCAATATCGGCGCGGTAATTTCCGCCGT ACCGCCAATGATTCAGGTGCTGCTGTTTAAT GGTGTTTACGAATGTATTCTGGTCGGCGCAT TGTTTTTAGTGGTCCATATGGTCATCGGCAA TATTTTAGAACCACGGATGATGGGCCATCG CCTGGGGATGTCCACCATGGTGGTATTTCTT TCATTGTTAATTTGGGGATGGCTGCTCGGCC CGGTAGGGATGCTACTTTCGGTGCCATTAAC CAGCGTGTGTAAAATCTGGATGGAAACCAC CAAAGGCGGTAGCAAACTGGCGATTTTACT GGGGCCGGGCAGACCGAAAAGTCGGTTACC GGGATGAGGCGACAAGTGATACGATACGCA CTTTCATTTTCCATTAAACGTTGGCCCTGAT ATGTATCGGATAGTTCTGGGGAAAGTTTCG ACCTTAAGCGCAGCTCCACTGCCACCGGGT TTACGCGAGCAA	gatctgcacccctactccctttacgctcaccgccg acagtgtgcaatggatcgattcagttaactgatccg cccacccgactgcccatctattgatccagaacag gtaatcagtatgacgaatacttaaaatcgtcatactt atttccgccatctattttaatccattgggttaccatg ctctccacactccgccgcactctatttgcgctgctg gcttgtgcgtctttatcgtccatgccgctgcacca gatgaaatcaccaccgcctggccggtgaatgtcg ggccactaaacccgcacctttacacgcctaacca gatgttcgcccagagcatggtttatgaaccattggt gaaatatcaggcagacggttcggtgatcccgtgg ctggcaaaaagctggactcattcagaagatggta aaacctggaccttcaccctgcgtgatgacgtgaaa ttctccaacggtgaaccgttcgatgccgaggcgg cggcaga
rpmF+ plsX+f abH+f abD+f abG+a cpP+fa bF	CGTAGTCAGTGTGGACAGTGATGTGGAATG CTCCATGTCGTTCGCTATCGATAACCAACGT CTCGCAGTGTTAAACGGCGATGCGAAGGTG ACGGTAACGCTCGAGTGTCAGCGTTGCGGG AAGCCGTTTACTCATCAGGTCTACACAACGT ATTGTTTTAGTCCTGTGCGTTCAGACGAACA GGCTGAAGCACTGCCGGAAGCGTATGAACC GATTGAGGTTAACGAATTCGGTGAAATCGA TCTGCTTGCAATGGTTGAAGATGAAATCATC	gaatacactctgtgtaactccttcggcttcggtggc actaatggttctttgatctttaaaaaagatctaagttgt cattttccacccttataaaaggtccgcttgcgggcc ttttttcttagcttttattccgacttgttccgtagtgaac atgctgccacactaacaattctctgataaggagcc ggtatgttcttaattaacggtcataagcaggaatcg ctggcagtaagcgatcgggcaacgcagtttggtg atggttgtttaccaccgccagagttatcgacggta aagtcagtttgttatcggcgcatatccagcgactac

| CTCGCCTTGCCGGTAGTTCCGGTGCATGATT CTGAACACTGTGAAGTGTCCGAAGCGGACA TGGTCTTTGGTGAACTGCCTGAAGAAGCGC AAAAGCCAAACCCATTTGCCGTATTAGCCA GCTTAAAGCGTAAGTAATTGGTGCTCCCCGT TGGATCGGGGATAAACCGTAATTGAGGAGT AAGGTCCATGGCCGTACAACAGAATAAACC AACCCGTTCCAAAC | aggatgcttgtcagcggttgatgatttcctgtgactt ctggcctcagcttgaacaagagatgaaaacgctg gcagcagaacagcaaaatggtgtgctgaaagtc gtgatcagtcgcggtagtggcgggcgagggtac agcacattgaacagcggaccggcaacgcggatt ctc |

Bibliography

1. Jarboe, L.R., et al., *Development of ethanologenic bacteria.* Adv Biochem Eng Biotechnol, 2007. **108**: p. 237-61.

2. Ingram, L.O., et al., *Metabolic engineering of bacteria for ethanol production.* Biotechnol Bioeng, 1998. **58**(2-3): p. 204-14.

3. Ingram, L.O., et al., *Genetic engineering of ethanol production in Escherichia coli.* Appl Environ Microbiol, 1987. **53**(10): p. 2420-5.

4. Clomburg, J.M. and R. Gonzalez, *Biofuel production in Escherichia coli: the role of metabolic engineering and synthetic biology.* Appl Microbiol Biotechnol, 2010. **86**(2): p. 419-34.

5. Zhang, F., S. Rodriguez, and J.D. Keasling, *Metabolic engineering of microbial pathways for advanced biofuels production.* Curr Opin Biotechnol, 2011. **22**(6): p. 775-83.

6. Steen, E.J., et al., *Microbial production of fatty-acid-derived fuels and chemicals from plant biomass.* Nature, 2010. **463**(7280): p. 559-62.

7. Peralta-Yahya, P.P., et al., *Microbial engineering for the production of advanced biofuels.* Nature, 2012. **488**(7411): p. 320-8.

8. Nielsen, J. and J.D. Keasling, *Engineering Cellular Metabolism.* Cell, 2016. **164**(6): p. 1185-1197.

9. Chou, H.H. and J.D. Keasling, *Synthetic pathway for production of five-carbon alcohols from isopentenyl diphosphate.* Appl Environ Microbiol, 2012. **78**(22): p. 7849-55.

10. Atsumi, S. and J.C. Liao, *Metabolic engineering for advanced biofuels production from Escherichia coli.* Curr Opin Biotechnol, 2008. **19**(5): p. 414-9.

11. Alper, H. and G. Stephanopoulos, *Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential?* Nat Rev Microbiol, 2009. **7**(10): p. 715-23.

12. Koppolu, V. and V.K. Vasigala, *Role of Escherichia coli in Biofuel Production.* Microbiol Insights, 2016. **9**: p. 29-35.

13. Zeng, Y., et al., *Lignin plays a negative role in the biochemical process for producing lignocellulosic biofuels.* Curr Opin Biotechnol, 2014. **27**: p. 38-45.

14. Wang, W., et al., *Connecting lignin-degradation pathway with pre-treatment inhibitor sensitivity of Cupriavidus necator.* Front Microbiol, 2014. **5**: p. 247.

15. Rude, M.A. and A. Schirmer, *New microbial fuels: a biotech perspective.* Curr Opin Microbiol, 2009. **12**(3): p. 274-81.

16. Yim, H., et al., *Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol.* Nat Chem Biol, 2011. **7**(7): p. 445-52.

17. White, A.K. and W.W. Metcalf, *Microbial metabolism of reduced phosphorus compounds.* Annu Rev Microbiol, 2007. **61**: p. 379-400.

18. Yang, D., et al., *Metabolic Engineering of Escherichia coli for Natural Product Biosynthesis.* Trends Biotechnol, 2020. **38**(7): p. 745-765.

19. Keasling, J.D. and H. Chou, *Metabolic engineering delivers next-generation biofuels.* Nat Biotechnol, 2008. **26**(3): p. 298-9.

20. Tucker, J.B. and R.A. Zilinskas, *The promise and perils of synthetic biology.* New Atlantis, 2006. **12**: p. 25-45.

21. Pleiss, J., *The promise of synthetic biology.* Appl Microbiol Biotechnol, 2006. **73**(4): p. 735-9.

22. Gibson, D.G. and J.C. Venter, *Synthetic biology: Construction of a yeast chromosome.* Nature, 2014. **509**(7499): p. 168-9.

23. Mukhopadhyay, A., et al., *Importance of systems biology in engineering microbes for biofuel production.* Curr Opin Biotechnol, 2008. **19**(3): p. 228-34.

24. Lee, S.Y. and H.U. Kim, *Systems strategies for developing industrial microbial strains.* Nat Biotechnol, 2015. **33**(10): p. 1061-72.

25. Esvelt, K.M. and H.H. Wang, *Genome-scale engineering for systems and synthetic biology.* Mol Syst Biol, 2013. **9**: p. 641.

26. Zecca, A. and L. Chiari, *Fossil-fuel constraints on global warming.* Energy Policy, 2010. **38**(1): p. 1-3.

27. Ward, J.D. and W.P. Nel, *Comment on Fossil-fuel constraints on global warming by A. Zecca and L. Chiari [Energy Policy 38 (2010) 1-3].* Energy Policy, 2011. **39**(11): p. 7464-7466.

28. Hill, J., et al., *Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels.* Proc Natl Acad Sci U S A, 2006. **103**(30): p. 11206-10.

29. Steinberg, M., *Fossil fuel decarbonization technology for mitigating global warming.* International Journal of Hydrogen Energy, 1999. **24**(8): p. 771-777.

30. Kerr, R.A., *Global warming. How urgent is climate change?* Science, 2007. **318**(5854): p. 1230-1.

31. Kerr, R.A., *Climate change. Another global warming icon comes under attack.* Science, 2007. **317**(5834): p. 28-9.

32. Zhu, Z., J. Jiang, and Y. Fa, *Overcoming the Biological Contamination in Microalgae and Cyanobacteria Mass Cultivations for Photosynthetic Biofuel Production.* Molecules, 2020. **25**(22).

33. Xue, C., et al., *Integrated butanol recovery for an advanced biofuel: current state and prospects.* Appl Microbiol Biotechnol, 2014. **98**(8): p. 3463-74.

34. Wu, W., Y.C. Lei, and J.S. Chang, *Life cycle assessment of upgraded microalgae-to-biofuel chains.* Bioresour Technol, 2019. **288**: p. 121492.

35. Ganesan, R., et al., *A review on prospective production of biofuel from microalgae.* Biotechnol Rep (Amst), 2020. **27**: p. e00509.

36. Davis, S.C., K.J. Anderson-Teixeira, and E.H. DeLucia, *Life-cycle analysis and the ecology of biofuels.* Trends in Plant Science, 2009. **14**(3): p. 140-146.

37. Dale, B.E., et al., *Take a Closer Look: Biofuels Can Support Environmental, Economic and Social Goals.* Environmental Science & Technology, 2014. **48**(13): p. 7200-7203.

38. Steinkraus, K.H., *Classification of fermented foods: worldwide review of household fermentation techniques.* Food Control, 1997. **8**(5-6): p. 311-317.

39. Becker, J. and C. Wittmann, *Advanced Biotechnology: Metabolically Engineered Cells for the Bio-Based Production of Chemicals and Fuels, Materials, and Health-Care Products.* Angewandte Chemie-International Edition, 2015. **54**(11): p. 3328-3350.

40. Armitage, J.P., et al., *Classic Spotlights: Selected Highlights from the First 100 Years of the Journal of Bacteriology.* J Bacteriol, 2017. **199**(13).

41. Naghshbandi, M.P., et al., *Metabolic Engineering of Microalgae for Biofuel Production.* Methods Mol Biol, 2020. **1980**: p. 153-172.

42. Liu, H., et al., *MEP pathway-mediated isopentenol production in metabolically engineered Escherichia coli.* Microb Cell Fact, 2014. **13**: p. 135.

43. Peskov, K., E. Mogilevskaya, and O. Demin, *Kinetic modelling of central carbon metabolism in Escherichia coli.* Febs Journal, 2012. **279**(18): p. 3374-3385.

44. Oliveira, A., et al., *A kinetic model of the central carbon metabolism for acrylic acid production in Escherichia coli.* Plos Computational Biology, 2021. **17**(3).

45. Kurata, H. and Y. Sugimoto, *Improved kinetic model of Escherichia coli central carbon metabolism in batch and continuous cultures.* Journal of Bioscience and Bioengineering, 2018. **125**(2): p. 251-257.

46. Bennett, B.D., et al., *Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli.* Nature Chemical Biology, 2009. **5**(8): p. 593-599.

47. Becker, J. and C. Wittmann, *Systems metabolic engineering of Escherichia coli for the heterologous production of high value molecules - a veteran at new shores.* Current Opinion in Biotechnology, 2016. **42**: p. 178-188.

48. Dai, Z.J. and J. Nielsen, *Advancing metabolic engineering through systems biology of industrial microorganisms.* Current Opinion in Biotechnology, 2015. **36**: p. 8-15.

49. San, K.Y., et al., *Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in Escherichia coli.* Metabolic Engineering, 2002. **4**(2): p. 182-192.

50. Metcalf, W.W., *Classic Spotlight: Metabolic Flux-Which Way To Go?* J Bacteriol, 2016. **198**(24): p. 3248-3249.

51. Lieber, D.J., et al., *A multienzyme complex channels substrates and electrons through acetyl-CoA and methane biosynthesis pathways in Methanosarcina.* PLoS One, 2014. **9**(9): p. e107563.

52. Henry, C.S., L.J. Broadbelt, and V. Hatzimanikatis, *Thermodynamics-based metabolic flux analysis.* Biophysical Journal, 2007. **92**(5): p. 1792-1805.

53. Atsumi, S., T. Hanai, and J.C. Liao, *Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels.* Nature, 2008. **451**(7174): p. 86-9.

54. Leuchtenberger, W., K. Huthmacher, and K. Drauz, *Biotechnological production of amino acids and derivatives: current status and prospects.* Appl Microbiol Biotechnol, 2005. **69**(1): p. 1-8.

55. Bongaerts, J., et al., *Metabolic engineering for microbial production of aromatic amino acids and derived compounds.* Metab Eng, 2001. **3**(4): p. 289-300.

56. Martin, V.J., et al., *Engineering a mevalonate pathway in Escherichia coli for production of terpenoids.* Nat Biotechnol, 2003. **21**(7): p. 796-802.

57.   McDaniel, R., et al., *Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products.* Proc Natl Acad Sci U S A, 1999. **96**(5): p. 1846-51.

58.   Voeste, T. and H. Buchold, *Production of Fatty Alcohols from Fatty-Acids.* Journal of the American Oil Chemists Society, 1984. **61**(2): p. 350-352.

59.   Ilina, A.I., Perchenk.Aa, and Marchenk.Ma, *Production of Alcohols from Intermediates of Production of Synthetic Fatty Acids.* International Chemical Engineering, 1970. **10**(3): p. 398-&.

60.   Lennen, R.M. and B.F. Pfleger, *Microbial production of fatty acid-derived fuels and chemicals.* Current Opinion in Biotechnology, 2013. **24**(6): p. 1044-1053.

61.   Steen, E.J., et al., *Microbial production of fatty-acid-derived fuels and chemicals from plant biomass.* Nature, 2010. **463**(7280): p. 559-U182.

62.   Dellomonaco, C., et al., *Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals.* Nature, 2011. **476**(7360): p. 355-9.

63.   White, S.W., et al., *The structural biology of type II fatty acid biosynthesis.* Annual Review of Biochemistry, 2005. **74**: p. 791-831.

64.   Ratledge, C., *Fatty acid biosynthesis in microorganisms being used for Single Cell Oil production.* Biochimie, 2004. **86**(11): p. 807-815.

65.   Liu, T.G., H. Vora, and C. Khosla, *Quantitative analysis and engineering of fatty acid biosynthesis in E. coli.* Metabolic Engineering, 2010. **12**(4): p. 378-386.

66.   Garwin, J.L., A.L. Klages, and J.E. Cronan, Jr., *Beta-ketoacyl-acyl carrier protein synthase II of Escherichia coli. Evidence for function in the thermal regulation of fatty acid synthesis.* J Biol Chem, 1980. **255**(8): p. 3263-5.

67.   Tehlivets, O., K. Scheuringer, and S.D. Kohlwein, *Fatty acid synthesis and elongation in yeast.* Biochim Biophys Acta, 2007. **1771**(3): p. 255-70.

68.   Toke, D.A. and C.E. Martin, *Isolation and characterization of a gene affecting fatty acid elongation in Saccharomyces cerevisiae.* J Biol Chem, 1996. **271**(31): p. 18413-22.

69.   Janssen, H.J. and A. Steinbuchel, *Fatty acid synthesis in Escherichia coli and its applications towards the production of fatty acid based biofuels.* Biotechnol Biofuels, 2014. **7**(1): p. 7.

70.   Henry, M.F. and J.E. Cronan, *Escherichia-Coli Transcription Factor That Both Activates Fatty-Acid Synthesis and Represses Fatty-Acid Degradation.* Journal of Molecular Biology, 1991. **222**(4): p. 843-849.

71.   Henry, M.F. and J.E. Cronan, *A New Mechanism of Transcriptional Regulation - Release of an Activator Triggered by Small Molecule-Binding.* Cell, 1992. **70**(4): p. 671-679.

72.   Dirusso, C.C., T.L. Heimert, and A.K. Metzger, *Characterization of Fadr, a Global Transcriptional Regulator of Fatty-Acid Metabolism in Escherichia-Coli - Interaction with the Fadb Promoter Is Prevented by Long-Chain Fatty Acyl Coenzyme-As.* Journal of Biological Chemistry, 1992. **267**(12): p. 8685-8691.

73.   Rock, C.O., et al., *Increased unsaturated fatty acid production associated with a suppressor of the fabA6(Ts) mutation in Escherichia coli.* Journal of Bacteriology, 1996. **178**(18): p. 5382-5387.

74.   Fujita, Y., H. Matsuoka, and K. Hirooka, *Regulation of fatty acid metabolism in bacteria.* Molecular Microbiology, 2007. **66**(4): p. 829-839.

75.   DiRusso, C.C. and T. Nystrom, *The fats of Escherichia coli during infancy and old age: regulation by global regulators, alarmones and lipid intermediates.* Molecular Microbiology, 1998. **27**(1): p. 1-8.

76.   Heath, R.J. and C.O. Rock, *Regulation of fatty acid elongation and initiation by acyl acyl carrier protein in Escherichia coli.* Journal of Biological Chemistry, 1996. **271**(4): p. 1833-1836.

77.   Magnuson, K., et al., *Regulation of Fatty-Acid Biosynthesis in Escherichia-Coli.* Microbiological Reviews, 1993. **57**(3): p. 522-542.

78.   Marrakchi, H., D. Patel, and C. Rock, *Regulation of fatty acid biosynthesis and degradation in Escherichia coli.* Faseb Journal, 2001. **15**(4): p. A192-A192.

79.   Nunn, W.D., D.L. Kelly, and M.Y. Stumfall, *Regulation of Fatty-Acid Synthesis during Cessation of Phospholipid Biosynthesis in Escherichia-Coli.* Journal of Bacteriology, 1977. **132**(2): p. 526-531.

80.   Zhang, F., J.M. Carothers, and J.D. Keasling, *Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids.* Nat Biotechnol, 2012. **30**(4): p. 354-9.

81.   Youngquist, J.T., et al., *Production of medium chain length fatty alcohols from glucose in Escherichia coli.* Metab Eng, 2013. **20**: p. 177-86.

82.   Liu, A., et al., *Fatty alcohol production in engineered E. coli expressing Marinobacter fatty acyl-CoA reductases.* Appl Microbiol Biotechnol, 2013. **97**(15): p. 7061-71.

83.   Akhtar, M.K., N.J. Turner, and P.R. Jones, *Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities.* Proc Natl Acad Sci U S A, 2013. **110**(1): p. 87-92.

84.   Haushalter, R.W., et al., *Development of an orthogonal fatty acid biosynthesis system in E. coli for oleochemical production.* Metab Eng, 2015. **30**: p. 1-6.

85.   Fatma, Z., et al., *Identification of long chain specific aldehyde reductase and its use in enhanced fatty alcohol production in E. coli.* Metab Eng, 2016. **37**: p. 35-45.

86.   Liu, Y., et al., *High production of fatty alcohols in Escherichia coli with fatty acid starvation.* Microb Cell Fact, 2016. **15**(1): p. 129.

87.   Chubukov, V., et al., *Engineering glucose metabolism of Escherichia coli under nitrogen starvation.* NPJ Syst Biol Appl, 2017. **3**: p. 16035.

88.   Fatma, Z., et al., *Model-assisted metabolic engineering of Escherichia coli for long chain alkane and alcohol production.* Metab Eng, 2018. **46**: p. 1-12.

89.   Hernandez Lozada, N.J., et al., *Production of 1-octanol in Escherichia coli by a high flux thioesterase route.* Metab Eng, 2020. **61**: p. 352-359.

90.   Akhtar, M.K., et al., *Microbial production of 1-octanol: A naturally excreted biofuel with diesel-like properties.* Metab Eng Commun, 2015. **2**: p. 1-5.

91.   Cao, Y.X., et al., *Biosynthesis of odd-chain fatty alcohols in Escherichia coli.* Metab Eng, 2015. **29**: p. 113-123.

92.   Jiang, W., et al., *Modular pathway engineering for the microbial production of branched-chain fatty alcohols.* Biotechnol Biofuels, 2017. **10**: p. 244.

93.   Krishnan, A., B.A. McNeil, and D.T. Stuart, *Biosynthesis of Fatty Alcohols in Engineered Microbial Cell Factories: Advances and Limitations.* Front Bioeng Biotechnol, 2020. **8**: p. 610936.

94.   Gupta, P. and S.C. Phulara, *Metabolic engineering for isoprenoid-based biofuel production.* Journal of Applied Microbiology, 2015. **119**(3): p. 605-619.

95.   Niu, F.X., et al., *Metabolic engineering for the microbial production of isoprenoids: Carotenoids and isoprenoid-based biofuels.* Synthetic and Systems Biotechnology, 2017. **2**(3): p. 167-175.

96.   Yang, J.M., et al., *Bio-isoprene production using exogenous MVA pathway and isoprene synthase in Escherichia coli.* Bioresource Technology, 2012. **104**: p. 642-647.

97.   Withers, S.T., et al., *Identification of isopentenol biosynthetic genes from Bacillus subtilis by a screening method based on isoprenoid precursor toxicity.* Appl Environ Microbiol, 2007. **73**(19): p. 6277-83.

98.   Vickers, C.E., et al., *Metabolic engineering of volatile isoprenoids in plants and microbes.* Plant Cell Environ, 2014. **37**(8): p. 1753-75.

99.   Vickers, C.E. and S. Sabri, *Isoprene.* Adv Biochem Eng Biotechnol, 2015. **148**: p. 289-317.

100.   Schewe, H., M.A. Mirata, and J. Schrader, *Bioprocess engineering for microbial synthesis and conversion of isoprenoids.* Adv Biochem Eng Biotechnol, 2015. **148**: p. 251-86.

101.   Roberts, S.C., *Production and engineering of terpenoids in plant cell culture.* Nature Chemical Biology, 2007. **3**(7): p. 387-395.

102.   Rodriguez-Concepcion, M. and A. Boronat, *Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics.* Plant Physiology, 2002. **130**(3): p. 1079-1089.

103.   Klein-Marcuschamer, D., P.K. Ajikumar, and G. Stephanopoulos, *Engineering microbial cell factories for biosynthesis of isoprenoid molecules: beyond lycopene.* Trends Biotechnol, 2007. **25**(9): p. 417-24.

104.   Wang, Z., et al., *Metabolic Engineering Escherichia coli for the Production of Lycopene.* Molecules, 2020. **25**(14).

105.   George, K.W., et al., *Integrated analysis of isopentenyl pyrophosphate (IPP) toxicity in isoprenoid-producing Escherichia coli.* Metab Eng, 2018. **47**: p. 60-72.

106.   Rohmer, M., *Isoprenoid biosynthesis via the mevalonate-independent route, a novel target for antibacterial drugs?* Prog Drug Res, 1998. **50**: p. 135-54.

107.   Rohmer, M., et al., *Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate.* Biochem J, 1993. **295 (Pt 2)**: p. 517-24.

108. Schwender, J., et al., *Biosynthesis of isoprenoids (carotenoids, sterols, prenyl side-chains of chlorophylls and plastoquinone) via a novel pyruvate/glyceraldehyde 3-phosphate non-mevalonate pathway in the green alga Scenedesmus obliquus.* Biochem J, 1996. **316 (Pt 1)**: p. 73-80.

109. Rudney, H., *The biosynthesis of beta-hydroxy-beta-methylglutaric acid.* J Biol Chem, 1957. **227**(1): p. 363-77.

110. Goyal, G., et al., *Parallel Integration and Chromosomal Expansion of Metabolic Pathways.* ACS Synth Biol, 2018. **7**(11): p. 2566-2576.

111. Tyo, K.E., P.K. Ajikumar, and G. Stephanopoulos, *Stabilized gene duplication enables long-term selection-free heterologous pathway expression.* Nat Biotechnol, 2009. **27**(8): p. 760-5.

112. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6640-5.

113. Yu, D., et al., *An efficient recombination system for chromosome engineering in Escherichia coli.* Proc Natl Acad Sci U S A, 2000. **97**(11): p. 5978-83.

114. Sharan, S.K., et al., *Recombineering: a homologous recombination-based method of genetic engineering.* Nat Protoc, 2009. **4**(2): p. 206-23.

115. Warner, J.R., et al., *Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides.* Nat Biotechnol, 2010. **28**(8): p. 856-62.

116. Costantino, N. and D.L. Court, *Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants.* Proc Natl Acad Sci U S A, 2003. **100**(26): p. 15748-53.

117. Posfai, G., et al., *Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome.* Nucleic Acids Res, 1999. **27**(22): p. 4409-15.

118. Porteus, M.H. and D. Baltimore, *Chimeric nucleases stimulate gene targeting in human cells.* Science, 2003. **300**(5620): p. 763.

119. Miller, J.C., et al., *An improved zinc-finger nuclease architecture for highly specific genome editing.* Nat Biotechnol, 2007. **25**(7): p. 778-85.

120. Sander, J.D., et al., *Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA).* Nat Methods, 2011. **8**(1): p. 67-9.

121. Wood, A.J., et al., *Targeted genome editing across species using ZFNs and TALENs.* Science, 2011. **333**(6040): p. 307.

122. Hockemeyer, D., et al., *Genetic engineering of human pluripotent cells using TALE nucleases.* Nat Biotechnol, 2011. **29**(8): p. 731-4.

123. Reyon, D., et al., *FLASH assembly of TALENs for high-throughput genome editing.* Nat Biotechnol, 2012. **30**(5): p. 460-5.

124. Sanjana, N.E., et al., *A transcription activator-like effector toolbox for genome engineering.* Nat Protoc, 2012. **7**(1): p. 171-92.

125. Boch, J., et al., *Breaking the code of DNA binding specificity of TAL-type III effectors.* Science, 2009. **326**(5959): p. 1509-12.

126. Moscou, M.J. and A.J. Bogdanove, *A simple cipher governs DNA recognition by TAL effectors.* Science, 2009. **326**(5959): p. 1501.

127. Christian, M., et al., *Targeting DNA double-strand breaks with TAL effector nucleases.* Genetics, 2010. **186**(2): p. 757-61.

128. Pougach, K., et al., *Transcription, processing and function of CRISPR cassettes in Escherichia coli.* Mol Microbiol, 2010. **77**(6): p. 1367-79.

129. Wu, H., et al., *Quantitatively relating gene expression to light intensity via the serial connection of blue light sensor and CRISPRi.* ACS Synth Biol, 2014. **3**(12): p. 979-82.

130. Ren, X., et al., *Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila.* Cell Rep, 2014. **9**(3): p. 1151-62.

131. Beloglazova, N., et al., *CRISPR RNA binding and DNA target recognition by purified Cascade complexes from Escherichia coli.* Nucleic Acids Res, 2015. **43**(1): p. 530-43.

132. Bassalo, M.C., et al., *Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli.* ACS Synth Biol, 2016. **5**(7): p. 561-8.

133. Kannan, K., et al., *One step engineering of the small-subunit ribosomal RNA using CRISPR/Cas9.* Sci Rep, 2016. **6**: p. 30714.

134. Sabri, S., et al., *Knock-in/Knock-out (KIKO) vectors for rapid integration of large DNA sequences, including whole metabolic pathways, onto the Escherichia coli chromosome at well-characterised loci.* Microb Cell Fact, 2013. **12**: p. 60.

135. Jiang, W., et al., *RNA-guided editing of bacterial genomes using CRISPR-Cas systems.* Nat Biotechnol, 2013. **31**(3): p. 233-9.

136. DiCarlo, J.E., et al., *Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.* Nucleic Acids Res, 2013. **41**(7): p. 4336-43.

137. Cobb, R.E., Y. Wang, and H. Zhao, *High-efficiency multiplex genome editing of Streptomyces species using an engineered CRISPR/Cas system.* ACS Synth Biol, 2015. **4**(6): p. 723-8.

138. Shan, Q., et al., *Targeted genome modification of crop plants using a CRISPR-Cas system.* Nat Biotechnol, 2013. **31**(8): p. 686-8.

139. Wang, Y., et al., *The CRISPR/Cas system mediates efficient genome engineering in Bombyx mori.* Cell Res, 2013. **23**(12): p. 1414-6.

140. Yu, Z., et al., *Highly efficient genome modifications mediated by CRISPR/Cas9 in Drosophila.* Genetics, 2013. **195**(1): p. 289-91.

141. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.

142. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.

143. Zhang, Q., et al., *CRISPR-Cas systems target a diverse collection of invasive mobile genetic elements in human microbiomes.* Genome Biol, 2013. **14**(4): p. R40.

144. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nat Protoc, 2013. **8**(11): p. 2281-2308.

145. Chen, J., et al., *DeviceEditor visual biological CAD canvas.* J Biol Eng, 2012. **6**(1): p. 1.

146. Hillson, N.J., R.D. Rosengarten, and J.D. Keasling, *j5 DNA assembly design automation software.* ACS Synth Biol, 2012. **1**(1): p. 14-21.

147. Ham, T.S., et al., *Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools.* Nucleic Acids Res, 2012. **40**(18): p. e141.

148. Gibson, D.G., *Enzymatic assembly of overlapping DNA fragments.* Methods Enzymol, 2011. **498**: p. 349-61.

149. Gibson, D.G., *Gene and genome construction in yeast.* Curr Protoc Mol Biol, 2011. **Chapter 3**: p. Unit3 22.

150. George, K.W., et al., *Correlation analysis of targeted proteins and metabolites to assess and engineer microbial isopentenol production.* Biotechnol Bioeng, 2014. **111**(8): p. 1648-58.

151. Juminaga, D., et al., *Modular engineering of L-tyrosine production in Escherichia coli.* Appl Environ Microbiol, 2012. **78**(1): p. 89-98.

152. Jiang, M. and B.A. Pfeifer, *Metabolic and pathway engineering to influence native and altered erythromycin production through E. coli.* Metabolic Engineering, 2013. **19**: p. 42-49.

153. Alper, H., et al., *Tuning genetic control through promoter engineering.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(36): p. 12678-12683.

154. Alper, H., et al., *Tuning genetic control through promoter engineering (vol 102, pg 12678, 2005).* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(8): p. 3006-3006.

155. Mutalik, V.K., et al., *Precise and reliable gene expression via standard transcription and translation initiation elements.* Nat Methods, 2013. **10**(4): p. 354-60.

156. Friehs, K., *Plasmid copy number and plasmid stability.* Adv Biochem Eng Biotechnol, 2004. **86**: p. 47-82.

157. Ublinskaya, A.A., et al., *A PCR-free cloning method for the targeted phi80 Int-mediated integration of any long DNA fragment, bracketed with meganuclease recognition sites, into the Escherichia coli chromosome.* J Microbiol Methods, 2012. **89**(3): p. 167-73.

158. Hook, C.D., et al., *A novel approach for Escherichia coli genome editing combining in vivo cloning and targeted long-length chromosomal insertion.* J Microbiol Methods, 2016. **130**: p. 83-91.

159. Kuhlman, T.E. and E.C. Cox, *Site-specific chromosomal integration of large synthetic constructs.* Nucleic Acids Res, 2010. **38**(6): p. e92.

160. Baumgartner, F., et al., *Construction of Escherichia coli strains with chromosomally integrated expression cassettes for the synthesis of 2'-fucosyllactose.* Microb Cell Fact, 2013. **12**: p. 40.

161. Martinez-Morales, F., et al., *Chromosomal integration of heterologous DNA in Escherichia coli with precise removal of markers and replicons used during construction.* J Bacteriol, 1999. **181**(22): p. 7143-8.

162. Sukhija, K., et al., *Developing an Extended Genomic Engineering Approach Based on Recombineering to Knock-in Heterologous Genes to Escherichia coli Genome.* Molecular Biotechnology, 2012. **51**(2): p. 109-118.

163. Haldimann, A. and B.L. Wanner, *Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria.* J Bacteriol, 2001. **183**(21): p. 6384-93.

164. Sauer, B., *Site-specific recombination: developments and applications.* Curr Opin Biotechnol, 1994. **5**(5): p. 521-7.

165. Mizuuchi, K., *Transpositional recombination: mechanistic insights from studies of mu and other elements.* Annu Rev Biochem, 1992. **61**: p. 1011-51.

166. Wei, X.X., et al., *A mini-Mu transposon-based method for multiple DNA fragment integration into bacterial genomes.* Appl Microbiol Biotechnol, 2010. **87**(4): p. 1533-41.

167. Song, C.W. and S.Y. Lee, *Rapid one-step inactivation of single or multiple genes in Escherichia coli.* Biotechnol J, 2013. **8**(7): p. 776-84.

168. Domi, A. and B. Moss, *Engineering of a vaccinia virus bacterial artificial chromosome in Escherichia coli by bacteriophage lambda-based recombination.* Nat Methods, 2005. **2**(2): p. 95-7.

169. Yang, J., et al., *High-efficiency scarless genetic modification in Escherichia coli by using lambda red recombination and I-SceI cleavage.* Appl Environ Microbiol, 2014. **80**(13): p. 3826-34.

170. Gu, P., et al., *A rapid and reliable strategy for chromosomal integration of gene(s) with multiple copies.* Sci Rep, 2015. **5**: p. 9684.

171. Kolter, R., M. Inuzuka, and D.R. Helinski, *Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K.* Cell, 1978. **15**(4): p. 1199-208.

172. Metcalf, W.W., W. Jiang, and B.L. Wanner, *Use of the rep technique for allele replacement to construct new Escherichia coli hosts for maintenance of R6K gamma origin plasmids at different copy numbers.* Gene, 1994. **138**(1-2): p. 1-7.

173. Chen, Y.Y., et al., *Chromosomal evolution of Escherichia coli for the efficient production of lycopene.* Bmc Biotechnology, 2013. **13**.

174. Jiang, P. and J.E. Cronan, Jr., *Inhibition of fatty acid synthesis in Escherichia coli in the absence of phospholipid synthesis and release of inhibition by thioesterase action.* J Bacteriol, 1994. **176**(10): p. 2814-21.

175. Davis, M.S. and J.E. Cronan, *Inhibition of Escherichia coli acetyl coenzyme A carboxylase by acyl-acyl carrier protein.* Journal of Bacteriology, 2001. **183**(4): p. 1499-1503.

176. Black, P.N., N.J. Faergeman, and C.C. DiRusso, *Long-chain acyl-CoA-dependent regulation of gene expression in bacteria, yeast and mammals.* J Nutr, 2000. **130**(2S Suppl): p. 305S-309S.

177. Bokinsky, G., et al., *Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered Escherichia coli.* Proc Natl Acad Sci U S A, 2011. **108**(50): p. 19949-54.

178. Hellenbrand, J., et al., *Fatty acyl-CoA reductases of birds.* BMC Biochem, 2011. **12**: p. 64.

179. Rowland, O. and F. Domergue, *Plant fatty acyl reductases: enzymes generating fatty alcohols for protective layers with potential for industrial applications.* Plant Sci, 2012. **193-194**: p. 28-38.

180. Teerawanichpan, P. and X. Qiu, *Molecular and functional analysis of three fatty acyl-CoA reductases with distinct substrate specificities in copepod Calanus finmarchicus.* Mar Biotechnol (NY), 2012. **14**(2): p. 227-36.

181. Opgenorth, P., et al., *Lessons from Two Design-Build-Test-Learn Cycles of Dodecanol Production in Escherichia coli Aided by Machine Learning.* ACS Synth Biol, 2019. **8**(6): p. 1337-1351.

182. Alonso-Gutierrez, J., et al., *Toward industrial production of isoprenoids in Escherichia coli: Lessons learned from CRISPR-Cas9 based optimization of a chromosomally integrated mevalonate pathway.* Biotechnol Bioeng, 2018. **115**(4): p. 1000-1013.

183. Blattner, F.R., et al., *The complete genome sequence of Escherichia coli K-12.* Science, 1997. **277**(5331): p. 1453-62.

184. Heintz, N. and S. Gong, *Two-Step Bacterial Artificial Chromosome (BAC) Engineering: Preparation and Verification of the Recombinant Shuttle Vector.* Cold Spring Harb Protoc, 2020. **2020**(4): p. 098061.

185. Sorokin, A., et al., *A new approach using multiplex long accurate PCR and yeast artificial chromosomes for bacterial chromosome mapping and sequencing.* Genome Res, 1996. **6**(5): p. 448-53.

186. Wang, H.H., et al., *Programming cells by multiplex genome engineering and accelerated evolution.* Nature, 2009. **460**(7257): p. 894-898.

187. Lee, T.S., et al., *BglBrick vectors and datasheets: a synthetic biology platform for gene expression.* Journal of biological engineering, 2011. **5**(1): p. 12.

188. Doublet, B., et al., *Antibiotic marker modifications of λ Red and FLP helper plasmids, pKD46 and pCP20, for inactivation of chromosomal genes using PCR products in multidrug-resistant strains.* Journal of microbiological methods, 2008. **75**(2): p. 359-361.

189. Thanbichler, M., A.A. Iniesta, and L. Shapiro, *A comprehensive set of plasmids for vanillate- and xylose-inducible gene expression in Caulobacter crescentus.* Nucleic Acids Res, 2007. **35**(20): p. e137.

190. Yu, B.J., et al., *Rapid and efficient construction of markerless deletions in the Escherichia coli genome.* Nucleic acids research, 2008. **36**(14): p. e84-e84.